



Bundesministerium
für Bildung
und Forschung

Steuerung im Bildungssystem

Implementation und Wirkung neuer
Steuerungsinstrumente im Schulwesen

Bildungsforschung Band 43

Steuerung im Bildungssystem

Implementation und Wirkung neuer
Steuerungsinstrumente im Schulwesen

Inhalt

Jasmin Tarkian, Felicitas Thiel: Einleitung: „Steuerung im Bildungssystem“ (SteBis) – Bilanz der Befunde aus der ersten Förderphase	3
1. Nutzung von Test- und Prüfungsergebnissen für evidenzbasiertes Handeln	7
Olga Zlatkin-Troitschanskaia, Lena M. Zimmer, Olga Mater, Bastian Laier, Anna Rosa Koch, Carmen Binnewies, Christian Dormann, Isabell van Ackeren, Marten Clausen, Peter Preisendörfer, Uwe Schmidt, Denise Demski, Daja Preuße, Martin Stump: Schulische und individuelle Einflussfaktoren auf das evidenzbasierte Handeln von Lehrkräften und Schulleitungen – eine mehrebenenanalytische Studie	8
Harm Kuper, Uwe Maier, Tanja Graf, Barbara Muslic, Carolin Ramsteck: Datenbasierte Schulentwicklung mit Vergleichsarbeiten aus der Pers- pektive von Lehrkräften, Fachkonferenzleitungen, Schulleitungen und Schulaufsichten – Qualitative Fallstudien aus vier Bundesländern	39
Marc Piopiunik, Guido Schwerdt, Ludger Wößmann: Zentrale Abschlussprüfungen als Steuerungsinstrument im Schulsystem: Arbeitsmarkteffekte und Wirkungsmechanismen	68
2. Externe Evaluation durch Schulinspektion: Steuerungsimpulse und Datenrezeption	85
Kathrin Dederling: Der Umgang von Schulen mit Schulinspektionen. – Zum „Projekt „Externe Evaluation/Schulinspektion und Qualitätsent- wicklung von Schulen“	86
Oliver Böhm-Kasper, Thomas Brüsemeister, Fabian Dietrich, Lisa Gromala, Martin Heinrich, Maike Lambrecht, Bianca E. Preuß, Matthias Rürup, Odette Selders, Jochen Wissinger: Schulinspektion als Steuerungsimpuls zur Schulentwicklung und seine Realisierungsbedingungen auf einzelschulischer Ebene – Ergebnisse eines triangulativ orientierten Verbundprojekts	110

3. Qualitätsentwicklung von Schulen in sozialräumlich deprivierter Lage 137

Isabell van Ackeren, Kathrin Racherbäumer, Christina Funke, Marten Clausen: Herausforderungen für und Bewältigungsstrategien von Lehrkräften an Schulen in sozialräumlich deprivierter Lage 138

Laura Fölker, Thorsten Hertel, Nicolle Pfaff: Schulische Praxis und Schulentwicklung unter Bedingungen sozialräumlicher Segregation – Befunde einer fallvergleichenden Studie zu zwei Schulen in benachteiligten Stadtteilen 161

4. Schulleitungen, Fachkonferenzleitungen und Lehrkräfte als Akteure Neuer Steuerung 177

Sebastian Wurster, Anabel Bach, Anna Schliesing, Katja Thillmann, Hans Anand Pant, Felicitas Thiel: Schulen als Steuerungsakteure im Bildungssystem – datenbasierte Schul- und Unterrichtsentwicklung aus der Perspektive von Schulleitungen, Fachkonferenzleitungen und Lehrkräften 178

Johannes Bellmann, Sebastian Schweizer, Corrie Thiel: Nebenfolgen Neuer Steuerung unter Bedingungen von „low-stakes“ und „no-stakes“ – Qualitative und quantitative Befunde einer Untersuchung in vier Bundesländern 208

Stefan Brauckmann: Schulleitungshandeln zwischen erweiterten Rechten und Pflichten (SHaRP) unter Berücksichtigung einschlägiger schulrechtlicher Bestimmungen 238

Autorinnen und Autoren 256

Impressum 259

Jasmin Tarkian, Felicitas Thiel

Einleitung: „Steuerung im Bildungssystem“ (SteBis) – Bilanz der Befunde aus der ersten Förderphase

Zentrale Fragen empirischer Schulforschung

In Reaktion auf die Ergebnisse in den ersten internationalen Schulvergleichsstudien (2001) haben die Länder der Bundesrepublik Deutschland damit begonnen, verschiedene Maßnahmen der Qualitätssicherung und -entwicklung im staatlichen Schulwesen zu implementieren. Neben der rückblickend als historisch bewerteten Leistung (Maritzen 2014) einer gemeinsamen Verständigung aller Bundesländer auf zentrale Bildungsstandards und deren Überprüfung im Ländervergleich zählt dazu insbesondere die Einführung neuer, extern administrierter Steuerungsinstrumente, wie Verfahren der Schulinspektion, zentrale Abschlussprüfungen und testbasierte Vergleichsarbeiten zur Feststellung von Schülerkompetenzen (Altrichter & Maag Merki 2012).

Mit den implementierten Maßnahmen ist der Anspruch verbunden, den Grad der Erreichung bildungspolitischer Zielstellungen für unterschiedliche Akteursgruppen transparent zu machen (Accountability) sowie Entwicklungsstatus und Entwicklungstrends auf unterschiedlichen Ebenen kontinuierlich zu erfassen, um auf dieser Grundlage belastbares Steuerungswissen für Akteure in Bildungspolitik und -administration zu generieren (Bildungsmonitoring). Zugleich bestand von Beginn an die Erwartung, dass die erhobenen Leistungs- und Prozessdaten auch zur formativ-reflektiven Selbststeuerung in den Schulen genutzt werden und damit als Ergänzung innerschulischer Selbstevaluation zu einer nachhaltigen Verbesserung schulischer Prozesse und Outputs beitragen (Entwicklungsfunktion). Steuerung bezieht sich in diesem Sinne also über eine staatlich-hoheitliche Regulierung hinaus auch auf das zielgerichtete Handeln professioneller Akteure in den Schulen selbst. Entsprechend ist, neben der Einführung systematischer Evaluationsverfahren, die Verlagerung von Entscheidungs- und damit Steuerungskompetenzen ein Hauptmerkmal des neuen Steuerungsmodells (Thiel, Cortina & Pant 2014).

Die Implementation der Maßnahmen und Instrumente war eingebunden in den konzeptionellen Rahmen einer bildungspolitischen Gesamtstrategie der Kultusministerkonferenz (2006; 2012), die gleichwohl Raum ließ für unterschiedliche länderspezifische Akzentuierungen. Einrichtungen zur Unterstützung der Bildungsministerien bei der Entwicklung und Durchführung der neuen Aufgaben wie Länderinstitute und Schulinspektorate sind entstanden bzw. funktional erweitert worden. Daraus ergeben sich für die traditionellen Akteure der Schulaufsicht zwangsläufig die Notwendigkeit der Neudefinition ihrer Aufgaben sowie ein spezifischer Qualifizierungsbedarf. Dass die Aktivitäten von Datenproduzenten, staatlichen Aufsichts-

behörden, Schulleitungen, Lehrpersonen und Eltern/Schülern koordiniert werden, ist eine Erfolgsbedingung des neuen Steuerungsmodells.

Welche Wirkungen die neu implementierten Instrumente und Verfahren im Schulsystem im Einzelnen auslösen, welche Faktoren eine konstruktive Nutzung der Datenrückmeldungen – mit Blick auf die innerschulische Koordination und Strukturbildung, die Unterrichtsentwicklung sowie insbesondere die Verbesserung von Lernergebnissen – begünstigen oder aber ihr entgegenstehen und welche nicht intendierten Effekte mit der Einführung der Maßnahmen verbunden sind, gilt es empirisch zu klären.

Der Forschungsschwerpunkt

In dem seit 2010 vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungsschwerpunkt „**Steuerung im Bildungssystem**“ (SteBis) werden diese Fragen von Wissenschaftlerinnen und Wissenschaftlern aus den Erziehungs- und Sozialwissenschaften, der Psychologie und der Bildungsökonomie verschiedener Universitäten und Forschungseinrichtungen aufgegriffen und mit Blick auf unterschiedliche Akteursebenen (z. B. Schulleitungen, Lehrkräfte, Schulaufsicht) unter Einsatz unterschiedlicher Studiendesigns erforscht.

Dabei stehen mit mehrebenenanalytischen, zum Teil über die Grenzen des Schulsystems hinausblickenden Untersuchungen zur *Nutzung von Test- und Prüfungsergebnissen für evidenzbasiertes Handeln* (1), mit ländervergleichenden Analysen zur Wirkung und Rezeption *externer Evaluation durch Schulinspektion* (2), mit Arbeiten, die den Blick gezielt auf die *Qualitätsentwicklung von Schulen in sozial-räumlich deprivierter Lage* richten und insofern Bedingungen für einen „Turnaround“ von Schulen in Brennpunktlagen rekonstruieren (3), sowie mit Beiträgen, die auf der Ebene der Schulorganisation Steuerungsaktivitäten von *Schulleitungen, Fachkonferenzleitungen und Lehrkräften als Akteure Neuer Steuerung* (4) untersuchen, vier Themenfelder Neuer Steuerung im Fokus der Forschungsprojekte.

Die Koordinierungsstelle

Zur Unterstützung der am Forschungsschwerpunkt beteiligten Projekte ist an der Freien Universität Berlin im Frühjahr 2011 die Koordinierungsstelle SteBis eingerichtet worden. Durch Maßnahmen der Vernetzung, der Öffentlichkeitsarbeit und der Organisation von Transferveranstaltungen fördert die Koordinierungsstelle parallel zu den Forschungsaktivitäten einen wechselseitigen wissenschaftlichen wie auch erfahrungsbezogenen Austausch zwischen den Projektnehmerinnen und Projektnehmern und anderen zentralen Akteuren in Bildungspolitik, -administration, -forschung und -praxis. Koordinierte Auftritte bei nationalen und internationalen Kongressen, gemeinsame fach- und praxisorientierte Veröffentlichungen sowie die Ausrichtung themenfokussierter Workshops und Tagungen mit zentralen Stakeholdern und (inter-)nationalen Experten bilden dabei Kernaufgaben, mit

denen der Dialog und der Wissenstransfer zwischen den Beteiligten kontinuierlich gefördert und ausgebaut wird.

Zur Publikation

Entsprechend diesem Aufgabentableau hat die Koordinierungsstelle auch die Herausgabe der vorliegenden Publikation organisatorisch unterstützt, die der abschließenden Gesamtdarstellung zentraler Projektergebnisse aller zwischen 2010 und 2014 geförderten zehn Forschungsprojekte dient. Diese waren überwiegend als Einzelvorhaben, zum Teil auch als standortübergreifende Verbundvorhaben mit mehreren Projektpartnern über eine jeweils dreijährige Laufzeit durchgeführt worden. Die Projekte stellen in dieser Publikation ihr Forschungsdesign sowie zentrale Befunde ihrer Untersuchungen vor. Die Gliederung des Bandes entspricht den vier oben beschriebenen Schwerpunktfeldern.

Über weitere Veröffentlichungen aus den Projekten, die laufenden Forschungsaktivitäten in den vom Bundesministerium in einer zweiten Phase anschlussgeförderten Vorhaben sowie schwerpunktübergreifende Befunde, Veranstaltungen und Veröffentlichungen zu Fragen der Steuerung im Schul- und Bildungssystem informiert die Koordinierungsstelle unter www.stebis.de bzw. über einen mehrmals jährlich erscheinenden Newsletter, für den Sie sich auf der Website registrieren können.

Kontakt Daten

SteBis – Koordinierungsstelle für den Forschungsschwerpunkt *Steuerung im Bildungssystem*

Freie Universität Berlin
Arbeitsbereich Schulpädagogik/Schulentwicklungsforschung
Habelschwerdter Allee 45
14195 Berlin
Telefon: 030 838 51663

E-Mail: stebis@fu-berlin.de
www.stebis.de

Ansprechpartnerin:
Jasmin Tarkian, Dipl.-Soz. M.A.

Leitung der Koordinierungsstelle:
Prof. Dr. Felicitas Thiel

Literatur

- Altrichter, H. & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 15–39, Wiesbaden: VS Verlag für Sozialwissenschaften.
- KMK (Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2006). *Gesamtstrategie der Kultusministerkonferenz – Erläuterungen zur Konzeption und Entwicklung*. München: Luchterhand.
- KMK (Sekretariat der ständigen 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: 13.07.2015.
- Maritzen, N. (2014). Glanz und Elend der KMK-Gesamtstrategie zum Bildungsmonitoring. Versuch einer Bilanz und eines Ausblicks. *Die Deutsche Schule*, 106 (4), S. 401–416.
- Thiel, F., Cortina, K. S. & Pant, H. A. (2014). Steuerung im Bildungssystem im internationalen Vergleich. In R. Fatke & J. Oelkers (Hrsg.), *Das Selbstverständnis der Erziehungswissenschaft: Geschichte und Gegenwart*, S. 123–138. Weinheim: Beltz.

1. Nutzung von Test- und Prüfungsergebnissen für evidenzbasiertes Handeln

*Olga Zlatkin-Troitschanskaia, Lena M. Zimmer,
Olga Mater, Bastian Laier, Anna Rosa Koch, Carmen
Binnewies, Christian Dormann, Isabell van Ackeren,
Marten Clausen, Peter Preisendörfer, Uwe Schmidt,
Denise Demski, Daja Preuß, Martin Stump*

Schulische und individuelle Einflussfaktoren auf das evidenzbasierte Handeln von Lehrkräften und Schulleitungen – eine mehrbenenanalytische Studie

1. Problemstellung und Ziele

Das Konzept der Evidenzbasierung¹ von Steuerung und professionellem Handeln erfährt als eine umfassende Innovation auch im Schulwesen zunehmende Beachtung. Kern dieses Konzeptes ist das Ziel, empirisch hinreichend belegte Erkenntnisse als Steuerungswissen für Politik und Praxis nutzbar zu machen, um eine wirksame und nachhaltige Schul- und Unterrichtsentwicklung zu unterstützen (vgl. Thiel et al. 2014a). Solches Wissen wird für die schulische Praxis in einer Reihe von Bereichen produziert. Neben dem Wissenschaftssystem (einschließlich der Lehrerbildung) sind in zunehmendem Maße auch staatliche Stellen (wie Schulinspektionen), internationale Organisationen (wie OECD) und privatwirtschaftliche Anbieter (wie Verlage) mit der Generierung und Vermittlung einschlägiger Wissensbestände befasst.²

1 Das Konzept orientiert sich an dem sog. Evidence-Based Management (EBM), das ausgehend von den Forschungen und praktischen Erfahrungen in der Medizin (z. B. Sackett et al. 2000) die Evidenzbasierung von Entscheidungen und Praktiken in Organisationen beschreibt und in zahlreichen nationalen und internationalen Studien zur Anwendung kam (s. Walshe & Rundall 2001; Pfeffer & Sutton 2006a, 2006b). Zudem findet man in der Forschung eine ganze Reihe von verwandten Konzepten wie etwa „research practice gap“ (z. B. Latham 2007; Rousseau 2006), „great divide“ (z. B. Anderson 2007), „organizational knowledge management“ (Miller & Niakanta 2001), „organizational knowledge formation“ (Song et al. 2008) oder „research utilization“ (Smith et al. 2007). Obgleich eindeutige Definitionen und klare Abgrenzungen zwischen den Konzepten nicht existieren (vgl. Ashkanasy 2007), ist ihnen gemeinsam, dass Managemententscheidungen und Praktiken in Organisationen auf den am besten geeigneten, verfügbaren wissenschaftlichen Fakten basieren sollten: „Translating principles based on best evidence into organizational practices“ (Rousseau 2006, S. 256).

2 Im Zusammenhang mit der Erforschung evidenzbasierten organisationalen Handelns wird der Evidenzbegriff aus dem Englischen entliehen und meint die Überprüf- bzw. Begründbarkeit von Annahmen (vgl. Jornitz 2008). Im Rahmen einschlägiger Forschungen wird Evidenz oft mit wissenschaftlichen Erkenntnissen gleichgesetzt. Letztlich geht es dabei um eine bestimmte (wissenschaftliche) Methode der Wissensgenerierung, d. h. um systematisches, empirisch gewonnenes und in seinen Voraussetzungen und seiner Gewinnung transparentes, objektiviertes Wissen (vgl. Maasen 1999).

Wie empirische Evidenzen zeigen, kann man jedoch nicht davon ausgehen, dass sich wissenschaftlich fundierte Erkenntnisse umstandslos in evidenzbasiertes Handeln umsetzen lassen (s. z. B. Gärtner et al. 2014; die Beiträge in Thiel et al. 2014b; Zlatkin-Troitschanskaia 2015). Bildungspolitisch intendierte Effekte lassen sich oftmals weder vollständig noch kurzfristig herbeiführen. Vielmehr gibt es empirisch fundierte Hinweise, dass vorhandene persönliche Dispositionen und Handlungsroutinen der Akteure im Schulwesen, organisationale Strukturen, Kulturen und Kontextfaktoren sowie nicht zuletzt auch die Struktur des zur Verfügung stehenden (Steuerungs- bzw. Schulentwicklungs-)Wissens die Realisierung bildungspolitisch angestrebter Ziele in der administrativen und pädagogischen Praxis erschweren (vgl. beispielhaft für Deutschland Maier & Kuper 2012; Thiel & Pant 2014b; Dederig & Müller 2011; für internationale Befunde vgl. z. B. Ehren et al. 2013; Breiter & Karbautzki 2012; Wayman et al. 2012; Schildkamp & Kuiper 2010).

Das BMBF-geförderte Forschungsprojekt EVIS („Evidenzbasiertes Handeln im schulischen Mehrebenensystem – Bedingungen, Prozesse und Wirkungen“) setzt an diesem gleichwohl wissenschaftlich als auch bildungspolitisch und -praktisch relevanten Desiderat an und untersucht Bedingungen, Prozesse und Wirkungen evidenzbasierten (Steuerungs-)Handelns im schulischen Mehrebenensystem, um belastbare Erkenntnisse für die Wirksamkeit und Wirkungen einer evidenzbasierten Steuerungspraxis im deutschen Schulwesen zu gewinnen. EVIS hat zum Ziel, auf der Basis von mehrperspektivischen theoretischen Analysezugängen und mithilfe unterschiedlicher qualitativer und quantitativer methodischer Ansätze die Ausprägungen und Einflussfaktoren evidenzbasierten Handelns in Schulen zu modellieren und zu beschreiben (vgl. van Ackeren et al. 2011).

Zur Untersuchung des evidenzbasierten Handelns in Schulen ist das Gesamtvorhaben EVIS unterteilt in ein Zentralprojekt und zwei Teilprojekte. Im Zentralprojekt liegt der Fokus auf der umfassenden Analyse der Bedingungen von evidenzbasiertem Handeln in Schulen, im Teilprojekt 1 auf der Betrachtung der Genese evidenzbasierten Handelns und im Teilprojekt 2 auf vertiefenden Interviews und Netzwerkanalysen an Schulen mit hoher vs. niedriger Evidenzbasierung.

In der Forschung zu evidenzbasierten Praktiken lassen sich im Hinblick auf den Transfer evidenzbasierten (Steuerungs-)Wissens in die organisationale Praxis zwei Forschungsstränge unterscheiden. Einerseits geht es darum, was Wissenschaftler und Wissenschaftlerinnen – z. B. über angepasste Kommunikationsformen und integrierte Interdisziplinarität – tun können, um evidenzbasiertes Handeln zu fördern (vgl. Bartunek 2007; Cummings 2007; Guest 2007). In seiner „theory of diffusion“ beschreibt Latham (2007), welche Informationen auf welche Weise und durch wen zur Verfügung gestellt bzw. gezielt rückgemeldet werden und welche Effekte hierdurch aufseiten der Empfänger evidenzbasierter Wissensbestände erzeugt werden. Daran anschließend identifizieren Cousins und Leithwood (1993; vgl. Visscher 2001) für die Nutzung extern generierten Wissens wichtige Kategorien der Informationsqualität: die wahrgenommene Qualität bzw. Beschaffenheit von Informationen einschließlich des technischen Anspruchs ihrer Generierung, die Glaubwürdigkeit der Informationsquelle sowie der Personen bzw. Organisationen, welche bestimmte Informationen rückmelden, die Relevanz für die (Schul-)

Praxis und die Kommunikationsqualität von Daten sowie die Zeitlichkeit von Informationen.

Das EVIS-Projekt bewegt sich eher in einem zweiten Strang, der sich in Anlehnung an Latham (2007) als „theory of absorption and infusion“ bezeichnen lässt. Hier geht es insbesondere darum, was Organisationen oder Manager und Managerinnen – hier Schulleitungen und Lehrerkollegien – tun können, um evidenzbasiertes Entscheiden und Handeln in ihrer Organisation zu etablieren und zu fördern. Dabei sind Lehrer und Lehrerinnen die zentralen Akteure in Schulen; die Qualität ihres professionellen Handelns ist entscheidend für den Organisationserfolg von Schulen. Dies deckt sich mit Forschungsergebnissen aus der Organisationsentwicklungs- bzw. Schulentwicklungsforschung, die insgesamt auf die zentrale Bedeutsamkeit des Lehrerhandelns für die Entwicklungen in der Schule verweisen. Diese Befunde gelten insbesondere im Kontext der erweiterten Schulautonomie und der organisationalen Selbststeuerung im Rahmen des neuen outputorientierten Steuerungsmodells (vgl. Zlatkin-Troitschanskaia & Förster 2009).

Im Fokus des Zentralprojekts und der beiden Teilprojekte stehen demzufolge die zentralen Akteure der Schulebene: Lehrerkollegien und Schulleitungen sowie deren Handlungskonstellationen. Strukturen und Prozesse auf dieser Ebene werden auf der Basis mehrperspektivischer theoretischer Analysezugänge modelliert und erfasst. Auf der Grundlage einer sorgfältigen Beschreibung der vorliegenden Praxis in Schulen sollen insbesondere solche personellen und organisationalen Faktoren und Faktorenkonstellationen in Schulen identifiziert werden, die das Wirksamwerden evidenzbasierten Steuerungswissens in der Schulpraxis begünstigen oder behindern (können). Die im Zentralprojekt und den beiden Teilprojekten gemeinsam erarbeiteten Befunde sollen ein wissenschaftlich fundiertes evidenzbasiertes Steuerungswissen bereitstellen, das auch wirksame Strategien des Wissenstransfers in die Schulen aufzeigt und so zur Förderung einer evidenzbasierten administrativen und pädagogischen Schulpraxis beiträgt.

2. Projektdesign

Im Mittelpunkt des Zentralprojekts steht die übergreifende Frage, inwieweit der gesellschaftspolitische Anspruch einer evidenzbasierten Schulentwicklung eingelöst wird. Die Identifikation und Analyse evidenzbasierten Handelns in Schulen sowie der hierfür zentralen Bedingungsfaktoren sind dabei sensibel für die unterschiedlichen Ebenen des Schulsystems. Auf der Ebene der Einzelschule (1) werden u. a. die strukturellen Rahmenbedingungen, Organisationsklima und -kultur auf ihren Einfluss auf den professionellen Umgang mit evidenzbasiertem Wissen untersucht. Der Schulebene sind spezifische Organisationsmerkmale sowie aggregierte individuelle Merkmale der Ebene der Lehrenden (3) zuzurechnen. Die Schulleitungsebene (2) hat eine besondere Funktion, da sie zwar auch auf der individuellen Ebene (3) angesiedelt werden kann, jedoch die organisationale Schulebene (1) in besonderem Maße beeinflusst und teilweise auch repräsentiert. Auf der Ebene der Lehrenden (3) ist u. a. von besonderem Interesse, ob evidenzbasiertes Wissen in die Unterrichts- und Schul-

praxis einfließt. Im Einzelnen werden Bedingungen und Prozesse von Rezeption, Weitergabe und Nutzung von u. a. auf Grundlage von internen Evaluationen, Schulinspektionen und Schulleistungsvergleichsstudien generierten Wissensbeständen untersucht.

Mithilfe einer Fragebogenerhebung wurden im Zentralprojekt Organisations- und Personenmerkmale inkl. Einstellungen und Verhaltensweisen von Lehrerkollegien und Schulleitungen in Rheinland-Pfalz (RLP) landesweit erfasst. Angestrebt wurde eine Vollerhebung aller Lehrkräfte und Schulleitungen an Berufsbildenden Schulen, Förderschulen, Integrierten Gesamtschulen, Gymnasien, Hauptschulen/ Realschulen Plus sowie Grundschulen. In einem ersten Schritt wurden alle Schulen aus dem rheinland-pfälzischen Schulverzeichnis per E-Mail zur Teilnahme eingeladen. In einem zweiten Akquiseschritt folgte eine telefonische Kontaktaufnahme mit jenen Schulen, die sich nach Versendung der E-Mail nicht zurückgemeldet hatten.³ Im Jahr 2011 konnten für die empirische Analyse schließlich Daten von insgesamt 2937 Personen (2640 Lehrkräfte sowie 297 [stellvertretende] Schulleiter) an 153 Schulen gewonnen werden. Da an vielen Schulen eine recht hohe Teilnehmerzahl zu verzeichnen war, konnten auch die Voraussetzungen für die mehrebenenanalytischen Auswertungen im Zentralprojekt erfüllt werden.⁴ Ferner konnten Daten über insgesamt 4025 Netzwerkpartner von 925 Lehrkräften mithilfe eines Netzwerkmoduls im Rahmen dieser Erhebung erfasst werden.

In den beiden Teilprojekten wurde darüber hinaus vertiefend den folgenden Fragen nachgegangen:

Teilprojekt 1 konzentrierte sich auf die Frage, wie sich der professionelle Umgang von (angehenden) Lehrenden mit evidenzbasierten Daten in den verschiedenen Professionalisierungsphasen entwickelt (Studierenden in den konsekutiven Lehramtsstudiengängen, Referendare und Referendarinnen im Vorbereitungsdienst, Berufseinsteiger im Schuldienst sowie erfahrene Lehrkräfte). Damit werden besonders die Entwicklung und Bedeutung individueller Einflussfaktoren wie relevantes Wissen, (motivationale) Einstellungen, Überzeugungen, selbstregulative Fähigkeiten etc. adressiert (s. auch Preußé 2015). In dem längsschnittlich angelegten Teilprojekt 1 erfolgten hierzu Befragungen von Studierenden und Referendaren; die Lehrkräfte im Schuldienst werden im Rahmen des Zentralprojekts erfasst. Der Schwerpunkt der Analysen im Teilprojekt 1 lag auf den kognitiven, motivationalen und selbstregulativen Dispositionen für einen professionellen Umgang mit Evaluationsdaten in Schulen. Ein besonderer Fokus lag auf der Entwicklung dieser Dispositionen im Professionalisierungsprozess im Verlauf der beiden zentralen Phasen der Lehrerbildung (universitäre Ausbildung und das Referendariat). Um diesen Fragestellun-

3 Für eine genauere Beschreibung des Studiendesigns vgl. auch Demski et al. (2012.); van Ackeren et al. (2013).

4 Um eine hohe Rücklaufquote und damit u. a. die hiervon abhängige Möglichkeit der Berechnung von Mehrebenenanalysen sicherstellen zu können, stand die Teilnehmermotivation und -gewinnung stets im Vordergrund forschungspraktischer Überlegungen. Daraus resultierte jedoch eine unbalancierte Stichprobe, was bei der Wahl einer angemessenen Analyseverfahren berücksichtigt werden musste (für den Umgang mit unbalancierten Stichproben bei Mehrebenenanalysen vgl. Hox 2010).

gen nachzugehen, wurden im Frühling 2011 und im Herbst 2013 insgesamt 1630 Personen im Rahmen zweier Erhebungsrounden quantitativ befragt. Zum ersten Erhebungszeitpunkt beteiligten sich 1300 Studierende des gymnasialen und berufsbildenden Lehramts (Bildungswissenschaften und Wirtschaftspädagogik) an der Johannes Gutenberg-Universität Mainz sowie 330 Referendarinnen und Referendare im Vorbereitungsdienst RLP für die Lehramtszugänge Allgemeines Gymnasium und Berufsbildende Schule. An der zweiten Befragung nahmen 1072 Studierende und 100 Referendare teil (zur Zusammensetzung und Beschreibung der Stichprobe s. Preußé 2015). Insgesamt zeichnet sich Teilprojekt 1 durch ein Vergleichs- und Kohortendesign aus. Mittels Vergleichsdesign wurden verschiedene Gruppen angehender Lehrkräfte beispielsweise nach ihrem Studiengang kategorisiert befragt (Wirtschaftspädagogik und Bildungswissenschaften). Das Kohortendesign dient der Untersuchung der Veränderung der individuellen Merkmale. Da im Rahmen des Förderzeitraumes mit zwei Erhebungszeitpunkten nur bedingt ein längsschnittliches Erhebungsdesign realisiert werden konnte, werden in Ergänzung alle Stadien der Lehrerausbildung vom Beginn des Studiums bis zum Ende des Vorbereitungsdienstes abgedeckt.

Teilprojekt 2 ging der Frage nach, welche Rolle organisationale Einflussfaktoren für evidenzbasiertes Handeln im Schulkontext spielen. Hierbei lag der Fokus auf der Erfassung und Analyse innerschulischer Gesamtnetzwerke an ausgewählten Schulen. Mithilfe von standardisierten Fragebögen wurde dabei speziell die Kommunikation über einzelne Evidenzquellen in den Mittelpunkt der Betrachtung gerückt. Darüber hinaus wurde den Anlässen und Motivationen der Nutzung bzw. Nichtnutzung der unterschiedlichen Informationsquellen mithilfe vertiefender leitfadengestützter Interviews nachgegangen. Auf Basis der quantitativen Befunde aus dem Zentralprojekt wurden für Teilprojekt 2 insgesamt acht Schulen mit extremer Ausprägung des evidenzbasierten Handelns ausgewählt. Ausgesucht wurden jeweils zwei Berufsbildende Schulen mit einer geringen bzw. einer ausgeprägten Evidenznutzung sowie jeweils zwei Gymnasien mit einer schwachen bzw. einer intensiven Nutzung evidenzbasierter Informationsquellen. Zwischen Mai und Dezember 2012 wurden an den ausgewählten Schulen 35 qualitative Interviews durchgeführt. Der Interviewleitfaden umfasste Aspekte der Schulkultur sowie der Evidenzbasierung. Es wurden insbesondere Gründe für die Nutzung respektive Nichtnutzung evidenzbasierter Wissensbestände erfragt. Die Interviews wurden vollständig transkribiert und inhaltsanalytisch (in Anlehnung an Mayring 2007) ausgewertet (s. hierzu Demski et al. 2015).

3. Untersuchungskonstrukte und Erhebungsinstrumente

Die Zusammenschau der in der Literatur vorzufindenden Einflussfaktoren auf den individuellen und organisationalen Umgang mit evidenzbasiertem (Steuerungs-) Wissen legt eine Unterscheidung von drei Formen von Einflussfaktoren (-konstellationen) nahe. Neben dem Einfluss persönlicher Einstellungen und von (Kompetenz-) Merkmalen (differenziert nach Schulleitung und Lehrkräften) lassen sich Auswirkungen des Organisationsklimas bzw. der Organisationskultur sowie organisations-

struktureller Bedingungen unterscheiden. Zur Betrachtung der Evidenzorientierung schulischer Akteure und der Evidenzbasierung ihres Handelns sowie zur Erfassung des dargelegten Einflussgefüges sind umfassendere, hinreichend komplexe Modelle nötig, die die Handlungsakteure im schulischen Mehrebenensystem in Interaktion miteinander wie auch in ihrem organisational-strukturellen und prozessualen schulischen Umfeld systematisch und hinreichend differenziert abbilden können (s. ausführlicher van Ackeren et al. 2011).⁵

3.1 Schulebene

a) Evidenzbasiertes Handeln

Als ein Kernelement evidenzbasierten Handelns kann die Bereitschaft der pädagogischen Akteure angesehen werden, evidenzbasierte Informationen und Steuerungsinstrumente aufzugreifen und handlungsleitend wirksam werden zu lassen (Imants et al. 2001). Dies wird als entscheidende Voraussetzung einer evidenzbasierten Selbststeuerungs- und Qualitätsentwicklungspraxis in Schulen verstanden (vgl. Jakku-Sihvonen 2008). Auch im Rahmen organisationswissenschaftlicher Forschung zum EBM („evidence based management“) fanden z. B. Guest und King (2004) Hinweise darauf, dass die Skepsis gegenüber Evidenz ein Hauptgrund dafür war, dass entsprechende Verfahren von Akteuren abgelehnt werden. Dies deckt sich mit den Ergebnissen der Einstellungsforschung (vgl. z. B. Ajzen 2001), die die Bedeutung von Einstellungen der Akteure bei der Erklärung tatsächlichen Verhaltens betonen.

Als abhängige Variablen auf der Organisationsebene werden demnach zwei zentrale Konstrukte fokussiert. Hierbei handelt es sich um die Operationalisierung des „evidenzbasierten Handelns“ mittels der Skalen „interne und externe Evidenzorientierung“. Diese beiden Skalen stellen Eigenentwicklungen dar, die an die internationale Forschung zur Operationalisierung des EBM anschließen und bereits im Rahmen von anderen Projekten validiert werden konnten (Dormann et al. 2010a, 2010b). Zudem wurde eine vertiefende Analyse auf Organisationsebene zur innerschulischen Verarbeitung und Nutzung von Ergebnissen der Schulinspektion in Rheinland-Pfalz vorgenommen.⁶

Um die Konstruktvalidität der Skalen zur Erfassung evidenzbasierten Handelns auf Schulebene zu überprüfen, wurden explorative sowie konfirmatorische Faktorenanalysen getrennt für die Daten der Schulleitungsmitglieder und der Lehrkräfte durchgeführt. Für beide Stichproben zeigten die Analysen die theoretisch angenommene Faktorenstruktur: Der erste Faktor entspricht der internen Evidenzorientie-

5 Im Rahmen des Projektes wurde ein umfassendes theoretisches Rahmenmodell entwickelt, das die Basis für die Herleitung eines Hypothesensystems sowie für die Operationalisierung der abhängigen und unabhängigen Untersuchungskonstrukte bildete und im weiteren Verlauf u. a. mithilfe von Mehrebenenanalysen geprüft worden ist. Zur ausführlichen Darstellung der theoretischen Grundlagen und Annahmen s. van Ackeren et al. (2011).

6 Für Verfahren der externen Evaluation von Schulen werden in den einzelnen Bundesländern unterschiedliche Begrifflichkeiten verwendet. In Rheinland-Pfalz bezeichnet dies die Schulinspektion, die von der Agentur für Qualitätssicherung, Evaluation und Selbstständigkeit von Schulen (AQS) durchgeführt wird.

rung. Items, die hohe Ladungen auf dem Faktor interne Evidenzorientierung haben, sind in Tabelle 1 exemplarisch dargestellt, ebenso für den zweiten Faktor, die externe Evidenzorientierung. Diese Faktorenstruktur repliziert im Wesentlichen auch die Ergebnisse von Stumm, Dormann und Ohlemann (2010). Insgesamt zeigen die Analysen, dass mit der Adaptation des EBM für den Schulkontext nicht nur die internationale Anschlussfähigkeit gewährleistet wird, sondern auch ein Instrument entwickelt wurde, das die Einstellungen von Lehrkräften und Schulleitungen zur Evidenzorientierung valide und reliabel erfasst (s. Dormann et al. 2015).

Tabelle 1: Faktoren der Evidenzorientierung: Bedeutung und Beispielimens im Kontext des EVIS-Projektes⁷			
Faktor	Bedeutung	Beispielimens Lehrkräfte	Beispielimens Schulleitung
Interne Evidenzorientierung	Berücksichtigung von Fakten zur eigenen Schule bei Entscheidungen	Bevor unsere Schule ein Verfahren von einer anderen Schule übernimmt, prüfen wir, inwiefern sich die Rahmenbedingungen in den Schulen ähneln.	Bevor ich ein Verfahren von einer anderen Schule übernehme, prüfe ich, inwiefern sich die Rahmenbedingungen in den Schulen ähneln.
Externe Evidenzorientierung	Berücksichtigung wissenschaftlicher Erkenntnisse bei Entscheidungen	Unsere Schulleitung sucht den direkten Kontakt zu Wissenschaftlern, um ihre Entscheidungsqualität zu verbessern.	Ich suche den direkten Kontakt zu Wissenschaftlern, um meine Entscheidungsqualität zu verbessern.
Quelle: van Ackeren et al. 2013, S. 63.			

b) Organisationale und strukturelle Faktoren

Bei den untersuchten unabhängigen Variablen auf Schulebene sind zunächst grundlegende organisationale und strukturelle Faktoren zu nennen. Einige davon wie etwa der Schultyp, die Größe des Kollegiums, die Sozialstruktur der Schülerschaft, die Betreuungsrelation u. v. m. konnten anhand von Daten des Statistischen Landesamtes Rheinland-Pfalz festgestellt werden. Andere Informationen wie etwa die Siedlungsstruktur im Einzugsgebiet (städtisch vs. kleinstädtisch vs. ländlich), die Altersstruktur des Kollegiums, die Teilnahme an Modellversuchen im Bereich Schul- und Unterrichtsentwicklung etc. wurden bei den Schulleitungen erfragt.

c) Organisationskulturelle bzw. -klimatische Bedingungen

Organisationspsychologische Untersuchungen im schulischen Bereich (vgl. z. B. Buchholzer 2003) zeigen, dass das organisationale Klima für Eigeninitiative und das psychologische Sicherheitsklima als Schlüsselfaktoren für erfolgreiches organisationales Lernen und den Erfolg neu eingeführter Prozesse betrachtet werden können (s. auch Baer & Frese 2003; Edmondson 1999). Mit Organisationsklima wird die auf

organisationaler Ebene aggregierte, subjektive Wahrnehmung von organisationalen Strategien, Praktiken und Abläufen durch die Organisationsmitglieder bezeichnet (Rentsch 1990). Auf der Betrachtungsebene von Schulkultur und Schulklima kann Evidenzorientierung i. S. der Öffnung der Schule für neuere Erkenntnisse und deren konkrete Umsetzung im Unterrichtsalltag als abhängig gesehen werden von der relativen Stärke alternativer Handlungsanforderungen und -bestrebungen der Organisation. Quinn und Rohrbaugh (1983) haben auf der Basis von theoretischen Überlegungen und Expertenbefragungen das competing values model (CVM) entwickelt, mit dem sie die Organisationskultur als ein Konglomerat von Werten beschreiben, d. h. als relative Gewichtung unterschiedlicher Zielkriterien oder Handlungsanforderungen, die von nicht schulischen wie schulischen Organisationen gleichermaßen verfolgt und balanciert werden müssen (Scheerens 2000).⁷

Im Bereich der organisationskulturellen bzw. -klimatischen Bedingungen werden demzufolge drei Konstrukte erfasst: das Klima für Eigeninitiative (Baer & Frese 2003), das psychologische Sicherheitsklima (Edmondson 1999) und die Organisationskultur, die mithilfe einer adaptierten Version des „Organizational Culture Assessment Instrument“ (Cameron & Quinn 2006) erhoben wurde (ausführlicher s. Demski et al. 2015).

Eine herausgehobene Position innerhalb der Schule nimmt die Schulleitung ein. Zahlreiche Befunde aus der Schulentwicklungsforschung belegen, dass die Persönlichkeit und das Handeln der Schulleitung für die Lehrer zentrale Kontextfaktoren darstellen, an denen sie ihr eigenes Handeln ausrichten (vgl. z. B. Bonsen 2010). Demnach werden im Rahmen der Untersuchung allen Lehrkräften auch Fragen zum Führungsstil (transformationale Führung nach Felfe 2006) sowie zur Unterstützung evidenzbasierten Handelns durch die Schulleitung gestellt (ausführlicher s. Stump et al. 2015).

3.2 Individualebene

a) Rezeption und Nutzung evidenzbasierter Quellen

Evidenz kann auf schulischer Ebene von der Schulleitung sowie von den Lehrkräften wahrgenommen, in strategische Planungen umgesetzt und schließlich für tatsächliche Handlungskonsequenzen nutzbar gemacht werden. Diese grundsätzliche Unterscheidung hinsichtlich der Empfänger – Schulleitung auf der einen und Lehrkräfte auf der anderen Seite – hat Visscher 1996 (in Anlehnung an Mintzberg 1989 und McPherson et al. 1986) vorgenommen. Die Schulakteure erhalten viele Informationen aus unterschiedlichen Quellen. Sie restrukturieren diese Informationen und nutzen sie für ihr Entscheidungshandeln. Darüber hinaus eröffnen sich auch formalen Gruppen wie Fachkonferenzen, Klassenteams, innerschulischen Steuergruppen und/oder der Schulkonferenz Nutzungsmöglichkeiten empirischer Evidenzen (vgl. insb. zu schulischen Steuergruppen Kadach & Zlatkin-Troitschanskaia 2009 sowie

⁷ Eine Integration der Ansätze in einem Modell zur Entwicklung der Schule findet sich bei Clausen (2008).

Berkemeyer et al. 2008). Im Kontext der Nutzbarmachung von Evaluationsergebnissen unterscheidet Johnson (1998) hinsichtlich möglicher Nutzungsstrategien nach einem instrumentellen, konzeptuellen, prozesshaften und symbolischen Gebrauch. Mit der instrumentellen Nutzungsform ist gemeint, dass Evaluationsergebnisse in direkter Konsequenz als Ausgangspunkt von Handlungen angesehen werden. In Abgrenzung zum direkten Anwendungsbezug von Evaluationen besteht auch die Möglichkeit, dass Entscheidungsprozesse von Akteuren in einem größeren Zeitrahmen bzw. mit größerem zeitlichen Abstand zum Evaluationsergebnis beeinflusst werden (konzeptuelle Nutzung; Askew & Lodge 2000). Die Evaluation kann ein Baustein neben anderen Einfluss nehmenden Informationen sein, die Einstellungen und Verhaltensweisen von Akteuren längerfristig und akkumulativ formen (Weiss 1984). Bei der symbolischen Nutzung entspringt der individuelle Informationsgebrauch einem politischen bzw. einem gezielten taktischen Interesse.

Zur Erfassung der abhängigen Variablen auf der Individualebene wurden vor diesem Hintergrund mehrere Fragebatterien eingesetzt, die die individuelle Rezeption und Nutzung eines ganzen Katalogs unterschiedlicher Evidenzquellen abbilden (s. van Ackeren et al. 2013). Im Rahmen der Einführung flächendeckender Schulleistungsmessungen entstand eine Vielzahl von Studien, die die Rezeption und Nutzung rückgemeldeter Daten analysieren (vgl. z. B. Schrader & Helmke 2004; Helmke & Hosenfeld 2005; Bach et al. 2014; Pöhlmann et al. 2014). Auch findet sich Literatur zu Wirkungsweisen externer Evaluationen (z. B. Kotthoff & Böttcher 2010; Dederling & Müller 2011; Lambrecht & Rürup 2012; Wurster et al. 2014), von Vergleichsarbeiten (Bach et al. 2014; Richter et al. 2014) oder zu Effekten zentraler Abschlussprüfungen (z. B. Maag Merki 2010; Kühn 2010; Kahnert et al. 2015). Nur wenige dieser existierenden Arbeiten untersuchen jedoch Rezeption und Nutzung verschiedener Evaluationsdaten (wie etwa bei Wurster et al. 2014), und die meisten Studien beschränken sich in der Regel auf einzelne Informationsquellen bzw. einzelne Instrumente evidenzbasierter Steuerung. Im EVIS-Projekt wird einem breiten Verständnis von potenziellen Evidenzquellen gefolgt (ausführlicher s. van Ackeren et al. 2011), um nicht möglicherweise wirksame Wege des Transfers unterschiedlicher Evidenzformen in den schulischen Alltag auszuschließen. Dabei erfolgte eine Systematisierung der aufgenommenen Quellen evidenzbasierter Wissensbestände anhand verschiedener Dimensionen (vgl. Tabelle 2; vgl. zu den folgenden Ausführungen ausführlich van Ackeren et al. 2013; Demski et al. 2012). Zum einen kann eine Unterscheidung entlang der Position bei der Wissensgenerierung erfolgen. Diese kann entweder intern beispielsweise über die Einholung von Schülerfeedbacks oder extern wie im Fall der Schulinspektionen verortet sein. Eine weitere Dimension bildet der Einsatz wissenschaftlicher Methoden und Instrumente bei der Gewinnung der Wissensbestände. Die Anwendung wissenschaftlicher Methoden ist evidenzbasierten Wissensbeständen grundsätzlich immanent. Folgt man jedoch einem weiten Verständnis von Evidenzen, kann eine Differenzierung bezüglich der Güte vorgenommen werden: „Das Spektrum reicht hier von wissenschaftlich-empirischen Studien bis hin zu kollegialen Hospitationen und Schülerfeedbacks, welche die einzelne Schule ggf. weniger standardisiert und nicht explizit unter Beachtung gängiger Gütekriterien der Datenerhebung durchführt“ (van Ackeren et al. 2013, S. 54 f.). Evidenzquellen mit einer aus-

geprägten Wissenschaftsorientierung, was zugleich mit zumeist extern generierten Datensätzen und Wissensbeständen einhergeht (vgl. auch Berkemeyer & Bos 2009), können demgemäß als Evidenzquelle im engeren Sinn klassifiziert werden (vgl. Tabelle 2). Zum anderen kann nach der Aussageebene differenziert werden. Das heißt, ob der Fokus der Evidenzen auf dem Bildungssystem im Allgemeinen, auf einer spezifischen Bildungsorganisation, Gruppe oder auf der Individualebene liegt. Darüber hinaus wird zwischen Evidenzquellen unterschieden, die expliziter mit einem Steuerungsanspruch verknüpft sind, wobei diese einerseits auf Verallgemeinerbarkeit und eher auf die Bestandsaufnahme zur Situation des Schulsystems insgesamt zielen können oder andererseits stärker an der einzelschulischen (Alltags-)Praxis ansetzen. Schließlich werden auch mögliche Evidenzquellen aufgegriffen, die nicht mit einer unmittelbaren Steuerungserwartung verknüpft sind, sondern vielmehr bereits vorhandene Wissensbestände aufgreifen und aufbereitet zur Verfügung stellen. Die Beschaffung und Wahrnehmung solcher vermittelter Evidenzquellen liegt allerdings stärker als bei den anderen genannten Kategorien in der Verantwortung der einzelnen Lehrkräfte.

Tabelle 2: Evidenzbasiertes Wissen: Herkunft des Wissens und zentrale Charakteristika		
	Evidenzquellen im engeren Sinne eher ausgeprägte Wissenschaftsorientierung, Beispiele	Evidenzquellen im weiteren Sinne eher geringere Wissenschaftsorientierung, Beispiele
Eher formalisiert, expliziter Steuerungsanspruch	<ul style="list-style-type: none"> – Schulleistungsstudien – Bildungsberichterstattung – Schulinspektion – Vergleichsarbeiten – Zentrale Abschlussprüfungen 	<ul style="list-style-type: none"> – Schülerfeedback zum Unterricht – Gemeinsame Unterrichtsentwicklungsmaßnahmen (wie kollegiale Hospitationen) – Schulinterne Befragungen – Schulstatistische Auswertungen – Parallelarbeiten
Eher informell, implizites Steuerungspotenzial	<ul style="list-style-type: none"> – Externe Bereitstellung von Aufgabensammlungen – Vermittelte Evidenz über Artikel in schulfachbezogenen oder überfachlichen schulbezogenen Zeitschriften 	<ul style="list-style-type: none"> – Bildungsteile von Tages-/Wochenzeitungen oder Magazinen

Quelle: van Ackeren et al. 2013, S. 55.

b) Sozialstrukturelle Merkmale und Netzwerke

Auf der Individualebene wurden Merkmale wie Alter, Geschlecht, Fachhintergrund, die berufliche Position der Befragten etc. erhoben. Im Hinblick auf die alltägliche Kommunikation über Arbeitsinhalte wurden zudem die egozentrierten Netzwerke der Schulleitungen und Lehrkräfte erfasst. Soziale Netzwerke stellen einerseits wich-

tige Informationskanäle zur Verbreitung professionellen Wissens dar. Andererseits beeinflussen sie die Selektion und Anerkennung evidenten Wissens als solches. Das im Zentralprojekt verwendete Netzwerkmodul bietet dabei eine Möglichkeit, die Unterschiedlichkeit der Gesprächspartner zu bestimmen, indem abgefragt wurde, mit wem wichtige berufliche Angelegenheiten besprochen werden. Dabei sollten die Lehrkräfte angeben, ob ihre Kontaktpartner Lehrer der eigenen Schule, Lehrer einer anderen Schule oder keine Lehrer sind. Da es in der Literatur Hinweise darauf gibt, dass in den sozialen Netzwerken wesentlich über die Akzeptanz und die Handlungsrelevanz evidenzbasierten (Steuerungs-)Wissens verhandelt und entschieden wird (Rogers 2003), wurden die Erkenntnisse aus dem Zentralprojekt in Teilprojekt 2 mit speziellem Fokus auf innerschulische Netzwerke vertiefend untersucht.

c) Individuelle Einstellungen und Kompetenzen

Im Kontext evidenzbasierter Steuerung wird in der Forschung zunehmend die Professionalität des Handelns pädagogischer Akteure, d. h. ihre erworbene Kompetenz zur Mitwirkung an evidenzbasierten Schulentwicklungsprozessen, kritisch hinterfragt (für einen Überblick vgl. Zlatkin-Troitschanskaia et al. 2009). Die wesentliche Bedeutung kognitiver Dispositionen sowie motivationaler Bereitschaften der Schulleitungen ist in der Literatur inzwischen unumstritten. Im Anschluss an die Befunde zu Bestimmungsfaktoren des innovativen Lehrerhandelns in Schulen (vgl. ausführlich Buske 2015) sowie der Lehrerkompetenzforschung nehmen Fragen zu individuellen Einstellungen und Kompetenzen der Lehrkräfte und Schulleitungen einen großen Teil des im Zentralprojekt eingesetzten Erhebungsinstrumentes ein. Zur Erfassung von Selbstwirksamkeitserwartungen der Schulleitungen und Lehrkräfte wurde u. a. auf das etablierte Konzept der *role breadth self-efficacy* von Parker (1998) zurückgegriffen. Kreativität und Innovation, d. h. die Generierung und Umsetzung neuer Ideen im Unterricht bzw. im Steuerungskontext Schule auf Basis evidenzbasierten Wissens, wurden durch eine Skala von Janssen (2000) erfasst. Etablierte Skalen kamen auch bei der Erfassung der allgemeinen Arbeitszufriedenheit (Baillod & Semmer 1994) und des Arbeitsengagements (Schaufeli et al. 2006) zum Einsatz. Diese wurden ergänzt durch Eigenentwicklungen zum wahrgenommenen Gestaltungsspielraum und Einstellungen gegenüber Evidenzen sowie einen Wissenstest zum Bereich Evaluation und Datenanalyse (ausführlicher s. Preuß 2015). Hierbei wurde ein Leistungstest im Multiple-Choice-Format für die Studie neu entwickelt und im Rahmen von Experteninterviews validiert. Für die epistemologischen Überzeugungen wurde eine bereits bewährte und validierte Skala von Urhahne und Hopf (2004) adaptiert⁸ und einem Pretest unterzogen.

⁸ Der naturwissenschaftliche Fokus in den Items wurde durch eine allgemeine Formulierung „Wissenschaft“ ersetzt. Weiterhin wurden einzelne Items ausgeschlossen bspw. wegen konkreten Schülerbezugs oder niedriger Faktorladungen in den Vorgängerstudien (vgl. Preuß 2015).

4. Ausgewählte Ergebnisse

Im Mittelpunkt des Zentralprojekts steht die forschungsleitende Frage, inwieweit der gesellschaftspolitische Anspruch einer evidenzbasierten Schulentwicklung eingelöst wird. Die Analyse des evidenzbasierten Handelns an Schulen und der Bedingungen von Einstellungen zu evidenzbasiertem Handeln bei Lehrkräften und Schulleitungen ist dabei differenziert für die unterschiedlichen Ebenen des Schulsystems vorgenommen worden. Neben der Lehrer- und Schulleiterebene fand dementsprechend auch die organisationale Ebene der Schule getrennt Beachtung. Im Folgenden werden ausgewählte Ergebnisse der Projektarbeit präsentiert.⁹

4.1 Evidenzbasiertes Handeln im Schulkontext

Die befragten Lehrkräfte und Schulleitungen sollten die Aussagen zur Erfassung der Evidenzorientierung ihrer Schule auf einer Skala von 1 (= stimme überhaupt nicht zu) bis 5 (= stimme voll zu) einschätzen (s. Abschnitt 3.1). Die Ergebnisse lassen auf deutliche Unterschiede in der Wahrnehmung der Evidenzorientierung in der jeweiligen Schule zwischen Schulleitung und Lehrkräften schließen. Rein deskriptiv ist zunächst erkennbar, dass die Mitglieder des Lehrerkollegiums die externe Evidenzorientierung an ihren Schulen durchschnittlich als relativ schwach ($M = 2,51$, $SD = 0,80$) und die interne Evidenzorientierung wesentlich stärker ($M = 3,43$, $SD = 0,72$) ausgeprägt sehen. Die Schulleitungsmitglieder schätzen die externe Evidenzorientierung an ihrer Schule dagegen höher ein ($M = 2,81$, $SD = 0,74$) als die Mitglieder des Kollegiums. Die interne Evidenzorientierung an ihrer Schule ($M = 4,15$, $SD = 0,47$) stuften die Schulleitungen ebenfalls höher ein als die externe (ausführlicher s. van Ackeren et al. 2013). Diesem Befund und insbesondere den Ursachen soll empirisch in weiteren Untersuchungen nachgegangen werden. Die unterschiedliche Wahrnehmung bezüglich der Evidenzorientierung der eigenen Schule zwischen den Mitgliedern des Lehrkollegiums und der Schulleitung scheint jedoch zunächst erwartungskonform zu sein. Lehrkräfte haben je nach Größe des Kollegiums teilweise nur einen punktuellen Einblick in bspw. laufende Projekte oder initiierte Maßnahmen an ihrer Schule und nehmen diese dadurch ggf. zeitverzögert bzw. weniger intensiv zur Kenntnis. Schulleitungsmitglieder haben hingegen einen umfassenden bzw. ganzheitlichen Blick auf die Aktivitäten an ihrer Schule.

4.2 Individuelle und organisationale Prädiktoren evidenzbasierten Handelns im Schulkontext

Das Hauptziel des Zentralprojekts war es, sowohl auf der Ebene der Lehrenden, der Schulleitung und der Organisation (= Ebene der Schule) Bedingungen von Einstel-

9 Für einen ausführlichen Überblick über die im Projekt erzielten Ergebnisse vgl. die Beiträge in Zlatkin-Troitschanskaia (2015).

lungen zu evidenzbasiertem Handeln und dem evidenzbasierten Handeln selbst zu identifizieren. Aus diesem Grund wurden Zusammenhänge zwischen individuellen und organisationalen Prädiktoren und der (1) Einstellung der Lehrenden gegenüber evidenzbasiertem Wissen sowie (2) der Evidenzorientierung an den Schulen bestimmt. Die Evidenzorientierung wurde dabei getrennt nach den beiden mittels einer Faktorenanalyse extrahierten Dimensionen externe Evidenzorientierung (EE) und interne Evidenzorientierung (IE) analysiert (vgl. ausführlich Abschnitt 3.1).

Betrachtet man zunächst die **Einstellungen der Lehrenden gegenüber evidenzbasiertem Wissen**, zeigt sich, dass diese positiv mit der spezifischen Selbstwirksamkeit der Lehrenden, auch Aufgaben außerhalb des Kernbereichs erfolgreich bewältigen zu können, zusammenhängen. Auch ein hohes Klima für Eigeninitiative, ein hohes psychologisches Sicherheitsklima sowie ein hohes Ausmaß transformationaler Führung sind positiv mit den Einstellungen der Lehrenden assoziiert. Darüber hinaus hängen die Einstellungen der Schulleitung gegenüber evidenzbasiertem Wissen positiv mit den Einstellungen der Lehrenden zusammen.

In Bezug auf die beiden **Dimensionen der Evidenzorientierung** zeigt sich korrelativ, dass hohe Selbstwirksamkeit, ein hohes Klima für Eigeninitiative, ein hohes psychologisches Sicherheitsklima sowie ein hohes Ausmaß transformationaler Führung positiv mit der externen und internen Evidenzorientierung zusammenhängen. Die Selbstwirksamkeit der Lehrenden ist nur mit der internen Evidenzorientierung assoziiert, die Einstellungen der Schulleitung nur mit der externen Evidenzorientierung.

Zur **Vorhersage der Einstellungen der Lehrenden gegenüber evidenzbasiertem Wissen** wurden Mehrebenenanalysen durchgeführt.¹⁰ Zuerst wurde ein Nullmodell berechnet, in dem noch keine möglichen Prädiktoren, sondern nur die abhängige Variable (Einstellungen der Lehrenden gegenüber evidenzbasiertem Wissen) eingegeben wurde. Dieses Nullmodell erlaubt es, die Varianz der abhängigen Variable in die Anteile der unterschiedlichen Ebenen zu zerlegen. Das Nullmodell zeigt, dass 10% der Varianz der Einstellungen der Lehrenden gegenüber evidenzbasiertem Wissen auf organisationale Variablen zurückgehen ($ICC = 0,103$) und somit die Mehrebenenanalysen angemessen sind (vgl. Snijders & Bosker 1999; Hox 2010).

In einem zweiten Modell wurden zur Vorhersage der Einstellungen der Lehrkräfte gleichzeitig individuelle Merkmale (Selbstwirksamkeit der Lehrenden) sowie Merkmale auf organisationaler Ebene (Klima für Eigeninitiative, psychologisches Sicherheitsklima, transformationale Führung und Einstellungen der Schulleitung) als Prädiktoren untersucht.¹¹ Es zeigt sich, dass auf der individuellen Ebene eine hohe spezifische Selbstwirksamkeit positiv mit den Einstellungen zusammenhängt. Auf der organisationalen Ebene lassen sich ein hohes Klima für Eigeninitiative, der transformationale Führungsstil des Schulleiters und die Einstellungen der Schulleitung als bedeutsame Prädiktoren identifizieren. Die Ergebnisse des Mehrebenenmodells sind in Tabelle 3 dargestellt. Daraus lässt sich schlussfolgern, dass Lehrende, die es sich zutrauen, auch Aufgaben außerhalb ihres Kernbereichs erfolgreich zu bewälti-

10 Die Berechnung der Mehrebenenmodelle erfolgte mittels MPlus.

11 Der Anteil erklärter Varianz auf Organisationsebene liegt bei 57 %, auf Individualebene bei 1 %.

gen, eine positivere Einstellung gegenüber evidenzbasiertem Wissen haben. Darüber hinaus ist die Einstellung an Schulen zu evidenzbasiertem Wissen unter Lehrenden positiver, wenn das Schulklima von hoher Eigeninitiative gekennzeichnet ist. Des Weiteren sind die Einstellungen gegenüber evidenzbasiertem Wissen positiver, wenn auch die Schulleitung eine positive Einstellung gegenüber evidenzbasiertem Wissen hat und das Kollegium transformational führt.

Tabelle 3: Mehrebenenmodell inklusive individueller und organisationaler Prädiktoren zur Vorhersage der Einstellungen der Lehrenden gegenüber evidenzbasiertem Handeln

	Individuelle Ebene (Level 1)				Organisationale Ebene (Level 2)			
	β	SE	R ²	SE	β	SE	R ²	SE
<i>Einstellung der Lehrenden gegenüber evidenzbasiertem Wissen</i>	,077**	,024	,006	,004			,570**	,162
Selbstwirksamkeit					,161	,154		
Psychologisches Sicherheitsklima					,413	,173		
Klima für Eigeninitiative					*	,117		
Transformationale Führung					,239	,099		
Einstellungen der Schulleitung					*			
					,243			
					*			

Anmerkung: Standardisierte Werte sind dargestellt. N = 122 Schulen und 1385 Lehrkräfte.
* p < ,05; ** p < ,01; *** p < ,001. Eigene Berechnungen und Darstellung.

Zur *Vorhersage der Dimensionen der Evidenzorientierung* wurden ebenfalls Mehrebenenanalysen durchgeführt. In Bezug auf die externe Evidenzorientierung zeigt das Nullmodell, dass 10 % der erklärten Varianz der externen Evidenzorientierung auf organisationale Variablen zurückgehen (ICC = 0,102). Bei der internen Evidenzorientierung können ca. 7 % der Varianz auf die Unterschiede zwischen den Schulen zurückgeführt werden (ICC = 0,066) (vgl. auch hier Snijders & Bosker 1999; Hox 2010).

Im weiteren Analyseschritt wurden zur Vorhersage der beiden Dimensionen der Evidenzorientierung gleichzeitig individuelle Merkmale (Selbstwirksamkeit der Lehrenden, Einstellungen der Lehrenden gegenüber evidenzbasiertem Wissen) sowie Merkmale auf organisationaler Ebene (Klima für Eigeninitiative, psychologisches Sicherheitsklima, transformationale Führung und Einstellungen der Schulleitung) in das Modell integriert.¹²

12 Der Anteil der erklärten Varianz auf Organisationsebene liegt mit 64 % bzw. 98 % deutlich höher als auf der Individualebene (jeweils knapp 10 %).

Betrachtet man die Ergebnisse zur externen Evidenzorientierung, so zeigt sich, dass auf individueller Ebene eine hohe spezifische Selbstwirksamkeit sowie eine positive Einstellung gegenüber evidenzbasiertem Wissen positiv auf die externe Evidenzorientierung wirken. Auf organisationaler Ebene gehen signifikante positive Effekte von einem Klima für Eigeninitiative und dem transformationalen Führungsstil der Schulleitung aus. Das Modell weist jedoch keinen Einfluss vom psychologischen Sicherheitsklima und den Einstellungen der Schulleitung auf die externe Evidenzorientierung aus (vgl. zu den Ergebnissen Tabelle 4).

In Bezug auf die interne Evidenzorientierung zeigt sich, dass auf der individuellen Ebene eine hohe spezifische Selbstwirksamkeit sowie positive Einstellungen gegenüber evidenzbasiertem Wissen positiv mit der internen Evidenzorientierung in Zusammenhang stehen. Auf organisationaler Ebene lassen sich ein hohes Klima für Eigeninitiative und der transformationale Führungsstil der Schulleitung als bedeutsame Prädiktoren identifizieren. Das psychologische Sicherheitsklima und die Einstellungen der Schulleitung zeigen keinen bedeutsamen Zusammenhang mit der internen Evidenzorientierung (vgl. Tabelle 4).

Tabelle 4: Mehrebenenmodell inklusive individueller und organisationaler Prädiktoren zur Vorhersage der externen und internen Evidenzorientierung								
	Individuelle Ebene (Level 1)				Organisationale Ebene (Level 2)			
	β	SE	R ²	SE	B	SE	R ²	SE
Externe Evidenzorientierung	,127***	,032	,097***	,016			,635***	,100
Selbstwirksamkeit	,275***	,026						
Einstellungen der Lehrenden					,052	,159		
Psychologisches Sicherheitsklima					,407**	,134		
Klima für Eigeninitiative					,448***	,122		
Transformationale Führung					,068	,120		
Einstellungen der Schulleitung								
Interne Evidenzorientierung	,173***	,030	,096***	,017			,979***	,088
Selbstwirksamkeit	,244***	,027						
Einstellungen der Lehrenden					,113	,139		
Psychologisches Sicherheitsklima					,450***	,122		
Klima für Eigeninitiative					,593***	,100		
Transformationale Führung					,106	,210		
Einstellungen der Schulleitung								

Anmerkung: Standardisierte Werte sind dargestellt. N = 122 Schulen und 1385 Lehrkräfte.
* p < ,05; ** p < ,01; *** p < ,001. Eigene Berechnungen und Darstellung.

Die Ergebnisse lassen die Schlussfolgerung zu, dass vor allem Lehrende, die es sich zutrauen, auch Aufgaben außerhalb ihres Kernbereichs erfolgreich zu bewältigen und eine positive Einstellung gegenüber evidenzbasiertem Wissen haben, über hohe interne und externe Evidenzorientierung verfügen. Darüber hinaus herrscht vor allem an jenen Schulen eine hohe externe und interne Evidenzorientierung, an denen ein hohes Klima für Eigeninitiative herrscht und die Schulleitung einen transformationalen Führungsstil hat.

4.3 Rezeption und Nutzung verschiedener Quellen für evidenzbasiertes Wissen

Weitere Unterschiede zwischen der Schulleitung und dem Lehrerkollegium zeigen sich in der Rezeption und Nutzung der einzelnen Informationsquellen und Instrumente zur Ausgestaltung der schulischen Arbeit. Anhand fünfstufiger Likert-Skalen wurden potenzielle Differenzen der Intensität der Auseinandersetzung (von 1 = gar nicht bis 5 = sehr intensiv) sowie der tatsächlichen Nutzung (von 1 = gar nicht bis 5 = in sehr hohem Maße) bestimmter Evidenzquellen zwischen Schulleitungen und Lehrkräften abgefragt (vgl. Demski et al. 2012; van Ackeren et al. 2013). Deskriptive Ergebnisse zeigen, dass sowohl für die Schulleitungen als auch für die Lehrkräfte interne Evidenzquellen wie das Schülerfeedback (Schulleitung: $M = 3,39$, $SD = 1,12$; Lehrkräfte: $M = 3,55$, $SD = 1,15$) oder kollegiale Unterrichtsentwicklungsmaßnahmen (Schulleitung: $M = 3,43$, $SD = 1,17$; Lehrkräfte: $M = 3,08$, $SD = 1,36$) die am häufigsten genutzten Instrumente sind. Deutliche Differenzen zwischen Schulleitungen und Kollegium zeigen sich insbesondere bezüglich der Nutzung externer Evidenzquellen, wie etwa der Ergebnismrückmeldung der Schulinspektion. Zwar ist die Auseinandersetzung der Lehrkräfte mit dieser Quelle als noch recht intensiv zu bewerten ($M = 2,87$, $SD = 1,36$), die durchschnittlich angegebene tatsächliche Nutzung fällt jedoch deutlich niedriger aus als bei den Schulleitungsmitgliedern (Schulleitung: $M = 2,90$, $SD = 1,32$; Lehrkräfte: $M = 2,00$, $SD = 1,11$). Auch insgesamt bewerteten die Schulleitungen die unterschiedlichen Evidenzquellen positiver und gaben im Durchschnitt auch eine höhere tatsächliche Nutzung an. Die Befunde lassen vermuten, dass vor allem die Lehrkräfte zentrale Elemente der evidenzbasierten Steuerung – wie Ergebnismrückmeldungen der Schulinspektion – in nur geringem Ausmaß für die eigene Arbeit nutzen (vgl. Tabelle 5; vgl. zu den Ergebnissen auch Demski 2014; van Ackeren et al. 2013).

Tabelle 5: Durchschnittliche Nutzung der abgefragten Evidenzquellen

Informationsquelle	Lehrkräfte		Schulleitung	
	M	SD	M	SD
Schülerfeedback	3,55	1,15	3,39	1,12
Kollegiale Unterrichtsentwicklungsmaßnahmen	3,08	1,36	3,43	1,17
Schulfachbezogene Zeitschriften	3,04	1,22	3,00	1,09
Parallelarbeiten	2,77	1,36	3,07	1,28
Eingesetzte Tests	2,70	1,31	3,27	1,26
Bildungsteil von Zeitungen und Magazinen	2,67	1,25	2,66	1,09
Überfachliche schulbezogene Zeitschriften	2,42	1,16	2,96	1,02
Schulstatistiken	2,22	1,14	2,83	1,22
Aufgabensammlungen von zentraler Stelle	2,10	1,18	2,35	1,19
Befunde der Schulinspektion	2,00	1,11	2,90	1,32
Landesweite Lernstanderhebungen	1,98	1,04	2,47	1,10
Schulleistungsvergleiche	1,91	1,00	2,38	1,05

Anmerkung: Fragetext: „In welchem Maße haben sie die Informationen aus den folgenden Quellen tatsächlich zur Ausgestaltung und Weiterentwicklung Ihrer eigenen Arbeit nutzen können?“. Ser-Skala von 1 „gar nicht“ bis 5 „in sehr hohem Maße“. N = 1230 Lehrkräfte und 297 Schulleitungen. Eigene Berechnungen und Darstellung.

Die Anlässe und Motivationen der Nutzung respektive Nichtnutzung der unterschiedlichen Informationsquellen waren Gegenstand vertiefender leitfadengestützter Interviews im Rahmen des Teilprojekts 2 (s. hierzu auch Abschnitt 2). Die Analysen zeigen, dass insbesondere mangelnde zeitliche Ressourcen von den befragten Lehrkräften und Schulleitungsmitgliedern als hinderlicher Faktor für eine intensive Nutzung zur Verfügung stehender Informationsquellen angeführt werden. Mit Blick auf den Ergebnisbericht der Schulinspektion schätzen die Befragten zudem teilweise die Passung zwischen diesem standardisierten Instrument und den Bedürfnissen der Einzelschule als problematisch ein. Begünstigt wird die Evidenznutzung durch den kollegialen Austausch in Schulentwicklungs-/Steuergruppen.

4.4 Schulform- sowie fachdisziplinspezifische Differenzen des evidenzbasierten Handelns

Des Weiteren wurden Zusammenhänge zwischen evidenzbasiertem Handeln und Schulmerkmalen untersucht. Hier zeigen sich zwar signifikante, jedoch relativ schwa-

che Differenzen hinsichtlich der Schulart/-form (wie z. B. zwischen berufsbildenden und allgemeinbildenden Schulen) sowohl bezüglich des evidenzbasierten Handelns selbst als auch der Art und Stärke seiner Prädiktoren (s. hierzu die Ergebnisse in Zlatkin-Troitschanskaia et al. 2013; Stump et al. 2013). Die Befunde aus Teilprojekt 1 unterstützen die schulformspezifischen Befunde auch für die Gruppe der angehenden Lehrkräfte. Bezogen auf die Ergebnisse aus dem Fachwissenstest zeigt sich bspw., dass der Einfluss des Lehramtszugangs im allgemeinbildenden oder berufsbildenden Bereich ca. 3 % der Varianzaufklärung des Wissensscores ausmacht. Dem Einfluss der Fachdisziplinen auf die Ausprägungen der Kompetenzdispositionen für den professionellen Umgang mit evidenzbasierten Erkenntnissen im Bereich der Schulentwicklung wurde in Teilprojekt 1 differenzierter nachgegangen.¹³ Dabei wurden die Studienteilnehmer in drei Fachdomänen unterteilt: Wirtschaft, Naturwissenschaften (NaWi) und Sozialwissenschaften (SoWi). Im Vergleichsgruppendesign wurden insgesamt fünf verschiedene Gruppen angehender Lehrkräfte untersucht: Wirtschaft (1), Erst- und Zweitfach NaWi (2), Erst- und Zweitfach SoWi (3), Erstfach NaWi und Zweitfach SoWi (4) sowie Erstfach SoWi und Zweitfach NaWi (5) (vgl. Preuße 2015).¹⁴ Um einen Einfluss vor Beginn der Hochschulsozialisation auszuschließen, wurden die domänenspezifischen Gruppen der Studienanfänger kovarianzanalytisch verglichen.¹⁵ Bei den Studienanfängern konnten keine signifikanten disziplinabhängigen Unterschiede festgestellt werden. Die Befunde zu den fortgeschrittenen Studierenden zeigen, dass tatsächlich beim Fachwissen zur Evaluationsmethodik signifikante Unterschiede nach der Disziplinzugehörigkeit vorliegen. Angehende Lehrkräfte des Faches Wirtschaft weisen bspw. das höchste Evaluationswissen¹⁶ ($M = 2,27, SE = 0,05$) auf gefolgt von den Studierenden mit dem Erstfach Naturwissenschaften ([2]: $M = 2,05, SE = 0,07$ bzw. [3]: $M = 1,95, SE = 0,06$) und den Studierenden mit dem Erstfach SoWi ([4]: $M = 1,88, SE = 0,03$ bzw. [5]: $M = 1,87, SE = 0,08$) (für detaillierte Ergebnisdarstellung vgl. Preuße 2015).

13 Vergleichende Analysen erfolgen hier insb. für die kognitiven Konstrukte „Wissen zum Evaluationsprozess“ und „Wissen zur Evaluationsmethodik“ sowie die affektiv-motivationalen Konstrukte „epistemologischen Überzeugungen“ und die „Selbstwirksamkeitsüberzeugungen“. Eine empirische Prüfung der theoretischen Dimensionsstruktur mittels konfirmatorischer Faktorenanalysen zeigt gute Passung des mehrdimensionalen Modells mit den Daten (REMSA: 0,041; CFI: 0,965; TLI: 0,961) (vgl. detaillierte Ergebnisdarstellung in Preuße 2015).

14 Die hier dargestellten qualitativen Ergebnisse aus Teilprojekt 1 beziehen sich exemplarisch auf die erste querschnittliche Befragung von 1600 angehenden Lehrkräften (1272 Studierende und 328 Referendare). 30 Befragungspersonen wurden als aktive Testverweigerer identifiziert und aus den Analysen ausgeschlossen (vgl. Abschnitt 2; Preuße 2015).

15 In den Analysen wurde die Abiturnote als Kontrollvariable berücksichtigt, sodass die angegebenen Mittelwerte adjustiert sind. Für weitere differenzierte Analysen zum Einfluss der Fachwissenschaft vgl. Preuße (2015).

16 Die maximal erreichbare Punktzahl für den Wissensscore liegt bei fünf Punkten für das „Wissen zum Evaluationsprozess“ bzw. bei drei Punkten für das „Wissen zur Evaluationsmethodik“. Beide Wissensdimensionen wurden mittels standardisierter Multiple-Choice-Items erfasst und mithilfe von Experteninterviews validiert (vgl. Preuße 2015).

4.5 Entwicklung des Evaluationswissens während der Lehrerausbildung

Die Anforderung an die Lehrkräfte, evidenzbasiert zu Handeln i. S. eines professionellen Umgangs mit schulbezogenen Evidenzen, birgt die Gefahr der Belastung und Überforderung. Zur Bewältigung sind bestimmte Ressourcen erforderlich, die u. a. in der Person selbst verankert sind. Unter den Kompetenzmerkmalen gilt das Professionswissen als wichtige persönliche Ressource für die Bewältigung beruflicher Herausforderungen (Klusmann 2011). Die Lehrerausbildung hat u. a. die Aufgabe, dieses Wissen herauszubilden. Die Ergebnisse des Teilprojekts 1 zeigen, dass dies für den Bereich der Schulentwicklung und Evaluation eher bedingt erfolgt. Kovarianzanalytisch zeigt sich bspw. kein signifikanter Effekt für die Phase der Professionalisierung als Einflussvariable auf die Höhe des Fachwissens zur Evaluationsmethodik. Auch für das Fachwissen zum Evaluationsprozess konnte nur ein marginal signifikanter Interaktionseffekt zwischen den Einflussvariablen Fachdisziplin und der Phase der Professionalisierung festgestellt werden. Im Kohortendesign wurden drei verschiedene Gruppen angehender Lehrkräfte untersucht: Studierende am Anfang der universitären Ausbildung (Kohorte A), fortgeschrittene Studierende am Ende der universitären Ausbildung (Kohorte B) sowie Referendare (Kohorte C) (vgl. Preuß 2015). Bemerkenswert ist der Befund, dass der im Studium (Vergleich zwischen Kohorte A und B) positive Entwicklungsverlauf in einigen Kompetenzdimensionen während der Ausbildung im Vorbereitungsdienst (Vergleich zwischen Kohorte B und C) stagniert und sogar rückläufige Tendenzen zeigt. Während für das Wissen zum Evaluationsprozess bei angehenden Lehrkräften mit Erstfach Natur- und dem Zweitfach Sozialwissenschaften die Gruppe fortgeschrittener Studierender ($M = 2,61$, $SE = 0,11$) im Vergleich zur Gruppe der Studienanfänger ($M = 1,96$, $SE = 0,12$) durchschnittlich höhere Testwerte erreicht, lässt sich ein Rückgang in den Testwerten für die Gruppe der Referendare ($M = 2,55$, $SE = 0,19$) gegenüber der Kohorte B feststellen (s. Preuß 2015). Auch für die Ausprägungen der Selbstwirksamkeitsüberzeugungen angehender Lehrkräfte wurde im Vorbereitungsdienst ein rückläufiger Verlauf deutlich. Hier zeigt sich ein signifikanter Interaktionseffekt für die Einflussfaktoren Phase der Professionalisierung und Fachdisziplin. Für die angehenden Lehrkräfte bspw. mit Erst- und Zweitfach Naturwissenschaften lässt sich zwischen den Kohorten A und B ($M_A = 3,73$, $SE_A = 0,11$; $M_B = 3,79$, $SE_B = 0,10$) ein leichter Anstieg in den Ausprägungen verzeichnen. Im Referendariat dagegen ($M_C = 3,28$, $SE_C = 0,10$) ist ein starker Rückgang in den Ausprägungen sogar unter das Ausgangsniveau feststellbar (s. Preuß 2015). Diese Befunde aus dem querschnittlich angelegten Untersuchungsdesign könnten als erste Hinweise auf einen Praxisschock während der praktischen Phase der Ausbildung interpretiert werden. In einer weiterführenden Längsschnittstudie soll diesem Effekt und seinen möglichen Erklärungen weiter nachgegangen werden und eine tiefer gehende Analyse erfolgen (vgl. Kapitel 5).

Die qualitativen Analysen ergänzen die quantitativen hinsichtlich einer tieferen Betrachtung der Wahrnehmung von Schulentwicklung und Evaluation als einer Kerndimension sowie die Integration dieser in den professionellen Aufgabenbereich

einer Lehrkraft (vgl. Magerl 2012).¹⁷ Dabei ist zu konstatieren, dass nur sehr wenige Studierende Schulentwicklung und Evaluation als zentrale Kerndimension der Lehrertätigkeit betrachten. Bei den Überzeugungen bezüglich der Aufgaben einer Lehrkraft werden Aufgaben aus dem Bereich Innovieren von den Studierenden nur sehr selten genannt. Studienanfänger erwähnen bspw. Aufgaben aus den Bereichen Erziehen und Unterrichten (durchschnittlich 1,56 bzw. 0,63 Nennungen) deutlich häufiger als jene, die dem Bereich Innovieren (durchschnittlich 0,08 Angaben) zugeordnet werden können. Erst Referendare führen deutlich mehr zum Aufgabenbereich Innovieren an (durchschnittlich 0,75 Nennungen). Bezeichnend ist neben der Veränderung des Stellenwertes auch die differenziertere Wahrnehmung der Themenbereiche der Schulentwicklung, die sich bei Referendaren deutlich zeigt. Während Studienanfänger und Studierende bei der Frage nach den Themenbereichen der Schulentwicklung durchschnittlich 1,69 bzw. 1,94 Angaben machen, liegt die durchschnittliche Anzahl der Antworten bei den Referendaren bei 2,77 Nennungen. Demnach kann der Wechsel von Studium in den Vorbereitungsdienst einhergehend mit der Konfrontation mit der schulischen Praxis als Anstoß für Veränderungen im Überzeugungssystem gesehen werden. Diese Veränderungen zeigen sich sowohl in einer Zunahme der Differenziertheit des Überzeugungssystems als auch in der zunehmenden Wahrnehmung der Bedeutung einzelner Kerndimensionen für die Praxisrelevanz.

4.6 Einfluss der sozialen Netzwerke auf evidenzbasiertes Handeln

Im Rahmen des Zentralprojekts konnten insgesamt 4025 Netzwerkpartner von 925 Lehrkräften mithilfe eines Netzwerkmoduls erfasst werden.¹⁸ Die Befragten wurden dabei zunächst gebeten, bis zu fünf Personen anzugeben, mit denen sie über wichtige berufliche Angelegenheiten sprechen. Zusätzlich wurden die Befragten gebeten, nähere Angaben zu den Netzwerkpartnern zu machen, wie z. B. den Grad der Bekanntschaft zwischen den genannten Personen. Die Angaben der befragten Lehrkräfte entsprechen einer durchschnittlichen Netzwerkgröße von 4,4 Personen. Das Interesse an der Netzwerkgröße lässt sich dadurch begründen, dass mit jedem weiteren Netzwerkpartner potenziell neue Informationen an die Lehrkräfte weitergeleitet werden können.¹⁹ Die Daten der egozentrierten Netzwerke zeigen jedoch, dass 73 % der angegebenen Netzwerkkontakte aus dem direkten Kollegenkreis der Befragten stammen. 13 % entfallen auf Lehrkräfte anderer Schulen, und die verbleibenden 14 % sind keine Lehrer. Die Daten des Zentralprojekts zeigen weiterhin, dass

17 Insgesamt wurden die angehenden Lehrkräfte angehalten, jeweils mindestens drei Aufgabenbereiche einer Lehrkraft sowie mindestens drei Themenbereiche der Schulentwicklung zu benennen.

18 Im Rahmen der hier berichteten Analysen wurden ausschließlich Fälle ausgewertet, für die vollständige Informationen zu den genannten Netzwerkpartnern vorliegen.

19 Einschränkung ist dabei anzumerken, dass aufgrund der Beschränkung auf maximal fünf Netzwerkpartner und der damit einhergehenden geringen Varianz der Netzwerkgröße keine direkten Zusammenhangsanalysen angebracht sind.

die Kommunikation mit den Lehrkräften der eigenen Schule am häufigsten stattfindet, während mit Lehrkräften anderer Schulen seltener über berufliche Angelegenheiten gesprochen wird. Dennoch lässt sich argumentieren, dass der Kontakt mit Lehrkräften anderer Schulen Vorteile bietet, da er den schulischen Akteuren Zugang zu neuen Informationen ermöglicht. Die betreffenden Netzwerkpartner aus den anderen Schulen verfügen über jeweils eigene soziale Netzwerke an ihren Schulen, in denen womöglich unterschiedliche Informationen (unter anderem auch Evidenzen) zirkulieren (Granovetter 1973).

Die Daten des Zentralprojekts unterstreichen auch die Bedeutung der im Teilprojekt 2 gewählten Herangehensweise zur Analyse schulischer Gesamtnetzwerke. Schließlich konnte gezeigt werden, dass berufliche Gespräche überwiegend innerhalb der eigenen schulischen Grenzen geführt werden, weshalb eine vertiefende Analyse der Beziehungs- und Kommunikationsstrukturen lohnenswert erscheint. Die Erhebung innerschulischer Gesamtnetzwerke bietet zudem gewisse methodische Vorzüge. Zum einen lässt sich innerhalb der Gesamtnetwerkerhebung die in der egozentrierten Netwerkerhebung des Zentralprojekts künstlich gesetzte Begrenzung von lediglich fünf Nennungen auflösen. Zum anderen können mithilfe der standardisierten Fragebögen spezielle Kommunikationsnetzwerke über einzelne Evidenzquellen erfasst werden. Im Rahmen des Teilprojekts 2 wurden alle Mitglieder der Lehrerkollegien von insgesamt fünf Schulen gebeten anzugeben, in welchem Maße sie verschiedene Evidenzquellen nutzen und mit wem sie sich über diese Quellen austauschen.²⁰ Zusätzlich wurden weitere Netzwerktypen erhoben, die die affektiven (z. B. persönliche Angelegenheiten besprechen) und arbeitsbezogenen Beziehungen (z. B. Ratschläge geben) erfassen. Die theoretischen Vorarbeiten deuteten darauf hin, dass Entscheidungen über die Relevanz und Akzeptanz einzelner Evidenzquellen nicht allein individuell, sondern im Austausch mit anderen als kollektiver Prozess getroffen werden (Rogers 2003).

Die Auswertung der Netzwerkdaten zeigt, dass affektive Netzwerkbeziehungen im Vergleich zu arbeitsbezogenen Beziehungen etwas höhere Dichtewerte aufzeigen. Die Dichte eines Netzwerks beschreibt den Grad der Vernetzung innerhalb eines bestimmten Netzwerks. Die niedrigsten Dichtewerte lassen sich bei den Kommunikationsnetzwerken über einzelne Evidenzquellen finden. Weitere Analysen des Datenmaterials geben einen Hinweis darauf, dass die Nutzung von Evidenzquellen für den eigenen Unterricht mit der Anzahl der angegebenen Gesprächspartner steigt, mit denen man über die betreffende Quelle spricht (ausführlicher s. Laier et al. 2015). Daraus lässt sich folgern, dass der Transfer von Evidenzen in die schulische Praxis befördert werden kann, indem innerhalb der Schulen Ressourcen, primär Zeit, für die Diskussion der verschiedenen Evidenzquellen bereitgestellt werden.

Die Befunde der qualitativen Interviews, die im Rahmen des Teilprojekts 2 zusätzlich durchgeführt wurden, untermauerten die Ergebnisse des Zentralprojekts und die der vertiefenden Netwerkerhebung. Die auf Basis der Fragebogenerhebung

²⁰ Von den acht Schulen, die für das Teilprojekt 2 ausgewählt wurden, haben sich fünf bereit erklärt, an der vertiefenden Netwerkerhebung teilzunehmen. Drei dieser Schulen wurden aufgrund einer hohen und zwei wegen einer niedrigen Evidenzorientierung ausgewählt, wobei die Schulen mit einer hohen Evidenzorientierung auch die höchsten Rücklaufquoten generierten.

als evidenzbasiert klassifizierten Schulen zeigten darüber hinaus ein hohes Maß an Aufgeschlossenheit für die vertiefenden Erhebungen. Die Kultur der ausgewählten evidenzorientierten Schulen zeichnete sich durch ein vergleichsweise hohes Maß an Kooperation und Flexibilität aus, während die aufgrund ihrer geringen Evidenzbasierung ausgewählten Schulen den Fokus eher auf Stabilität und Wettbewerbsorientierung gerichtet hatten. Die vertiefenden Interviews verdeutlichten des Weiteren, dass für die befragten Schulleitungsmitglieder und Lehrkräfte Zeitmangel der größte Hinderungsgrund für die Nutzung evidenzbasierter Wissensbestände darstellt. Zudem ist vielen Akteuren der spezifische Nutzen der Informationsquellen für die eigene schulische Arbeit nicht ersichtlich.

5. Fazit und Ausblick

In den vergangenen Jahren wurden zahlreiche Maßnahmen initiiert, um auf Grundlage neu eingeführter Steuerungsinstrumente (wie die Teilnahme an internationalen Vergleichsstudien, die Einführung von Bildungsstandards und darauf bezogene Lernstandserhebungen oder die Verfahren der Schulinspektion) eine verbesserte Schulsystemsteuerung zu fördern. Bisher mangelt es jedoch an empirisch fundierten Erkenntnissen darüber, welche (förderlichen) Wirkungen von den eingeführten Innovationen tatsächlich in die Schulpraxis ausgehen und ob diese in Bezug auf die bildungspolitisch anvisierten Ziele als erfolgreich gelten können.

Eine evidenzbasierte Bildungspolitik wie auch evidenzbasiertes Handeln in Schulen können als zentrales bildungspolitisches Anliegen auf allen Steuerungsebenen betrachtet werden. Aus den bildungspolitischen Reformen der vergangenen Jahre ergeben sich für die Schulen erhöhte Anforderungen an eine effiziente und effektive evidenzbasierte Selbststeuerung. Demnach müssen sich Schulen zu einem wesentlichen Anteil selbst steuern können, um in ihrem spezifischen Kontext und den sich wandelnden Rahmenbedingungen erfolgreich zu sein. Erfolg sollte sich insbesondere dann einstellen, wenn Steuerungsmechanismen und -entscheidungen evidenzbasiert sind. Damit geht die Forderung einer hohen Evidenzorientierung in Schulen einher. Gleichzeitig ist ein erhöhter Bedarf an Reflexion auch aufseiten der Bildungspolitik zu beobachten, wo man insbesondere an Untersuchungen zur Wirkung der eingeleiteten Reformen interessiert ist.

Das EVIS-Projekt knüpft hieran an, indem es einen Beitrag zur empirischen Bewertung des Umgangs mit evidenzbasiertem (Steuerungs-)Wissen in der Schulpraxis leistet und dezidiert die Bedingungen evidenzbasierten Handelns im schulischen Mehrebenensystem untersucht. Unter schulischer Evidenzorientierung werden dabei Vorgehensweisen und Praktiken in Schulen verstanden, die der Bereitstellung bzw. Beschaffung und Umsetzung der verfügbaren Wissensbestände für schulisches Entscheidungshandeln dienen. Schulisches evidenzbasiertes Handeln ist demnach dann gegeben, wenn die schulische Evidenzorientierung der Organisation Schule insgesamt hoch ausgeprägt ist (van Ackeren et al. 2013).

Im Rahmen des EVIS-Projektes wurde zunächst untersucht, inwiefern die Schulen diesem Anspruch nach eigener Selbsteinschätzung gerecht werden. Dabei wurde

die Ebene der Schule als Organisation fokussiert. Dem folgend wurden sowohl die Schulleitungen als auch die Lehrerkollegien um ihre Einschätzung gebeten, ob und inwiefern ihre Schule als Organisation u. a. ein evidenzorientiertes Klima aufweist und inwieweit die verschiedenen Evidenzquellen rezipiert und zur tatsächlichen Unterrichtsgestaltung genutzt werden. Im Sinne der förderpolitischen Zielsetzung wurde des Weiteren der Frage nachgegangen, wie sich ein solches evidenzbasiertes Handeln ausprägt und vor allem durch welche individuellen Voraussetzungen und strukturellen bzw. organisationalen Faktoren eine Beeinflussung bzw. Förderung und gezielte Unterstützung erfolgen kann. Damit schließt das Projekt direkt an den aktuellen Stand der nationalen und internationalen Forschung an (s. o.) und erweitert diesen um verschiedene methodische und inhaltliche Aspekte. Zum einen ermöglicht die Befragung von Schulleitungen und Lehrkräften einen direkten und kontrollierten Vergleich der Einstellungen und Einschätzungen zwischen diesen beiden zentralen Akteursgruppen. Zum anderen erlauben die gewonnenen Daten – wie bereits ausgeführt – die Berechnung von Mehrebenenanalysen, sodass die hierarchische Struktur der Organisation Schule angemessen berücksichtigt wird. Schließlich wird im Projekt einem breiten Verständnis evidenzbasierten Handelns gefolgt, wobei die international etablierten Konstrukte einer internen und externen Evidenzorientierung auf den deutschen Schul- bzw. Steuerungskontext adaptiert wurden sowie zudem der Vergleich zwischen Rezeption und Nutzung ganz unterschiedlicher Informationsquellen bzw. Instrumente evidenzbasierter Steuerung ermöglicht wurde. Auf diese Weise wurden im Projekt alle wesentlichen und theoretisch wirksamen Wege des Transfers unterschiedlicher Evidenzformen in die schulische Praxis berücksichtigt.

Die EVIS-Ergebnisse zeigen in der Gesamtschau, dass der bildungspolitisch angestrebte Wissenstransfer in die Schulpraxis oft nicht hinreichend erfolgt, was durch eine Vielzahl organisationaler und personeller Bestimmungsfaktoren erklärt werden kann.²¹ Indem unterschiedliche Formen der Evidenzorientierung und Nutzung verschiedener Evidenzquellen parallel für die beiden Akteursgruppen Schulleitungen und Lehrkräfte vergleichend betrachtet wurden, konnte eine differenzierte, an den internationalen Forschungsstand gut anschlussfähige Befundlage ermittelt werden. So zeigt sich, dass die interne Evidenzorientierung im Vergleich zur externen Evidenzorientierung sowohl bei Schulleitungen als auch bei Lehrkräften deutlich ausgeprägter ist. Das heißt, im Durchschnitt berücksichtigen Schulleitungen und Lehrkräfte bei Entscheidungen zwar häufig Wissensbestände zur eigenen Schule und Unterricht, greifen jedoch deutlich seltener auf allgemeine schulübergreifende wissenschaftliche Erkenntnisse zurück. Der Vergleich zwischen den beiden Akteursgruppen verdeutlicht zudem – theoretisch konforme – Unterschiede im durchschnittlichen Niveau der Evidenzorientierung. So weisen Schulleitungen eine ausgeprägtere interne sowie externe Evidenzorientierung auf. Dies gilt auch für die Befunde zur Rezeption und Nutzung einzelner Evidenzquellen. Insgesamt greifen

21 Kritisch anzumerken ist jedoch, dass die Datengewinnung über Selbstauskünfte der Lehrerschaft und Schulleitungen erfolgte. So kann angenommen werden, dass aufgrund von Effekten sozialer Erwünschtheit sowie einer potenziellen Selbstrekrutierung evidenzbasiert handelnder Schulen ein tendenziell positives Bild im Rahmen des Projektes gezeichnet wird.

sowohl Schulleitungen als auch Lehrkräfte zur Ausgestaltung und Weiterentwicklung ihrer eigenen Arbeit am häufigsten auf interne Evidenzquellen wie das Schülerfeedback oder kollegiale Unterrichtsentwicklungsmaßnahmen zurück. Hingegen werden externe Evidenzquellen wie bspw. die Ergebnisberichte der Schulinspektion zwar stark rezipiert, jedoch vergleichsweise selten in der Schul- und Unterrichtspraxis genutzt. Der Vergleich zwischen Schulleitungen und Lehrkräften zeigt jedoch auch hier deutliche Unterschiede. So ist über alle Evidenzquellen hinweg eine im Durchschnitt ausgeprägtere Nutzung durch Schulleitungen zu konstatieren.

Hinsichtlich der Erklärung der Einstellungen der Lehrkräfte gegenüber evidenzbasiertem Handeln sowie der internen und externen Evidenzorientierung konnte – unter Rückgriff auf den Forschungsstand (s. o.) – in der ersten Förderphase des EVIS-Projektes (EVIS I) eine Reihe von individuellen Merkmalen (wie z. B. Wissen, Selbstwirksamkeitsüberzeugungen) und organisationalen Faktoren (wie z. B. Führungsstil der Schulleitung, Klima für Eigeninitiative) berücksichtigt werden. Als zentraler Einflussfaktor auf der Individualebene zeigt sich die Selbstwirksamkeit der Lehrkräfte. Auf der Organisationsebene konnten der Führungsstil der Schulleitung sowie das Klima für Eigeninitiative als bedeutsame Prädiktoren identifiziert werden. So wirkt ein transformationaler Führungsstil der Schulleitung sowohl positiv auf die Einstellung der Lehrkräfte gegenüber evidenzbasiertem Handeln als auch auf deren interne sowie externe Evidenzorientierung. Die Einstellung der Schulleitungen gegenüber evidenzbasiertem Handeln hat zudem einen positiven Einfluss auf die Einstellung der Lehrkräfte. Diese Befunde sind konform mit zahlreichen Studien, in denen das Schulleiterhandeln als zentraler Einflussfaktor auf das Lehrerhandeln identifiziert werden konnte (vgl. z. B. Wayman et al. 2012; Stump et al. 2015).

Diese Erkenntnisse werden im Rahmen der sich anschließenden zweiten Förderphase (EVIS II) vertieft und erweitert. So weisen die Ergebnisse aus EVIS I auf eine besondere Stellung und die großen Einflussmöglichkeiten der Schulleitung hin. Daher soll in EVIS II Aufschluss über die Kausalität erlangt werden, indem die bisherigen Ergebnisse aus der querschnittlichen EVIS-I-Studie in einem Längsschnitt überprüft werden. Dies ist vor allem auch in Bezug auf die praktischen Implikationen relevant, da nur hierdurch die zentrale Frage geklärt werden kann, ob Interventionen primär auf der Ebene der Schulleitung angesiedelt werden sollten (z. B. wenn sich zeigt, dass das Handeln der Schulleitung das Schulklima bedingt) oder auf der Ebene des Lehrerkollegiums bzw. der Schule als Ganze (z. B. wenn sich zeigen sollte, dass das Schulklima das Handeln der Schulleitung bedingt). Zum anderen soll die Rolle der Schulleitung und ihr Handeln in verschiedenen Akteurskonstellationen (d. h. in Zusammenarbeit mit Lehrkräften, Einrichtungen der Bildungsverwaltung u. a.) näher untersucht werden. Die Rolle der Schulleitung soll auch im Hinblick auf die Rezeption und Nutzung einzelner Evidenzquellen in den Blick genommen werden. Insbesondere bei der externen Evidenzquelle der Schulinspektionen klaffen der Grad der Auseinandersetzung und die tatsächliche Nutzung weit auseinander (s. o.).

Des Weiteren standen im EVIS-I-Projekt vor allem individuelle und organisationale Ressourcen im Forschungsfokus. Im Folgeprojekt sollen vor dem Hintergrund der vorliegenden Befundlage (s. o.) neben den Ressourcen auch insbesondere potenzielle Belastungsfaktoren auf individueller und organisationaler Ebene

(z. B. Zeitdruck, hohe Selbstregulationsanforderungen auf individueller Ebene, hoher Krankenstand, schlechte finanzielle Ausstattung, hohe Belastung durch Lernstanderhebungen auf organisationaler Ebene) sowie spezifische Ressourcen für den Umgang mit Belastungen (z. B. hohe Selbstregulationskompetenz, Unterstützung durch die Schulleitung) untersucht werden. So deuten die Befunde des Teilprojekts 1 darauf hin, dass sich die Kompetenz angehender Lehrkräfte beim Eintritt in den Schuldienst sogar rückläufig entwickelt und die angehenden Lehrenden eine Vielzahl von Belastungen angeben. Die erforderliche Analyse des Zusammenspiels der individuellen Kompetenz- und Bewältigungsprofile mit den organisationalen Umgebungsbedingungen sind somit auch das Ziel weiterer Untersuchungen in EVIS II.

Aus dem querschnittlichen Design, das im Rahmen des EVIS-I-Projektes realisiert werden konnte, können insgesamt noch keine Rückschlüsse auf Kausalitäten gezogen werden. Basierend auf der Erhebung der ersten Förderphase werden im Rahmen von EVIS II die teilnehmenden Schulen für eine zweite quantitative Erhebung gewonnen. Die Wiedergewinnung der Schulen aus der ersten Erhebungsphase ermöglicht auf Ebene der Schulleiter und der Schulen (= Organisationseinheiten) eine Längsschnittuntersuchung. So können mit diesem Design bei der in der zweiten Förderphase fokussierten Schulleiterrolle Ergebnisse aus der ersten Förderphase im Längsschnitt validiert werden und Hinweise zur Kausalität gewonnen werden.

Literaturverzeichnis

- Ackeren, I. van, Binnewies, C., Clausen, M., Dormann, C., Preisendörfer, P., Rosenbusch, C., Schmidt, U. & Zlatkin-Troitschanskaia, O. (2011). Evidenzbasierte Schulentwicklung. Ein Forschungsüberblick aus interdisziplinärer Perspektive. *DDS – Die Deutsche Schule, Jg. 103, H. 2, S. 170–184.*
- Ackeren, I. van, Binnewies, C., Clausen, M., Demski, D., Dormann, C., Koch, A. R., Laier, B., Preisendörfer, P., Preuße, D., Rosenbusch, C., Schmidt, U., Stump, M. & Zlatkin-Troitschanskaia, O. (2013). Welche Wissensbestände nutzen Schulen im Kontext von Schulentwicklung? Theoretische Konzepte und erste Befunde des EviS-Verbundprojektes im Überblick. *DDS – Die Deutsche Schule, 12. Beiheft, S. 51–73.*
- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, H. 52, S. 27–58.*
- Anderson, N. (2007). The practitioner-researcher divide revisited: Strategic-level bridges and the role of IWO psychologists. *Journal of Occupational and Organizational Psychology, H. 80, S. 175–183.*
- Ashkanasy, N. (2007). From the Editor. Evidence-based inquiry, learning, and education: What are the pros and cons? *Academy of Management Learning & Education, H. 6, S. 5–8.*
- Askew, S. & Lodge, C. (2000). Gifts, ping-pong and loops – linking feedback and learning. In S. Askew (Hrsg.), *Feedback for Learning, S. 1–17.* London und New York.

- Bach, A., Wurster, S., Thillmann, K., Pant, H. A. & Thiel, F. (2014). Vergleichsarbeiten und schulische Personalentwicklung – Ausmaß und Voraussetzungen der Datennutzung. *Zeitschrift für Erziehungswissenschaft*, 17 (1), S. 61–84.
- Baer, M. & Frese, M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior*, H. 24, S. 45–68.
- Baillod, J. & Semmer, N. K. (1994). Fluktuation und Berufsverläufe bei Computerfachleuten. *Zeitschrift für Arbeits- und Organisationspsychologie*, 38, S. 152–163.
- Bartunek, J. (2007). Academic-practitioner collaboration need not require joint or relevant research: toward a relational scholarship of integration. *Academy of Management Journal*, H. 50, S. 1323–1333.
- Berkemeyer, N., Feldhoff, T. & Brüsemeister, T. (2008). Schulische Steuergruppen – ein intermediärer Akteur zur Bearbeitung des Organisationsdefizits der Schule? In R. Langer (Hrsg.), *Warum tun die das? Governanceanalysen zum Steuerungshandeln in der Schulentwicklung*, S. 149–172. Wiesbaden: VS.
- Berkemeyer, N. & Bos, W. (2009). Professionalisierung im Spannungsfeld von interner und externer Evaluation. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität – Bedingungen, Genese, Wirkungen und ihre Messung*, S. 529–542. Weinheim: Beltz.
- Bonsen, M. (2010). Schulleitungshandeln. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 277–294. Wiesbaden: VS.
- Breiter, A. & Karbautzki, L. (2012): Data-Use in Schools – A Cross-Country Study. Institute for Information Management, University of Bremen. *International Congress for School Effectiveness and Improvement*. Malmö.
- Buchholzer, A. (2003): *Das Innovationsklima in Schulen. Eine empirische Studie zur Innovation „Schulen mit Profil“ in den Primarschulen des Kantons Luzern* (Schriftenreihe/Lehrerbildung Sentimatt Luzern, 13). Aarau: Bildung Sauerländer.
- Buske, R. (2014). *Die Entwicklung kollektiver Innovationsbereitschaft von Lehrkollegien – Eine theoretische Modellierung und empirische Untersuchung im Längsschnitt*. Landau: Verlag Empirische Pädagogik.
- Cameron, K. S. & Quinn, R. E. (2006). *Diagnosing and changing organizational culture: based on the competing values framework* (2. überarbeitete Aufl.). San Francisco: Jossey-Bass.
- Clausen, M. (2008). *Zur Wahl der konkreten weiterführenden Einzelschule aus der Sicht von Bildungsnachfragenden und Bildungsanbietenden*. Habilitationsschrift an der Fakultät für Sozialwissenschaften der Universität Mannheim.
- Cousins, J. B. & Leithwood, K. (1993). Enhancing knowledge utilization as a strategy for school improvement. *Knowledge: Creation, Diffusion, Utilization*, H. 3, S. 305–333.
- Cummings, T. (2007). Quest for an engaged Academy. *Academy of Management Review*, H. 32, S. 355–350.
- Dederig, K. & Müller, S. (2011). School Improvement through Inspections? International research findings and first insights from Germany. *Journal of Educational Change*, 12, S. 301–322.
- Demski, D. (2014): Which Data Do Principals and Teachers Use to Inform Their Practice? Evidence From Germany With a Focus und the Influence of School

- Culture. In A. J. Bowers, A. R. Shoho & B. G. Barnett (Hrsg.): *Using Data in Schools to Inform Leadership and Decision Making*, S. 121–150. Charlotte, NC: Information Age.
- Demski, D., Rosenbusch, C., Ackeren, I. van, Clausen, M. & Schmidt, U. (2012). Steuerung von Schule durch evidenzbasierte Einsicht? Konzeption und erste Befunde des Forschungsverbundes EviS. In S. Hornberg & M. Parreira do Amaral (Hrsg.), *Deregulierung im Bildungswesen*, S. 131–150. Münster: Waxmann.
- Demski, D., van Ackeren, I. & Clausen, M. (2015). Zum Zusammenhang von Schulkultur und evidenzbasiertem Handeln – Befunde einer Erhebung mit dem „Organizational Culture Assessment Instrument“. In O. Zlatkin-Troitschanskaia (Hrsg.), *Evidence-based Actions within the Multilevel System of Schools – Requirements, Processes, and Effects*. Topical Issue on *Journal for Educational Research* (accepted).
- Dormann, C., Binnewies, C., Koch, A. R., van Ackeren, I., Clausen, M., Preisendörfer, P., Schmidt, U. & Zlatkin-Troitschanskaia, O. (2015). Transferring Best Evidence into Practice: Assessment of Evidence-based School Management. In O. Zlatkin-Troitschanskaia (Hrsg.), *Evidence-based Actions within the Multilevel System of Schools – Requirements, Processes, and Effects*. Topical Issue on *Journal for Educational Research* (accepted).
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *H. 44*, S. 350–383.
- Ehren, M. C. M., Altrichter, H., McNamara, G. & O’Hara, J. (2013): Impact of school inspections on improvement of schools – describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, *H. 1*, S. 3–43.
- Felfe, J. (2006). Validierung einer deutschen Version des „Multifactor Leadership Questionnaire“ (MLQ Form 5x Short) von Bass und Avolio (1995). *Zeitschrift für Arbeits- und Organisationspsychologie*, *H. 50*, S. 61–78.
- Gärtner, H., Wurster, S. & Pant, H. A. (2014). The effect of school inspections on school improvement. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *25* (4), S. 489–508.
- Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78* (6), S. 1360–1380.
- Guest, D. (2007). Don’t shoot the messenger: A wake-up call for academics. *Academy of Management Journal*, *H. 50*, S. 1020–1026.
- Guest, D. & King, Z. (2004). Power, innovation and problem-solving: The personnel managers’ three steps to heaven? *Journal of Management Studies*, *H. 41*, S. 401–423.
- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In G. Brägger, B. Bucher & N. Landwehr (Hrsg.), *Schlüsselfragen zur externen Schulevaluation*, S. 127–151. Bern: h.e.p. Verlag.
- Hox, J. J. (2010). *Multilevel Analysis. Techniques and Applications*. Quantitative Methodology Series (Second Edition). New York: Routledge.
- Imants, J., Slegers, P. & Witziers, B. (2001). The Tension between Organisational Substructures in Secondary Schools and Educational Reform. *School Leadership & Management*, *H. 21/3*, S. 289–308.

- Jakku-Sihvonen, R. (2008). Use of Evidence – Development of an evaluation culture in education. *BMBF 2008: Wissen für Handeln – Forschungsstrategien für eine evidenzbasierte Bildungspolitik*, S. 85–91.
- Janssen, O. (2000). Job demands, perceptions of effort-reward fairness and innovative work behavior. *Journal of Occupational and Organizational Psychology*, 73, S. 287–302.
- Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, H. 1, S. 93–110.
- Jornitz, S. (2008). Was bedeutet eigentlich „evidenzbasierte Bildungsforschung“? Über den Versuch, Wissenschaft für Praxis verfügbar zu machen, am Beispiel der Review-Erstellung. *DDS – Die deutsche Schule*, H. 2, S. 206–216.
- Kadach, A. & Zlatkin-Troitschanskaia, O. (2009). Wirksamkeit der internen Evaluation als einzelschulisches Steuerungsinstrument. In W. Böttcher, J. N. Dicke & H. Ziegler (Hrsg.), *Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis*, S. 143–154. Münster: Waxmann.
- Kahnert, J., Eickelmann, B., Lorenz, R. & Bos W. (2015). Die Steuerungsfunktion von zentralen Abiturprüfungen. Analysen und kontroverse Einschätzungen der Aufgabenschwierigkeit und mögliche Rückkopplungen auf den Unterricht. In H. J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungssystem. Analysen zur Mehrebenenperspektive, Steuerung und Koordination. Reihe: Educational Governance, Band 26*, S. 89–116. Wiesbaden: VS.
- Klusmann, U. (2011). Allgemeine berufliche Motivation und Selbstregulation. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, S. 277–294. Münster: Waxmann.
- Kotthoff, H.-G. & Böttcher, W. (2010). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 295–325. Wiesbaden: VS.
- Kühn, S. (2010). *Steuerung und Innovation durch Abschlussprüfungen?* Wiesbaden: VS.
- Laier, B., Demski, D., van Ackeren, I., Clausen, M. & Preisendörfer, P. (2015). Die Bedeutung sozialer Netzwerke von Lehrkräften für evidenzbasiertes Handeln im schulischen Kontext. In O. Zlatkin-Troitschanskaia (Hrsg.), *Evidence-based Actions within the Multilevel System of Schools – Requirements, Processes, and Effects. Topical Issue on Journal for Educational Research* (accepted).
- Lambrecht, M. & Rürup, M. (2012). Bildungsforschung im Rahmen einer evidence based policy: Das Beispiel „Schulinspektion“. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung*, S. 57–77. Wiesbaden: VS.
- Latham, G. (2007). A speculative perspective on the transfer of behavioral science findings to the workplace: „The times they are a-changin“. *Academy of Management Journal*, H. 50, S. 1027–1032.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 145–169. Wiesbaden: VS.

- Maasen, S. (1999). *Wissenssoziologie*. Bielefeld: Transcript.
- Magerl, R. (2012). *Die Veränderung individueller Kognitionen im Rahmen der Aus- und Weiterbildung. Eine vergleichende Analyse am Beispiel von zukünftigen Lehrkräften* (unveröffentlichte Diplomarbeit). Johannes Gutenberg-Universität Mainz.
- Maier, U. & Kuper, H. (2012): Vergleichsarbeiten als Instrumente der Qualitätsentwicklung an Schulen. Überblick zum Forschungsstand. *Die Deutsche Schule, H. 1*, S. 89–100.
- McPherson, R. B., Crowson, R. & Pitner, N. J. (1986). *Managing uncertainty. Administrative theory and practice in education*. Columbus: C. E. Merrill Pub.
- Miller, L. L. & Nilakanta, S. (2001). A Committee-Based Model for Supporting Organizational Knowledge Management. *Annual Conference of the International Academy for Information Management*, S. 317–326.
- Mintzberg, H. (1989). *Mintzberg on management*. New York: Free Press.
- Mohr, N., Stumm, S. & Dormann, C. (2010). Wissen ist Macht! Vorteile eines evidenzbasierten Managements. Vortrag auf dem 56. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft. Darmstadt, 24. –26. März 2010.
- Parker, S. K. (1998). Enhancing role breadth self-efficacy: The roles of job enrichment and other organizational interventions. *Journal of Applied Psychology, H. 83*, S. 835–852.
- Pfeffer, J. & Sutton, R. (2006a). *Hard facts, dangerous half-truths and total nonsense*. Boston, MA: Harvard Business School Press.
- Pfeffer, J. & Sutton, R. (2006b). Management by Fakten. *Harvard Business Manager, H. 2 (April)*, S. 44–63.
- Pöhlmann, C., Pant, H. A., Frenzel, J., Roppelt, A. & Köller, O. (2014). Auswirkungen einer Intervention auf die Auseinandersetzung und Arbeit mit Bildungsstandards bei Mathematik-Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 17 (1)*, S. 113–133.
- Preuß, D. (2015). *Wissen und Überzeugungen von Lehrkräften – eine Analyse für den Kontext schul- und unterrichtsbezogener Evaluationen* (in press).
- Quinn, R. E. & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science, 29 (3)*, S. 363–377.
- Rentsch, J. R. (1990). Climate and culture: interaction and qualitative differences in organizational meanings. *Journal of Applied Psychology, H. 75*, S. 661–668.
- 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. *Zeitschrift für Pädagogik, 60 (2)*, S. 225–244.
- Rogers, E. M. (2003). *Diffusion of Innovations* (5. Aufl.). New York: Free Press.
- Rousseau, D. (2006). Is there such thing as „Evidence-based management“? *Academy of Management Review, H. 31*, S. 256–269.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W. & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM*. New York: Churchill Livingstone.

- Schaufeli, W. B., Bakker, A. B. & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross national study. *Educational and Psychological Measurement*, 66, S. 701–716.
- Scheerens, J. (2000). *Improving school effectiveness*. Paris: UNESCO.
- Schildkamp, K. & Kuiper, W. (2010): Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, H. 3, S. 482–496.
- Schrader, F.-W. & Helmke, A. (2004). *Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER: Ergebnisse der Lehrerbefragung*. *Empirische Pädagogik*, 18 (1), S. 140–161.
- Smith, C., Bingman, B. & Beall, K. (2007). *Research utilization in the field of adult learning and literacy: Lessons learned by NCSALL about connecting practice, policy, and research* (NCSALL Occasional Paper). Cambridge, MA: National Center for the Study of Adult Learning and Literacy.
- Snijders, T. & Bosker, R. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. London: SAGE.
- Song, J. H., Chermack, T. J. & Kim, H. M. (2008). An integrating individual learning processes and organizational knowledge formation: A perspective of continuous organizational performance improvement. In T. J. Chermack, J. Storberg-Walker & C. M. Graham (Hrsg.), *Academy of Human Resource Development (AHRD) conference proceedings*, S. 941–948. Panama City, FL: Academy of Human Resource Development.
- Stumm, S., Dormann, C. & Ohlemann, S. (2010). Evidenzbasiertes Management als Schlüssel für den Transfer zwischen Wissenschaft und unternehmerischer Praxis. Vortrag auf dem 47. Kongress der Deutschen Gesellschaft für Psychologie (DGPs). Bremen, 26. –30. September 2010.
- Stump, M., Zlatkin-Troitschanskaia, O. & Happ, R. (2013). Evidenzbasiertes professionelles Handeln in berufsbildenden Schulen – Forschungsstand und neue Perspektiven. In G. Niedermair (Hrsg.), *Facetten berufs- und betriebspädagogischer Forschung: Grundlagen, Herausforderungen und Perspektiven*. Schriftenreihe *Berufs- und Betriebspädagogik*. Band 8, S. 179–192. Linz: Trauner.
- Stump, M., Zlatkin-Troitschanskaia, O. & Mater, O. (2015). The Effects of Transformational Leadership on Teachers' Data Use. In O. Zlatkin-Troitschanskaia (Hrsg.), *Evidence-based Actions within the Multilevel System of Schools – Requirements, Processes, and Effects*. Topical Issue on *Journal for Educational Research* (in review).
- Thiel, F., Cortina, K. S. & Pant, H. A. (2014a). Steuerung im Bildungssystem im internationalen Vergleich. In R. Fatke & J. Oelkers (Hrsg.), *Das Selbstverständnis der Erziehungswissenschaft: Geschichte und Gegenwart (Zeitschrift für Pädagogik, Beiheft 60)*, S. 123–138. Weinheim: Beltz.
- Thiel, F., Hannover, B. & Pant, H. A. (2014b). Nutzung und Effekte zentraler Abschlussprüfungen und standardbasierter Schulleistungstests als Instrumente der Neuen Steuerung. *Zeitschrift für Erziehungswissenschaft*, H. 1, S. 3–6.
- Visscher, A. J. (1996). The Implications of how school staff handle information for the usage of school information systems. *International Journal of Educational Research*, H. 4, S. 323–334.

- Visscher, A. J. (2001). Publieke schoolprestatie-indicatoren: de problemen op een rij. In A. B. Dijkstra, S. Karsten, R. Veenstra & A. J. Visscher (Hrsg.), *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties*, S. 54–61. Assen: van Gorcum.
- Walshe, K. & Rundall, T. (2001). Evidence-based management: from theory to practice in health care. *Milbank Quarterly*, 79 (3), S. 429–457.
- Wayman, J. C., Jimerson, J. B. & Cho, V. (2012): Organizational considerations in establishing the Data-Informed District. *School Effectiveness and School Improvement*, H. 2, S. 159–178.
- Weiss, C. H. (1984). Toward the future of stakeholder approaches in evaluation. In R. F. Conner, D. G. Altman & C. Jackson (Hrsg.), *Evaluation Studies Review Annual*, 9. Beverly Hills: Sage.
- Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2014). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. In C. Fischer (Hrsg.), *Damit Unterricht gelingt. Von der Qualitätsanalyse zur Qualitätsentwicklung*, S. 53–84. Münster: Waxmann.
- Zlatkin-Troitschanskaia O., Beck K., Sembill D., Nickolaus R. & Mulder R. (2009). *Lehrprofessionalität – Bedingungen, Genese, Wirkungen und ihre Messung*. Weinheim: Beltz.
- Zlatkin-Troitschanskaia, O. & Förster, M. (2009). Wirkung der Lehrprofessionalität auf Schulorganisation und Schulentwicklung. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität – Bedingungen, Genese, Wirkungen und ihre Messung*, S. 729–741. Weinheim: Beltz.
- Zlatkin-Troitschanskaia, O., Seidel, J. & Stump, M. (2013). Teachers' evidence-based actions. A comparison of different school types. In K. Beck & O. Zlatkin-Troitschanskaia (Hrsg.), *From Diagnostics to Learning Success. Proceedings in Vocational Education and Training*, S. 281–294. Rotterdam: Sense Publishers.
- Zlatkin-Troitschanskaia, O. (Hrsg.) (2015). Evidence-based Actions within the Multi-level System of Schools – Requirements, Processes, and Effects. Topical Issue on *Journal for Educational Research* (accepted, in preparation).
- Zlatkin-Troitschanskaia, O., Förster, M., Preusse, D., Mater, O. & Stump, M. (2015). Influence of Communication, Cooperation, and Participation Structures at Schools on Teachers' Evidence-Based Actions. In O. Zlatkin-Troitschanskaia (Hrsg.), *Evidence-based Actions within the Multilevel System of Schools – Requirements, Processes, and Effects*. Topical Issue on *Journal for Educational Research* (in review).

*Harm Kuper, Uwe Maier, Tanja Graf, Barbara Muslic,
Carolin Ramsteck¹*

Datenbasierte Schulentwicklung mit Vergleichsarbeiten aus der Perspektive von Lehrkräften, Fachkonferenzleitungen, Schulleitungen und Schulaufsichten – Qualitative Fallstudien aus vier Bundesländern

Zusammenfassung

Mit Vergleichsarbeiten (VERA) soll geprüft werden, ob Schülerinnen und Schüler kurz vor Abschluss eines Bildungsabschnitts über zentrale Kompetenzen verfügen. Bisherige Rezeptionsstudien zeigen allerdings, dass die Nutzung der Testdaten an den Schulen hinter den Erwartungen zurückbleibt. Ziel dieser Studie ist es, die vorwiegend quantitativen Befunde der bisherigen Rezeptions- und Nutzungsstudien durch qualitative, fallorientierte Analysen zu erweitern. Welche Formen der instrumentellen Nutzung werden von Lehrkräften berichtet, wie werden diese didaktisch begründet, welche schulinternen Formen der Kommunikation unterstützen die Nutzung? Welche Rolle spielt das Verhältnis zur Schulaufsicht?

Die empirische Basis für diese Studie² bilden insgesamt 19 Gymnasien (Fälle) in vier Bundesländern mit jeweils unterschiedlichen Regelungsstrukturen bezüglich VERA. Es wurden 229 Interviews mit Lehrkräften, Fachbereichsleitungen, Schulleitungen und den zuständigen Schulaufsichten geführt. Die Ergebnisse bestätigen insgesamt ein eher geringes Nutzungsniveau. Formen der instrumentellen Nutzung sind vorwiegend inhaltsorientiert. Ein an Kompetenzstufen orientierter Umgang war nicht erkennbar. Eine systematische, schulinterne Kommunikation (vor allem auf Fachbereichsebene) ist selten, kann die instrumentelle Nutzung aber unterstützen. Auf Ebene der Schulaufsicht spielt VERA entweder keine Rolle, oder es besteht ein diffuses Verhältnis zwischen schulischer Datenhoheit sowie Kontrolle und Unterstützung. Dies kann nur teilweise durch die institutionellen Regelungsstrukturen der unterschiedlichen Bundesländer erklärt werden. Eine ausgeprägte Kommunikation – nicht nur über organisatorische Belange – sowie Rollenklarheit

- 1 Freie Universität Berlin (Harm Kuper, Tanja Graf, Barbara Muslic), Pädagogische Hochschule Schwäbisch Gmünd (Uwe Maier, Carolin Ramsteck), unter Mitarbeit von Dr. Annette Strößner und Thomas Reß, M. A.
- 2 Das Projekt wurde vom Bundesministerium für Bildung und Forschung unter dem Forschungsschwerpunkt „Steuerung im Bildungswesen“ gefördert und im Verbund von Wissenschaftlern der Freien Universität Berlin, der Friedrich-Alexander-Universität Erlangen-Nürnberg und der Pädagogischen Hochschule Schwäbisch Gmünd durchgeführt.

seitens der Schulaufsicht wirken sich günstig auf die innerschulische Datennutzung aus.

1. Forschungsstand

In Deutschland wurden vor etwa einem Jahrzehnt nach entsprechenden Beschlüssen der Kultusministerkonferenz (KMK) in allen Bundesländern zentrale Vergleichsarbeiten (VERA) für die Primar- und die Sekundarstufe I implementiert. Mit VERA soll geprüft werden, welche fachlichen Kompetenzen die Schülerinnen und Schüler in den Kernfächern bis kurz vor Abschluss der jeweiligen Schulstufe erworben haben (vgl. KMK 2006). In der Sekundarstufe I werden die vom *Institut für Qualitätsentwicklung im Bildungswesen (IQB)* entwickelten Vergleichsarbeiten in der Jahrgangsstufe 8 für Deutsch, Mathematik und die erste Fremdsprache seit dem Schuljahr 2009/2010 in mindestens einem Fach obligatorisch durchgeführt (vgl. KMK 2012). Einzige Ausnahme ist Baden-Württemberg, das bis zum Schuljahr 2014/2015 noch eigenständig entwickelte Vergleichsarbeiten zu Beginn der Jahrgangsstufen 7 und 9 durchführte (ab 2015 Teilnahme an VERA 8).

Primäres Ziel von VERA ist die Schul- und Unterrichtsentwicklung (vgl. KMK 2012). Bildungspolitiker erwarten, dass Lehrkräfte, Fachbereichsleitungen und Schulleitungen die Vergleichsarbeitsrückmeldungen sowie die fachdidaktischen Handreichungen heranziehen, um den Austausch über die Unterrichtsqualität anzustoßen und gegebenenfalls Veränderungen des Unterrichts vorzunehmen. In einer Reihe von empirischen Studien wurden Lehrkräfte zu Akzeptanz, Verständlichkeit und Nutzung von VERA-Rückmeldungen befragt. In den folgenden Abschnitten sollen wichtige Befunde entlang der für die Rezeption und Nutzung zentralen Akteure zusammengefasst werden (vgl. Maier & Kuper 2012). Wenn angebracht und sinnvoll, wird dabei auch auf internationale Befunde zu Effekten von zentralen Tests in Ländern mit einem High-Stakes-Kontext verwiesen (vgl. Maier 2010).

1.1 Rezeption und Nutzung externer Testdaten auf Ebene der Lehrkräfte und des Unterrichts

Lehrkräfte müssen unterrichten, Schülerleistungen diagnostizieren, bei Bedarf einzelne Schülerinnen und Schüler individuell fördern und Noten vergeben. Gemeinsam mit den Bildungsstandards könnte VERA dazu beitragen, dass sich Lehrkräfte auf wichtige fachliche Lernziele und Kompetenzen konzentrieren. Lehrkräfte erhalten mit VERA zudem ein externes Instrument, um die fachliche Kompetenzentwicklung der Klasse mit den eigenen Einschätzungen abzugleichen. Auf individueller Ebene sollen vor allem die didaktischen Handreichungen wertvolle Hinweise zur Analyse von Schülerfehlern liefern. Ebenso könnte die enge Anbindung der VERA-Tests sowie der didaktischen Handreichungen an aktuelle fachdidaktische Innovationen dazu beitragen, dass Lehrkräfte über inhaltliche oder methodische Alternativen für ihren Unterricht nachdenken.

Empirische Studien zeigen, dass die Mehrheit der Lehrkräfte der Primarstufe eine eher positive Einstellung gegenüber den zentralen Tests äußert (Frühwacht 2012). Langfristige Trendanalysen aus Surveystudien zu VERA 3 (vgl. Groß Ophoff 2013a) und den Thüringer Kompetenztests (vgl. Nachtigall & Hellrung 2013) decken jedoch konträre Einstellungen auf. Sowohl Akzeptanz als auch Rezeption und Nutzung gehen mit den Jahren insgesamt zurück. Die Durchführung und Auswertung der Vergleichsarbeiten werden von vielen Lehrkräften als zeitaufwendig beklagt, bereiten aber generell keine Schwierigkeiten. Ebenso geben Lehrkräfte an, dass sie bei der Interpretation der Rückmeldedaten keine Verständnisschwierigkeiten haben (vgl. Groß Ophoff et al. 2006; Nachtigall 2005; Nachtigall & Jantowski 2007). Wenn Verständnisprobleme auftreten, dann ist dies vermehrt bei Lehrkräften in sprachlichen Fächern der Fall (vgl. Kühle 2010). Experimentelle Studien zeigen allerdings, dass die Testdurchführung und Testauswertung durch Lehrkräfte zu Beurteilereffekten führen können. Im Vergleich zu externen Testleitern und externen Beurteilern bewerten Lehrkräfte die Vergleichsarbeiten ihrer Klasse eher milde, obwohl im deutschen Schulsystem keine ernsthaften Konsequenzen bei unterdurchschnittlichen Leistungsergebnissen drohen (vgl. Graf, Emmrich, Harych & Brunner 2013; Koch & Hosenfeld 2013; Spoden, Fleischer & Leutner 2014).

Die Nutzung der VERA-Daten wird in den meisten Studien als sehr kritisch beschrieben (vgl. z. B. Bensen et al. 2006; Groß Ophoff et al. 2006; Nachtigall & Jantowski 2007; Kuper & Hartung 2007; Maier 2009; Diemer 2013). Eine systematische Nutzung als Evaluationsinstrument wird von vielen Lehrkräften vor allem unter Hinweis auf die mangelnde curriculare Validität der Tests abgelehnt. Sie findet am ehesten dann statt, wenn die Nützlichkeit als sehr hoch eingeschätzt wird (vgl. Bensen et al. 2006; Kühle 2010). Lehrkräfte sehen VERA eher als weiteres Diagnoseinstrument, um die Leistungen der eigenen Klasse einschätzen zu können. Als Begründung für die mangelnde Nutzung der Tests wird oft der Wunsch nach individueller Förderung einzelner Schülerinnen und Schüler genannt, was allerdings aufgrund der geringen Aussagekraft der VERA-Daten auf individueller Ebene nicht möglich scheint und auch nicht intendiert ist (vgl. Maier 2009).

In einigen Studien wurde versucht, bestimmte Rezeptions- und Nutzungstypen zu identifizieren (vgl. Groß Ophoff et al. 2007; Jäger 2012; Groß Ophoff 2013a). Groß Ophoff et al. (2007) unterscheiden auf der Basis von quantitativen VERA-Befragungsdaten drei Rezeptionstypen. Rezeptionstyp 1 fällt auf durch die stärkere Beachtung der Ergebnisse auf Individual- und Klassenebene. Typ 2 setzt sich kaum mit den Ergebnissen auseinander, schätzt den Nutzen dennoch hoch ein. Der eher kritische Rezeptionstyp 3 zeichnet sich vor allem dadurch aus, dass er die Nützlichkeit der Ergebnisse sehr niedrig einstuft. Interessant ist hierbei, dass sich diese Rezeptionstypen auch hinsichtlich der realisierten Unterrichtsentwicklungsmaßnahmen unterscheiden.

Die Nutzung der VERA-Daten hängt mit bestimmten Faktoren zusammen. Die Rückmeldungen sollten recht schnell nach der Testauswertung vorliegen und sich auf faire Landesvergleichswerte beziehen (vgl. Nachtigall 2005; Nachtigall & Jantowski 2007). Ein weiterer kritischer Faktor ist das Verständnis der kriterialen Bezugsnorm. Wenn Lehrkräfte einen Bezug zwischen Vergleichsarbeitsrückmeldungen

und den auf Kompetenzmodellen basierenden Niveaustufen herstellen können (kri-
teriale Bezugsnorm), wurde auch eher von einer Veränderung des Unterrichts auf-
grund von VERA berichtet (vgl. Groß Ophoff 2013b).

An dieser Stelle sollen einige gut abgesicherte Befunde aus Ländern mit High-
Stakes-Kontext erwähnt werden, um die sich in Deutschland abzeichnende Befund-
lage einordnen zu können. Ein empirisch gut abgesichertes Ergebnis der vor allem
US-amerikanischen Forschung zu Testeffekten ist die Reduktion der verfügbaren
Unterrichtszeit durch Testvorbereitungsaktivitäten (vgl. Stecher 2002; Cheng & Cur-
tis 2004; Herman 2004; Watanabe 2004). Lehrkräfte setzen veröffentlichte Aufgaben
früherer Tests oder kommerzielle Materialien ein, um Schülerinnen und Schüler auf
die Prüfungen vorzubereiten. Nicht getestete Fächer und Inhalte verlieren an Bedeu-
tung. Ebenso gibt es in den USA eine breite Befundlage zu unseriösen Praktiken der
Erhöhung von Testwerten (z. B. leistungsschwache Schülerinnen und Schüler vom
Test ausschließen). In Deutschland sind derartige Praktiken nicht belegt, allerdings
deuten einige Befunde tendenziell in diese Richtung. Dazu zählen: milde Bewertungen
durch Lehrkräfte, Zunahme von Publikationen für die Testvorbereitung oder
Üben von Testaufgabenformaten.

Ebenfalls international gut untersucht sind Effekte von sog. testbasierten Unter-
richtsreformen (vgl. Popham 1987). Gemeint ist damit die gezielte Veränderung von
Unterricht über bestimmte Testaufgabenformate, v. a. in Abschlussprüfungen (z. B.
innovative Schreibaufgaben), aber auch in zentralen Tests (z. B. mathematische Pro-
blemlöseaufgaben, Portfolio-Assessments), d. h. im Grunde genommen ein positives
„teaching to the test“. In Studien konnten Effekte staatlicher Testreformen auf Un-
terrichtsprozesse gezeigt werden (vgl. Firestone et al. 2000; Lane et al. 2002; Cheng
2003; Parke et al. 2006; Au 2007; Luxia 2007). Allerdings wird in diesen Publikationen
immer wieder darauf hingewiesen, dass die Lehrkräfte einzelne neuartige Testaufga-
benformate zwar übernehmen, ihren Unterricht deswegen aber nicht grundlegend
überdenken oder gar ändern würden.

1.2 Rezeption und Nutzung von Testdaten auf Ebene von Fachkonferenzen

Fachkonferenzen sind eine zweite Ebene in der Organisation der Schule, auf der eine
Auseinandersetzung mit den VERA-Testrückmeldungen erwartet wird. In Fachkon-
ferenzen sollten vor allem schuljahresübergreifende Fragen der fachlichen Kompe-
tenzentwicklung der Schülerinnen und Schüler diskutiert werden. Dieses gemein-
sam geteilte Verständnis von Lernentwicklung in einer Fachdomäne bildet eine
wichtige Grundlage für die gemeinsame Ausgestaltung des Fachunterrichts an einer
Schule (Anschaffung von Materialien, Schulbüchern, neue Themen und Methoden,
gemeinsame Standards für die Leistungsbewertung, Planung von Fortbildungen
usw.). VERA-Testdaten könnten in Kombination mit weiteren Leistungsdaten (z. B.
Abschlussprüfungen) genutzt werden, um die fachliche Kompetenzentwicklung an
einer Schule gemeinsam zu reflektieren.

Was sagen die bisherigen empirischen Studien über die Rezeption und Nutzung von VERA-Ergebnissen in Fachkonferenzen? Zunächst einmal ist festzuhalten, dass es viele Schulen gibt, an denen sich die Fachkonferenzen überhaupt nicht mit VERA auseinandersetzen oder lediglich organisatorische Fragen zur Durchführung der Tests thematisieren (vgl. Maier 2012; Bach et al. 2014). Wenn VERA-Daten Thema in Fachkonferenzen sind, findet man unterschiedliche Muster der Reflexion. Es dominieren kriteriale (in Bezug auf Niveaustufen) und sozial-vergleichende (in Bezug auf landesweite Vergleichswerte) Reflexionsmuster. Eine an einzelnen Schülerinnen und Schülern orientierte Reflexion ist eher selten, was allerdings der Funktion von Fachkonferenzen durchaus entspricht (vgl. Kühle & Peek 2007; Kühle 2010). Ebenso findet praktisch keine zusätzliche Informationsbeschaffung zur besseren Dateninterpretation statt (vgl. Kühle & Peek 2007).

Wenn Diskussionen über VERA in Fachkonferenzen stattfinden, dann beziehen sich diese eher auf Inhalte und Aufgabenformate, weniger dagegen auf Möglichkeiten der Schülerförderung oder der Unterrichtsgestaltung (vgl. Kühle & Peek 2007; Kühle 2010). Dies korrespondiert mit den bei Lehrkräften festgestellten Nutzungsformen. VERA-Rückmeldungen tragen damit im Prinzip zu einer Inputsteuerung von Unterricht bei. Eine Nutzung im Sinne der Evaluation des eigenen Unterrichts tritt in den Hintergrund. VERA-Ergebnisse werden auch nur in wenigen Fällen zur Fortbildungsplanung herangezogen (vgl. Bach et al. 2014). Wenn eine Planung von Fortbildungen aufgrund von VERA-Rückmeldungen in Fachkonferenzen stattfindet, hängt dies von der Unterstützung durch die Schulleitung und einer kooperativen Datenauswertung ab (ebd.). Diese beiden Kontextfaktoren bestätigten sich auch in anderen Studien. Beispielsweise berichten Lehrkräfte eher von einer Auseinandersetzung mit VERA in Fachkonferenzen, wenn gleichzeitig die kollegiale Kooperation, die selbst erlebte Autonomie der Fachkonferenz und das Innovationspotenzial der Schule hoch eingeschätzt werden (vgl. z. B. Maier et al. 2012; Zeitler 2010).

Internationale Studien zeigen, dass gezielte Fortbildungsprogramme auf Fachebene zur Nutzung von Testdaten erfolgreich sein können. Wikeley, Stoll und Lodge (2002) berichten beispielsweise, dass Lehrerinnen und Lehrer durch die Teilnahme an fachspezifischen Fortbildungen zusehends in die Lage versetzt wurden, externe Leistungsdaten schülerspezifisch zu interpretieren. In einer experimentellen Studie in verschiedenen US-Bundesstaaten konnte sogar gezeigt werden, dass sich ein Fortbildungsprogramm zur Rezeption und Nutzung von Testdaten positiv auf die Schülerleistungen auswirken kann (vgl. Carlson et al. 2011). Diese Fortbildungsprogramme fokussieren jedoch sehr stark auf fachdidaktische Fragen und Fragen der Schülerförderung. Die bisher in Deutschland evaluierten Fortbildungsprogramme versuchen dagegen, die Dateninterpretationskompetenz der Lehrkräfte zu fördern, und zielen damit nicht auf den eigentlich wichtigen Schritt der fachlichen Unterrichtsentwicklung (vgl. z. B. Koch 2011).

1.3 Prozesse der Rezeption und Nutzung von Testdaten in Einzelschulen

Bei der Rezeption und Nutzung externer Schulleistungsdaten an Einzelschulen muss auch die Rolle der Schulleitungen in Bezug auf die anderen Akteure in den Blick genommen werden. Im Rahmen der Mitarbeiterführung könnten Schülerleistungsdaten bei sensiblem Gebrauch sowohl für die Kontrolle als auch für Beratung und Unterstützung durch die Schulleitung herangezogen werden. Wenn es um die Planung und Durchsetzung von Maßnahmen geht (z. B. Raum- und Materialausstattung, pädagogisches Konzept usw.), könnten Leistungsdaten als weitere Argumentationsgrundlage gegenüber Eltern, Schulträger und Schulverwaltung dienen. Ebenso müsste man erwarten, dass die Schulleitung den systematischen Umgang mit externen Evaluationsdaten an ihrer Schule organisiert und von den anderen Akteuren Berichte einfordert.

In einigen Rezeptionsbefragungen zeigte sich, dass Schulleiterinnen und Schulleiter im direkten Vergleich mit Lehrkräften immer höhere Akzeptanzwerte für die zentralen Tests angeben (vgl. Kühle 2010; Nachtigall & Hellrung 2013). Ähnlich wie bei Lehrkräften sinken diese Akzeptanzwerte jedoch mit der Zeit. Es scheint, dass Schulleitungen zumindest aufgrund ihrer Position eine größere Affinität zu testbasierten Evaluationsformen entwickeln. Nach Wurster et al. (2013) schätzen Schulleitungen allerdings die Nützlichkeit von VERA wesentlich geringer ein als die Nützlichkeit der externen oder internen Evaluation. Zudem findet eher keine systematische Moderation der Nutzung von VERA-Daten durch Schulleitungen statt. Die Kommunikation über VERA-Ergebnisse ist im Vergleich zur Kommunikation über externe und interne Evaluation am geringsten. Bach et al. (2014) zeigten in einer ebenfalls quantitativen Befragung von Schulleitungen, dass VERA-Daten für Schulentwicklungsprozesse und Personalentwicklung im Vergleich zur Unterrichtsentwicklung kaum eine Rolle spielen. Wenn an Schulen VERA-Daten dann doch für die Personalentwicklung genutzt werden, hängt dies sehr stark mit der wahrgenommenen Nützlichkeit der Testrückmeldungen zusammen.

Zur Rolle der Schulleitung für die testdatenbasierte Schulentwicklung liegen einige internationale Studien vor (vgl. z. B. Yang et al. 1999; Saunders 2000; Louis et al. 2005; Schildkamp & Kuiper 2010; Schildkamp & Visscher 2010). Das durch die Schulleitung beförderte Innovationsklima an Schulen ist eine zentrale Kontextvariable für die datenbasierte Schulentwicklung. In Schulen mit geringen Veränderungskapazitäten beispielsweise standen die entwickelten Testrückmeldesysteme recht unverbunden anderen Initiativen der Qualitätsentwicklung gegenüber (vgl. Wikeley, Stoll & Lodge 2002). Vanhoof et al. (2012) untersuchten, wie belgische Schulleiterinnen und Schulleiter den Prozess der Rezeption und Nutzung von Testdaten moderieren. Die Datennutzung ist dann hoch, wenn die Schulleitung generell Wert auf die Entwicklung einer Schulkultur legt und die Kooperation der Lehrkräfte in den Fachkonferenzen befördert. Die Datennutzung wird auch befördert, wenn Schulleitungen zusätzliche Evaluationsdaten erfassen wie z. B. Schülerbefragungen zum Unterricht oder Befragungen der Lehrkräfte zu möglichen Ursachen für die Leistungsdaten (vgl. Schildkamp & Kuiper 2010).

1.4 Die Rolle der Schulaufsicht beim Umgang mit Testdaten

Bisher gibt es erst wenige Arbeiten zur Rezeption und Nutzung von Vergleichsarbeitsdaten durch Schulaufsichtsvertreterinnen und -vertreter (vgl. Ramsteck, Muslic, Graf, Maier & Kuper 2015). Ein Grund hierfür ist, dass die Rezeption und Nutzung von Testdaten in Deutschland allein innerhalb der Schulen stattfinden soll und deshalb die innerschulischen Akteure zunächst im Blickfeld der Forschung standen. Diese Sichtweise greift allerdings zu kurz, weil die Rolle der Schulaufsicht bei der Rezeption und Nutzung von VERA-Daten in vielen Bundesländern sowohl rechtlich als auch konzeptionell nicht immer eindeutig ist und dies auch innerhalb der Schulen zu Unsicherheit über die mögliche Verwendung der Daten führen kann.

Es gibt konzeptionelle Überlegungen, wie die Schulaufsicht aktiv eingreifen könnte, um die schulinterne Nutzung von Testdaten zu befördern, beispielsweise über gezielte Schulungen oder eine Moderation von schulinternen Netzwerken (vgl. Butzke-Rudzynski & Bachorz 2013). Oder es wird von Schulen gefordert, dass sie ihre Auseinandersetzung mit VERA-Rückmeldungen im Rahmen einer externen Evaluation dokumentieren. Diese Dokumentation kann dann wiederum die Schulaufsicht einsehen, so wie es bspw. in Baden-Württemberg der Fall ist. In Berlin und Brandenburg erhalten die Bildungsverwaltungen auf Schulebene aggregierte VERA-Ergebnisse und können auf Anfrage an die einzelnen Bezirksschulaufsichten weitergegeben werden. In Thüringen hingegen gibt es keine Weitergabe der schulbezogenen VERA-Daten. Die VERA-Ergebnisse werden lediglich auf Schulumtsebene aggregiert und in Berichtsform zur Verfügung gestellt.

Aus Ländern mit High-Stakes-Kontext gibt es zahlreiche empirische Befunde, die erfolgreiche Möglichkeiten der Unterstützung von datenbasierter Schulentwicklung durch die Schulaufsicht aufzeigen (vgl. z. B. Yang et al. 1999; Hayes & Rutt 1999; Demie 2003; Peng et al. 2006; Levin & Datnow 2012). Ein wichtiges Ergebnis dieser Studien ist, dass die schulinterne Nutzung externer Leistungsdaten von außen gezielt unterstützt werden muss (vgl. Opfer, Henry & Masburn 2008). Die Kooperation innerhalb des Kollegiums und mit der Schulaufsicht sind weitere Voraussetzungen für eine sinnvolle Dateninterpretation auf Schulebene (vgl. Rudd & Davies 2002).

Als besonders hilfreich stellte sich heraus, wenn lokale Schulbehörden sinnvolle Leistungsindikatoren auswählen, mit weiteren organisationalen Variablen verknüpfen und in verständlicher Form den Schulen zur Verfügung stellen (vgl. Demie 2003; Louis et al. 2005). Die Schulen bei der Bildung von Netzwerken zu unterstützen ist erfolversprechender, als lediglich Zielvereinbarungen abzuschließen (vgl. Lee et al. 2012). Eine effektive Schnittstelle zwischen Schulverwaltung und Einzelschule sind Experten, die sich auf die Dateninterpretation und -nutzung spezialisiert haben (vgl. Hayes & Rutt 1999). Gerade innerhalb der Einzelschule kommt es darauf an, dass Lehrkräfte mit einer statistischen Expertise bei der Übersetzung und Interpretation von Daten helfen können (vgl. Harris et al. 1997; Yang et al. 1999; Wikeley, Stoll & Lodge 2002).

2. Theoretische Modelle der Datennutzung und Fragestellungen

Wie lassen sich die bisherigen Befunde zur Rezeption und Nutzung von Vergleichsarbeiten theoretisch einordnen, und welche Fragen bleiben weiterhin offen? In einem idealtypischen Rezeptions- und Nutzungsmodell (vgl. Helmke 2004; Groß Ophoff 2013a) wird angenommen, dass externe Testrückmeldungen zunächst einmal von den schulischen Akteuren rezipiert und reflektiert werden und sich daran eine Phase der Nutzung anschließt, die wiederum evaluiert wird. Die in den Studien aufgezeigten Brüche und Probleme der Rezeption von Testdaten können mit diesem Modell jedoch nicht aufgezeigt oder erklärt werden. In internationalen Publikationen wird häufig das Modell für *School Performance Feedback Systems* (SPFS-Modell) herangezogen (vgl. Visscher & Coe 2003; Van Petegem & Vanhoof 2007; Verhaeghe et al. 2010). Das SPFS-Modell beschreibt weniger den idealtypischen Fall der Rezeption und Nutzung von Testdaten, sondern fasst vielmehr die empirischen Befunde übersichtlich zusammen. Es werden vor allem schulische Kontextfaktoren, aber auch Merkmale des Testsystems bzw. der Testimplementation als wichtige Einflussfaktoren für die Nutzung von Testdaten beschrieben.

Im Rahmen des SPFS-Modells ist vor allem die Unterscheidung zwischen instrumenteller, konzeptueller und strategischer bzw. symbolischer Nutzung von Testdaten relevant (vgl. Visscher & Coe 2003; Schildkamp & Kuiper 2010; Verhaeghe et al. 2010). Instrumentelle Nutzung bedeutet, dass Maßnahmen sachlich mit Evaluationsdaten begründet werden. An einer Schule wird beispielsweise ein Programm zur Förderung des selbstständigen Übens entwickelt, wenn sich in den Tests herausstellte, dass Grundfertigkeiten schlecht ausgeprägt sind. Die Lehrkräfte haben in diesem Fall ein Konzept, wie etwas zu verändern ist, und die Testdaten sind ein hinreichender Anlass, aufgrund dieser Theorie eine bestimmte Maßnahme zu ergreifen. Eine konzeptuelle Nutzung liegt dann vor, wenn es nicht direkt zur Maßnahmenableitung kommt, die Akteure jedoch generell bei ihren Entscheidungen die aktuellen Leistungsdaten der Schülerschaft mit reflektieren. Strategisch bzw. symbolisch werden die Daten genutzt, wenn damit sachfremde Begründungen für andere Zielsetzungen konstruiert werden.

Betrachtet man insgesamt den Stand der bisherigen Forschung zu Vergleichsarbeiten in Deutschland, dann findet man vor allem auf Ebene der Lehrkräfte eine Vielzahl von Einzelbefunden zur Rezeption und Nutzung von externen Schülerleistungsdaten, die insgesamt nur geringfügig auf eine konstruktive Nutzung hinweisen. Eine genaue Analyse der Nutzungsformen wird jedoch in den meisten Studien nicht geleistet, weil mit quantitativen Fragebögen die tatsächliche Nutzung und ihr didaktischer Begründungszusammenhang kaum erfasst werden können. In vielen Studien werden Kontextfaktoren einer intendierten Nutzung auf den Ebenen der Schule oder des Testsystems beschrieben (z. B. Innovationsklima der Schule, kollegiale Datenreflexion, Sofortrückmeldungen, faire Vergleiche, Verständnis der kriterialen Bezugsnorm usw.). Auch diese Befunde sollten kritisch hinterfragt werden, weil beispielsweise eine generell schulsystemkritische Haltung einer Lehrkraft sowohl die Bewertung der Testdatennutzung als auch die Sicht auf das Innovationsklima an der eigenen Schule beeinflussen könnte.

Unklar bleiben dagegen die sehr viel komplexeren Kommunikations- und Handlungsstrukturen in einzelnen Schulen. Hier spielen die Perspektiven verschiedener Akteure an einer Schule auf den Gegenstand „datenbasierte Schulentwicklung“ eine wesentliche Rolle. Gerade wenn Vergleichsarbeiten auf Schulebene als Evaluationsdaten rezipiert und idealerweise instrumentell genutzt werden sollen, kommt es auf die komplexen Aushandlungsprozesse innerhalb einzelner Schulen und die für diese Schule spezifischen Kontextbedingungen an.

Forschungsdesiderate bestehen hinsichtlich der Verarbeitung von VERA-Rückmeldungen in der Mehrebenenstruktur der schulischen Organisation. Ziel dieses Forschungsprojektes ist es, die stark quantitativ und variablenorientiert ausgerichtete Rezeptions- und Nutzungsforschung durch Befunde zu ergänzen, die aus qualitativen Fallstudien stammen und die Mehrebenenstruktur in Schulen als Grundlage für die Modellierung der Datennutzung berücksichtigen. In diesem Forschungsprojekt sollen deshalb folgende Fragen beantwortet werden:

1. Findet auf Ebene der Lehrkräfte und Fachkonferenzen eine kriteriale Interpretation der Leistungsdaten statt und führt diese zu einer instrumentellen Nutzung, beispielsweise zu einer Ableitung von Maßnahmen, die sich an den in den didaktischen Handreichungen beschriebenen Kompetenzmodellen und Kompetenzstufen orientiert?
2. In vielen Studien werden kollegiale Kooperation und Innovationspotenziale einer Schule als wichtige Voraussetzungen für die Datennutzung beschrieben. Es werden allerdings nur selten konkrete Positivbeispiele auf Schulebene untersucht. Es soll deshalb gefragt werden, welche konkreten Formen der Kommunikation und Interaktion innerhalb einer Schule (zwischen Schulleitung, Fachkonferenzen und Lehrkräften) die instrumentelle Nutzung von Testrückmeldungen unterstützen können?
3. Internationale Studien deuten darauf hin, dass eine aktiv die Datennutzung unterstützende Schulaufsicht eher eine datenbasierte Schulentwicklung an Einzelschulen befördern kann als eine Schulaufsicht, die lediglich auf Kontrolle, Zielvereinbarungen und Sanktionen setzt. Gibt es hierfür auch Hinweise im deutschen Schulsystem? Wie sehen Schulaufsichtsvertreterinnen und -vertreter die Datennutzung an den Schulen, und wie agieren sie vor dem Hintergrund der jeweils bundeslandspezifischen Regelungen, um die instrumentelle Nutzung von Vergleichsarbeitsrückmeldungen zu fördern? Wie nehmen Schulleitungen im Gegenzug die Rolle und die Aktivitäten der Schulaufsicht wahr?

3. Methodisches Vorgehen

3.1 Untersuchungsdesign

Das Untersuchungsdesign orientiert sich an internationalen Studien, die Prozesse der Rezeption und Nutzung von externen Testdaten in Einzelschulen mithilfe verschiedener, vor allem qualitativer Datenquellen aus unterschiedlichen Akteursperspektiven beschreiben (vgl. z. B. Wikeley, Stoll & Lodge 2002; Schildkamp & Kuiper 2010; Schildkamp & Visscher 2010; Zeitler 2010). Es wurden relevante Akteure (Fachlehrkräfte, Fachkonferenzleitungen und Schulleiterinnen bzw. Schulleiter) innerhalb einzelner Gymnasien (Fälle) in vier Bundesländern (Brandenburg, Thüringen, Baden-Württemberg und Berlin) befragt. Ebenso wurde die für die jeweilige Schule zuständige Person in der Schulaufsichtsbehörde interviewt. Insgesamt fanden zwei Befragungswellen mit Abstand etwa eines Schuljahres statt, um bestimmte Sichtweisen und Praktiken an den Schulen bestätigen zu können. Als Schulart wurde das Gymnasium gewählt, weil es sich hierbei um die einzige zwischen Bundesländern vergleichbare Schulart handelt. Ein weiterer Grund ist, dass die meisten Gymnasien eine fest etablierte Fachschaftsstruktur vorweisen können. Die Fachschaften gelten als Ankerpunkt für die schulinterne Datennutzung.

Die Studie wurde in vier Bundesländern durchgeführt, um unterschiedliche Regelungen in Bezug auf VERA berücksichtigen zu können. Baden-Württemberg führte zum Zeitpunkt der Datenerhebung noch eigens entwickelte Diagnose- und Vergleichsarbeiten zu Beginn der Klassen 7 und 9 durch. In Berlin werden alle verfügbaren Tests seit dem Schuljahr 2010/2011 verbindlich durchgeführt. In Baden-Württemberg, Thüringen und Brandenburg gibt es verpflichtende und optionale Tests.

Auch die Auswertungs- und Rückmeldemodalitäten unterscheiden sich zwischen den vier Bundesländern. In Berlin und Brandenburg übernimmt das Institut für Schulqualität (ISQ) die Auswertung der Lernstandserhebungen. Zudem erarbeitet das ISQ eine ausführliche Rückmeldung zu den Ergebnissen. In Baden-Württemberg hingegen kamen bis zum Jahr 2014 selbstausswertende Excel-Tabellen mit automatisch generierten Rückmeldungen auf Basis einer Pilotierungsstichprobe zum Einsatz. Ebenso werden in Baden-Württemberg Niveaustufen in Form von Quartilsbereichen rückgemeldet und keine Kompetenzstufen im Sinne der KMK-Kompetenmodelle. In den anderen drei Bundesländern werden kriteriale und soziale Vergleichsdaten auf Basis der Landeswerte zur Verfügung gestellt. In Berlin und Thüringen werden auf Grundlage verschiedener Verfahren (vgl. Fiege et al. 2011) den Lehrkräften faire Vergleichsdaten zurückgemeldet.

Im Hinblick auf die Datenhoheit gibt es folgende Unterschiede: In Baden-Württemberg und Thüringen werden die Ergebnisse aus VERA nur den Schulen zurückgemeldet. Bei der externen Schulevaluation sollen die Schulen darstellen, wie sie mit den Ergebnissen umgehen. In Berlin und Brandenburg müssen die auf Schulebene aggregierten Ergebnisse auch an die Schulverwaltung gemeldet werden.

3.2 Stichprobe

Insgesamt wurden 229 Interviews mit schulischen Akteuren aus 19 Gymnasien (Fälle) und ihren zuständigen Schulaufsichtsbehörden geführt (fünf Schulen in Brandenburg und Baden-Württemberg sowie je vier Schulen in Berlin und Thüringen). Bei der Kontaktaufnahme mit den Schulen mussten letztendlich auch aus Gründen der Genehmigungspflicht die Schulaufsichtsbehörden mit einbezogen werden. Die finalen Auswahlkriterien waren ein Mindestmaß an organisatorischer Differenzierung (d. h. eine Schülerzahl von mind. 600 und/oder mind. dreizügige 8. Klassenstufe) sowie Erfahrungen mit Vergleichsarbeiten in der Klassenstufe 8. In Abstimmung mit den Schulaufsichtsbehörden wurden Gymnasien kontaktiert, bei denen in Bezug auf die VERA-Nutzung neben einem Minimum an Aktivität auch eine gewisse Auskunftsbereitschaft zu erwarten war. Es handelt sich also um eine hinsichtlich der VERA-Nutzung positiv selektierte Stichprobe, um Muster der Nutzung hervortreten zu lassen.

Die Schulen waren mehrheitlich (groß-)städtisch, wenige ländlich gelegen. Die Schulleitungen waren in etwas mehr als der Hälfte der Fälle weiblich besetzt. Die Schulleiterinnen und Schulleiter waren mindestens 1,5 Jahre bis 12 Jahre an den untersuchten Schulen in der Position der Schulleitung tätig. Die Einzugsgebiete der Schulen weisen vor allem in Berlin und teilweise in Baden-Württemberg eine sozial heterogene Struktur auf, in Brandenburg und Thüringen lassen sich diese als sozial homogen beschreiben.

3.3 Datenerhebung und Datenauswertung

Die Daten wurden mithilfe problemzentrierter, leitfadengestützter Einzelinterviews in zwei Erhebungszeitpunkten (1. Quartal Schuljahr 2010/2011; 2. Quartal Schuljahr 2011/2012) erhoben. Die Interviewleitfäden orientierten sich an dem *Phasenmodell der Nutzung von Ergebnissen aus Lernstandserhebungen* nach Helmke (2004). Entsprechend diesem Modell werden bei der Verarbeitung von Evaluationsergebnissen idealtypisch die Phasen Rezeption, Reflexion, Aktion und Evaluation durchlaufen. In den Interviews wurden neben Fragen zur Einstellung und Wissen über die Konzeption der Vergleichsarbeiten vor allem Fragen zur Verwendung der Ergebnisse, insbesondere der Testergebnisse, in der Abstimmung mit der jeweils übergeordneten Ebene gestellt. Zudem wurden Fragen nach Maßnahmen infolge der Ergebnismrückmeldung aus Vergleichsarbeiten und nach der Evaluation dieser Maßnahmen formuliert. Schulinterne und -externe Rahmenbedingungen, die die Nutzung der Vergleichsarbeiten und Ergebnismrückmeldungen begünstigen oder behindern, wurden ebenso erfragt und lassen sich theoretisch in das Modell von Visscher und Coe (2003) einordnen, das wichtige Einflussfaktoren auf eine Datennutzung im Rahmen von *School Performance Feedback Systems* darstellt. Dabei wurde gefragt, welche Einflüsse zwischen den Akteurebenen bezüglich der Nutzung der Rückmeldungen erfahren wurden.

Die Interviewdaten wurden alltagssprachlich transkribiert und inhaltsanalytisch aufbereitet. Eine Grundausswertung der Daten wurde online gestellt (vgl. Ramsteck et al. 2013). Die Themenkomplexe des Interviewleitfadens dienten als Grundlage für ein sowohl induktiv als auch deduktiv entwickeltes Kategoriensystem (vgl. Mayring 2010). Es zeigte sich, dass die Interviewdaten – auch aufgrund der Anzahl der Interviews – zu einer inhaltlichen Sättigung führten.

Um die Reliabilität der inhaltsanalytischen Auswertung im Rahmen des Verbundprojektes zu überprüfen, wurde Cohens Kappa (κ) berechnet. Cohens Kappa gibt den Grad der Übereinstimmung der Kodierung bei zwei Beurteilern an (Interkoderreliabilität). Die Interkoderreliabilitäten wurden für die Auswertung der Interviews zu den einzelnen Akteursebenen nach Bundesländern getrennt ermittelt. Pro Bundesland und Akteursebene wurden drei zufällig ausgewählte Interviews von zwei Mitarbeiterinnen und Mitarbeitern kodiert, um die Güte der entwickelten Kategoriensysteme zu überprüfen. Es fanden sich mehrheitlich gute bis sehr gute ($0.66 \leq \kappa \leq 0.83$) und mit einer Abweichung mindestens akzeptable ($0.44 \leq \kappa \leq 0.58$) Reliabilitätskoeffizienten. Das Kategoriensystem bezogen auf die Lehrkräfte stellte die Ausnahme dar. Es ist mit sieben Überkategorien, die sich jeweils in mehrere Unterkategorien bis zu sieben Ebenen tief gliedern, das komplexeste; das hat möglicherweise dazu geführt, dass die Codings nicht angemessen eingeschätzt und den Kategorien zugeordnet werden konnten. Es fand eine Präzisierung der Kategorien sowie der Kodierregeln für die anschließenden Kodiervorgänge statt.

Die Validität der Daten lässt sich wie folgt einschätzen: Das Risiko qualitativer Interviewdaten, in der Erhebung subjektiven Verzerrungen und Erinnerungseffekten ausgesetzt zu sein, lässt sich nicht vollständig kontrollieren. Im Gegensatz zu standardisierten Instrumenten werden Prozesse der schulinternen Kommunikation und Nutzung von VERA-Daten jedoch wesentlich konkreter erfasst. Gerade die alltagssprachlich formulierten Antworten der Lehrkräfte geben interessante Hinweise auf die Denk- und Handlungsmuster im Bezug auf VERA. Ein weiterer Aspekt, der für eine gewisse Validität der Daten spricht, ist die fallorientierte Herangehensweise. Aussagen einzelner Akteure können mit den Aussagen anderer Akteure an derselben Schule verglichen werden. Damit ist zumindest ansatzweise eine Unterscheidung zwischen tatsächlich erfolgten Aktivitäten und sozial erwünschten Darstellungen möglich.

4. Empirische Befunde

In diesem Bericht werden lediglich die Befunde der qualitativen Inhaltsanalyse in Bezug auf die Fragestellungen zusammenfassend dargestellt. Die inhaltsanalytische Auswertung und die genaue Beschreibung von Kategorien mit zugeordneten Fallbeispielen und beispielhaften Interviewziten kann in der im Internet veröffentlichten Grundausswertung nachgelesen werden (vgl. Ramsteck et al. 2013). Lediglich zur Veranschaulichung der Befunde werden in den nachfolgenden Abschnitten ausgewählte Interviewaussagen zitiert oder paraphrasiert.

4.1 Formen der instrumentellen Nutzung bei Lehrkräften und in Fachkonferenzen

Zur Beantwortung der Forschungsfrage 1 wurde auf Interviews mit Fachlehrkräften und Fachkonferenzleitungen zurückgegriffen. Die Aussagen zur Datennutzung waren erwartungsgemäß sehr heterogen. Ein Teil der befragten Lehrkräfte lehnte die Auseinandersetzung mit den VERA-Ergebnissen vollständig ab, weil beispielsweise die VERA-Tests nicht zu den eigenen Unterrichtsinhalten passten. In diesen Fällen konnte auch nicht über eine Nutzung berichtet werden. Insgesamt gelang es jedoch, verschiedene Nutzungsformen zu kategorisieren und zu beurteilen, ob eine instrumentelle Nutzung vorliegt und wie diese von den Lehrkräften begründet wird.

In den Interviews zeigte sich sehr deutlich, dass die in den didaktischen Handreichungen sehr ausführlich erläuterten Kompetenzstufen in den einzelnen Kompetenzbereichen (z. B. Lesen, Rechtschreibkompetenz etc.) praktisch keine Rolle spielen. Hier gibt es auch keine bundesländerspezifischen Unterschiede zwischen den Gymnasien. In einzelnen Fällen greifen Lehrkräfte die Bezeichnungen von Kompetenzbereichen auf (z. B. Hörverstehen als Teil der fremdsprachlichen Kompetenz). Es gibt allerdings auch in diesen Fällen keine Anzeichen für eine an den Kompetenzstufenbeschreibungen orientierte Interpretation der Schülerleistungsdaten.

Dagegen ließen sich sehr viele Beispiele für eine dichotome und zugleich an Inhalten orientierte Einordnung und Bewertung der Schülerleistungen nach folgenden Mustern finden: Die Klasse hat den Inhalt A verstanden vs. nicht verstanden; im Bereich B war die Klasse besser als im Bereich C. Eine weitere, sehr häufig zu beobachtende Interpretationsebene sind die Schülerleistungen bei einzelnen Testaufgaben. Ein über- oder unterdurchschnittliches Abschneiden bestimmter Schülergruppen bei einzelnen Testaufgaben wurde jedoch wiederum mit den im Curriculum verankerten Inhaltsbereichen verknüpft, nicht mit bestimmten Aspekten der Kompetenzmodellierung.

Die Interpretation der Leistungsdaten wirkt sich recht deutlich auf die Ableitung und Begründung der Maßnahmen aus. Auch bei den von den Lehrkräften genannten VERA-Konsequenzen findet man eine sehr große Bandbreite. Zunächst einmal gibt es eine Gruppe von Lehrkräften, die die VERA-Ergebnisse mit dem eigenen Unterricht nicht in Verbindung bringen. Entweder wird die Aussagekraft der Tests abgelehnt, oder es werden die Ergebnisse komplett mit den Lernanstrengungen und Begabungen der Schülerinnen und Schüler erklärt.

Didaktische Ideen und Maßnahmen im Anschluss an VERA, die der Definition von instrumenteller Nutzung entsprechen, wurden wie folgt kategorisiert (vgl. ausführlich bei Maier, Ramsteck und Frühwacht 2013):

- Am häufigsten wurden Übungen oder Wiederholungen als Folgemaßnahmen genannt und nach folgendem Muster begründet: Bei VERA hat meine Klasse in einem Bereich X (z. B. Kopfrechnen, Grammatik etc.) schlecht abgeschnitten. Aus diesem Grund wird das Thema wiederholt, oder es werden weitere Übungsaufgaben zur Verfügung gestellt.
- Einzelne Testaufgabenformate werden von Lehrkräften als innovativ wahrgenommen und verstärkt in den Unterricht integriert. Zum Beispiel wurden von

einer Lehrerin Testaufgaben aus dem Bereich Hörverstehen in Englisch in den Unterricht eingebunden, weil diese gut aufbereitet sind, als CD zur Verfügung stehen und Informationen nach dem Hören von den Schülerinnen und Schülern verarbeitet werden müssen.

- Die Ergebnisse bei VERA fielen unterdurchschnittlich aus, weil die Schülerinnen und Schüler die Aufgabenstellungen nicht genau genug gelesen hatten. Aus diesem Grund wird in Zukunft das genaue Lesen von Arbeitsanweisungen verstärkt geübt.
- Bei einzelnen Schülerinnen und Schülern wurden Defizite in einzelnen Bereichen bestätigt. Aus diesem Grund werden verfügbare Förderstunden darauf ausgerichtet. Allerdings waren die Aussagen in dieser Kategorie sehr häufig auch nur vage Andeutungen bzw. Absichtserklärungen, oder es werden mangelnde Ressourcen dafür verantwortlich gemacht, dass eine gezielte Förderung zwar notwendig wäre, jedoch nicht stattfinden kann.
- Einige Aussagen ließen eine fachdidaktische Argumentation erkennen. Eine Lehrkraft möchte mehr Veranschaulichungen im Mathematikunterricht nutzen, weil in den VERA-Testaufgaben verschiedene Darstellungsformen von Daten (in Diagrammen, Tabellen, Texten) miteinander verknüpft werden müssen und den Schülerinnen und Schülern dies schwerfiel. Eine weitere Lehrkraft wird das Thema Novelle, bei dem die Schülerinnen und Schüler schlecht abgeschnitten hatten, noch einmal aufgreifen und anhand dieser Textsorte exemplarisch Kriterien wiederholen, die Erzähltexten zugrunde liegen.

Diese Beispiele können als instrumentelle Nutzung bezeichnet werden, weil sich die Folgemaßnahme inhaltlich auf die Interpretation der VERA-Rückmeldung bezieht. Wir fanden jedoch keine einzige Aussage, die darauf hinwies, dass die VERA-Interpretation oder die Begründung der Folgemaßnahme etwas mit den in den didaktischen Handreichungen ausführlich beschriebenen Kompetenzstufen und Aufgabenanalysen zu tun gehabt hätte. Die Lehrkräfte greifen vor allem auf sehr alltagsbewährte Handlungstheorien zurück und argumentieren durchweg lehrplanorientiert und nicht kompetenzorientiert.

Ebenso gab es Hinweise, dass sich die fachdidaktisch begründeten Maßnahmen deutlich widersprechen können. Eine Lehrkraft möchte in Zukunft die einzelnen Grammatikthemen gesondert unterrichten und dabei kleinschrittig vorgehen, um die Regelkenntnis zu sichern. Eine andere Lehrkraft dagegen sah ebenfalls Defizite der eigenen Klasse im Bereich Grammatik, zog allerdings die Konsequenz, Grammatikthemen in Zukunft stärker miteinander zu verknüpfen.

4.2 Strukturierung des Umgangs mit VERA-Daten in Einzelschulen durch Schulleitung und Fachkonferenzen

Auf Schulebene fanden wir beispielhafte Muster von Kommunikationsstrukturen, die zum einen die instrumentelle Datennutzung förderten, zum anderen konnten wir auch Muster von Kommunikationsvorgängen identifizieren, die sich hinderlich auf die

instrumentelle Datennutzung auswirkten. Im Folgenden werden wir einige Fälle präsentieren, um diese Muster zu verdeutlichen und Forschungsfrage 2 zu beantworten:

Institutionalisierung der Datennutzung über eine Funktionsstelle

An einem Gymnasium des Bundeslandes B (Fall 11) wurde eine Oberstudienratsstelle speziell mit dem Schwerpunkt Vergleichsarbeiten eingerichtet (genannt: Koordinierungsstelle). Ein Motiv für die Einrichtung dieser Funktionsstelle war die Entlastung der Schulleitung von der Aufgabe der Koordination der Vergleichsarbeiten und der Maßnahmenableitung. Die Unterstützung bei der Auswertung wurde seitens des Lehrerkollegiums aktiv angenommen und geschätzt: „Hilfe habe ich von der Koordinierungsstelle erhalten. Sie hat uns wunderbar betreut bspw. bei der Frage, wie die VERA-Aufgaben aussehen, die Aufgabenformate. Da war eine große Unsicherheit“ (11L1A, 112–115)³. In der Fachkonferenz Deutsch wurde über die Klausurengestaltung gesprochen. Es wurde beispielsweise vereinbart, ein gemeinsames Diktat zu schreiben, bei dem Aspekte, die „in der Vergleichsarbeit zutage getreten sind“, berücksichtigt werden (L11L3A).

Die Schulleitung scheint sich trotz dieser Funktionsstelle nicht aus der Diskussion über die Datennutzung zu verabschieden. Sie wird in Bezug auf die Vergleichsarbeiten als „up to date“ bezeichnet (L11L3A). Als weitere instrumentelle Nutzung von VERA ist im Fall 11 laut Schulleitung eine Anpassung des Schulcurriculums denkbar (in Bundesland B können Schulen einen kleineren Teil der zu lernenden Inhalte selbst festlegen; dies wird Schulcurriculum genannt), wenn sich durch VERA „auffällige Lücken [zeigen], die bei allen Klassen einer Jahrgangsstufe vorhanden sind“ (11SL1A). Im Rahmen der Elternberatung spielt VERA ebenfalls eine Rolle. „Speziell bei meiner Klasse gibt es jetzt diese Woche einen Elternabend außer der Reihe. Weil es eben Handlungsbedarf gibt. Und man kann mit den VERA-Ergebnissen zeigen, dass Defizite da sind bei vielen“ (11FL1B, 91 ff.).

Förderung ebenenübergreifender Kooperationen und Kommunikationen mit Ansätzen einer schulweiten Verarbeitung der VERA-Ergebnisse

Im Fallbeispiel 22 zeigten sich in besonderer Weise förderliche strukturelle Voraussetzungen für die Datennutzung aus Vergleichsarbeiten. Hierbei stellte die Schulleitung angemessene Rahmenbedingungen bereit: Die Vergleichsarbeiten wurden als ein gesamtschulisches, alle Lehrkräfte betreffendes Thema verankert. Die Leistungsdaten wurden insbesondere als Argument für die Beantragung zusätzlicher Ressourcen bei der zuständigen Schulaufsicht sowie für die Stärkung von Gremienstrukturen herangezogen. Insbesondere die Fachkonferenz der in VERA getesteten Fächer ist dabei von zentraler Bedeutung; sie bekommt von der Schulleitung professionelle Verantwortung und Autonomie für die Auseinandersetzung mit den Vergleichsarbeiten übertragen (22SL1A). Gleichwohl bleibt die Schulleitung in diese Prozesse involviert, indem sie versucht, „auch von Zeit zu Zeit an Fachkonferenzen teilzunehmen. Aber ich denke, man kann das wirklich

3 Die Signatur der Interviewzitate verweist auf das Bundesland, die Schule, den befragten Akteur (L, FL, SL und SA) und den Erhebungszeitpunkt (A/B).

auch gut delegieren“, oder auch durch die Berichtslegung der Fachbereichsleitungen (z. B. durch Protokolle über stattgefundene Konferenzsitzungen). Darüber hinaus werden die VERA-Ergebnisse auf der Gesamtlehrerkonferenz vorgestellt. Auf diese Weise fördert die Schulleitung die Vernetzung der verschiedenen schulischen Ebenen und die Bedeutung dieses Instruments in der gesamten Schule (22SL1B). Hieraus resultieren auch auf Grundlage von Vergleichsarbeiten kollektive und verbindliche Entscheidungen, die das gesamte Lehrerkollegium betreffen. Dabei ist der Schulleitung zufolge „im Bewusstsein aller Lehrkräfte, ganz gleich welcher Fachrichtung, verankert, dass der Schwerpunkt der Schule auf Sprachförderung liegt“ (22SL1A).

Pflichtbewusster Umgang mit VERA ohne systematische Strukturierung der Datennutzung

Ein weiteres Muster im Umgang mit VERA zeichnet sich durch einen aktiven Austausch auf Schulebene über ausschließlich organisatorische Belange der Vergleichsarbeiten aus. Vor allem der Zeitaufwand für die Eingabe der Daten oder Mängel bei den Aufgabenstellungen werden kritisch diskutiert. Die Nutzung der Daten wird auf Schulebene dagegen nicht systematisch organisiert und geprüft. Fachbereichsleitungen konstatieren zudem, dass es keine Veränderung der Kommunikation durch die Einführung der Vergleichsarbeiten gab. In einigen Fällen suchte die Schulleitung zwar den Kontakt zu den Fachkonferenzen, meist jedoch nur dann, wenn schlechte Schülerleistungen bei VERA zurückgemeldet wurden.

Ein Beispiel hierfür ist der Fall 13. Die Aufgabe bezüglich der Vergleichsarbeiten wird an dieser Schule durch die Schulleitung „so definiert, dass wir natürlich die Diagnose durchführen. Wir haben auch alles schön abgelegt, abgeheftet.“ Eine Ableitung von Maßnahmen aufgrund der Feedbackdaten unterbleibt jedoch, da nach Ansicht der Schulleitung die Tests nicht reliabel sind. „Man geht immer vom Idealfall aus, dass VERA tatsächlich misst, was es messen soll. Dass es reliabel misst, dass der Schüler/die Schülerin motiviert ist, dass die Aufgaben auch wirklich von allen Schülern verstanden werden und so weiter und so fort. Und da die Voraussetzungen nicht gegeben sind, können auch die Schlussfolgerungen nicht gezogen werden. So ist unsere Ansicht der Schule“ (13SL2A, 135).

Im Rahmen von Fachkonferenzen und auch innerhalb des Lehrerkollegiums findet dagegen durchaus eine Besprechung der Ergebnisse statt. „Es ist fester Bestandteil, dass [die Daten] im Fachkonvent [Fachkonferenz] angesprochen werden. Aber eine vertiefte Weiterarbeit damit, wer macht jetzt was mit diesen Ergebnissen, das läuft dann sehr individuell“ (13FL1A, 57). Die einzelnen von der VERA-Durchführung betroffenen Lehrkräfte arbeiten mit den Daten z. T. instrumentell (Wiederholung bestimmter Aufgabenbereiche). Jedoch wurden auch zeitweise strategische Maßnahmen wie das *teaching to the test* umgesetzt: „Leider hatte sich das Unterrichten negativ verändert. Ganz am Anfang, weil man natürlich dachte: ‚Oh Gott, wie schrecklich.‘ Und ich habe plötzlich nur noch Grammatik gepaukt und Rechtschreibung in Klasse fünf und sechs. Und das war sogar fachschaftsweit der Fall, dass wir irgendwie angefangen haben, zu pauken“ (13FL1A, 139).

Keinerlei Auseinandersetzung mit VERA auf Schulebene, dennoch Formen der instrumentellen Nutzung auf Lehrerebene

Ein weiteres Muster ist eine komplette Verweigerungshaltung der Schulleitung, sich mit VERA auseinanderzusetzen. Es findet weder eine inhaltliche noch eine organisatorische Kommunikation über VERA statt. Diese wird von den Schulleiterinnen bzw. Schulleitern an die Fachkonferenz delegiert. Auch seitens der Fachbereichsleitungen besteht ein hohes Maß an Kritik und Vorbehalten gegenüber den Vergleichsarbeiten, vor allem strukturelle Mängel sowie den Aufwand (Zeit) und den daraus folgenden Nutzen (wenig Erkenntnisgewinn) betreffend.

Ein Beispiel hierfür ist der Fall 44. Eine Kommunikation zwischen Schulleitung und Lehrerkollegium bezüglich der VERA-Aufarbeitung fand nicht statt. Die Schulleitung delegierte diese an die Fachschaften. Die Fachschaftsleitungen förderten allerdings auch keinen Austausch zwischen den Lehrkräften: „Ich geh' da nicht schnüffeln. Das mache ich nicht, nein. Um Gottes Willen. Wir haben eigentlich genügend Verantwortungsbewusstsein, dass wir uns dieser Aufgabe widmen“ (44FL1A). Auf der Ebene der Lehrkräfte zeigte sich in diesem Fall, dass die VERA-Ergebnisse auch an dieser Schule durchaus instrumentell genutzt werden (v. a. Übernahme neuer Aufgabenformate). Doch bleibt dieses Lehrerhandeln isoliert und ist nicht in einen fachschafts- oder schulinternen Diskussionsprozess eingebunden: „Alles, was die Arbeit mit VERA angeht, läuft im Moment auf rein fachlicher Ebene ab. Also, das haben wir jetzt noch nicht in größere Konzepte mit eingebunden“ (44SL2B). VERA im Sinne der Schulentwicklung oder als Selbstevaluationsinstrument zu nutzen, „das halte ich für ein bisschen weit gegriffen“ (44SL2B).

4.3 Die Rolle der Schulaufsicht für die datenbasierte Schulentwicklung

Forschungsfrage 3 zielt darauf ab, den Umgang von Schulaufsichtsvertreterinnen und -vertretern in Bezug auf VERA zu analysieren. Wie handeln die von uns befragten Schulaufsichtsvertreterinnen und -vertreter in den jeweiligen Bundesländern? Im Folgenden werden unsere Ergebnisse aus den vier Bundesländern geschildert:

Bundesland A

In Bundesland A werden die Ergebnisse auf Schulumtsebene in einem Bericht zusammengestellt. Die Ergebnisse einzelner Schulen lassen sich daraus jedoch nicht ablesen. Dies beklagen die vier von uns befragten Schulaufsichtspersonen einhellig. Das Schulumt kann zwar von einzelnen Schulen Einsicht in die Ergebnisse fordern, was von einer der befragten Schulaufsichtspersonen (4SA1A) jedoch als „Umweg“ bezeichnet wird. Man hätte lieber die „direkte Linie“, d. h. einen Bericht mit „schulscharfen“ Ergebnissen. Es wird beklagt, dass dies vom Ministerium aber „nicht erwünscht“ ist. Um Schulen beurteilen zu können, greifen die Schulämter auf Daten zu Abschlussprüfungen zurück. Von Bedeutung ist in diesem Zusammenhang, dass die von uns befragten Schulleiterinnen und Schulleiter nicht ausreichend über die Datenhoheit der Schulen informiert sind. Ein von uns befragter Schulleiter vermu-

tet, dass die Schulleiter schulscharfe Daten erhalten könnten. Ein anderer Schulleiter sagt, dass die Schulleiter genau dieselben Ergebnisse erhalten wie einzelne Schulen.

Im Rahmen von Schulleiterdienstbesprechungen oder bei konkreten Schulbesuchen lässt sich eine der befragten Schulaufsichtspersonen von den Schulleiterinnen und Schulleitern oder Fachschaften über Kompetenztestergebnisse und abgeleitete Maßnahmen berichten. Interessant ist, dass in diesem Fall eine gehörige Portion Misstrauen mitschwingt: „Ob wir da immer bis ins letzte Detail die Wahrheit erfahren, wage ich zu bezweifeln“ (4SA1A). Die Kontrollfunktion wird dabei argumentativ mit Unterrichtsentwicklung verknüpft. Hätte man einen direkten Zugang zu den Schulergebnissen, könnte man „die Guten loben“ und „die Schlechten nicht tadeln, sondern auffordern, besser zu werden“.

In einem anderen Schulamtsbezirk hatten die Schulen die Testdaten nur schleppend vom zentralen Server abgerufen. Deshalb wurden regionale Konferenzen einberufen, auf denen Schulleiterinnen und Schulleiter über den Umgang mit den Testdaten berichten sollten. Durch den direkten Kontakt mit den Schulen kamen laut Schulaufsicht gewisse Unterstützungsmöglichkeiten in Gang. Der von uns befragte Schulleiter berichtet dagegen nichts von Unterstützungsmöglichkeiten, sondern lediglich von einem einmaligen Anruf des Schulamts, bei dem gefragt wurde, ob man den Bericht abgerufen habe. Weitere Unterstützungsmöglichkeiten werden lediglich als Ideen geäußert, z. B. eine Netzworkebildung zwischen Schulen fördern und anregen. Laut Aussagen aller befragten Schulaufsichtspersonen gibt es in Bundesland A keine gezielten Fortbildungen zur Datennutzung auf Schulamtssebene. Dies bestätigen auch die Schulleitungen. Die zentralen Fortbildungsangebote sind zwar bekannt, werden jedoch nicht genutzt.

Die Bedeutung der Kompetenztests bei Zielvereinbarungen mit den Schulen wird unterschiedlich gesehen. Einmal spielen die Tests keine Rolle bei konkreten Zielvereinbarungen mit Schulen. In einem anderen Schulaufsichtsbezirk sind sie jedoch eine wichtige Grundlage. Die dort genannten Beispiele sind allerdings sehr weit entfernt von VERA-Kompetenzen: z. B. Sozialkompetenztraining, Methodentraining. Auch aus Sicht der Schulleitungen spielen Kompetenztests bei Zielvereinbarungen im Rahmen der externen Evaluation keine Rolle, u. a. auch weil andere Leistungsdaten (Abiturnoten) zur Verfügung stehen.

Insgesamt wird die Zurückhaltung des Schulamts von den Schulleiterinnen und Schulleitern eher positiv bewertet und als Vertrauensbeweis gegenüber den Schulen gesehen.

Bundesland B

Die Vergleichsarbeiten werden von den zwei in Bundesland B befragten Schulaufsichtspersonen als wichtiges Diagnoseinstrument für die Schul- und Unterrichtsentwicklung betrachtet. Wenn es jedoch um die konkrete Arbeit mit Vergleichsarbeiten geht, sieht sich die Schulaufsicht „außen vor“ (1SA1A). Sie geht diesbezüglich nicht aktiv auf die Gymnasien zu. Alle Anfragen zu den Vergleichsarbeiten werden ans Landesinstitut des Bundeslandes B weitergeleitet. Bei großen Schulleitertagungen, an denen die Schulaufsicht beteiligt war, wurde die Thematik einmal aufgegriffen.

Es wurde ein Referent vom Landesinstitut zur Frage des Umgangs mit Vergleichsarbeiten eingeladen. Weitere Fortbildungs- und Unterstützungsangebote gibt es nicht. Schulen könnten jedoch Fachberaterinnen und Fachberater auch anfordern, wenn Vergleichsarbeiten hierfür Anlass geben. Im ganzen Bezirk war dies für den gymnasialen Bereich jedoch noch nicht der Fall.

Ebenso spielen die Vergleichsarbeiten keine Rolle bei Schulbesuchen oder bei der Fremdevaluation: „Also bei allen Fremdevaluationsberichten, die ich bisher gelesen habe, sind die Testergebnisse speziell der Vergleichsarbeiten eigentlich nicht erwähnt“ (1SA1A). Der Schulaufsichtsvertreter weist in diesem Zusammenhang darauf hin, dass die Schulen sehr viel im Vorfeld organisieren müssten, um die Vergleichsarbeiten gewinnbringend einsetzen zu können. Beispielsweise müssten Lehrkräfte über mehrere Jahrgangsstufen hinweg dieselben Klassen unterrichten. Nur so könnte man die Ergebnisse mit dem Unterricht in Verbindung bringen. Aufgrund vieler Reformbaustellen haben die Schulen hierfür jedoch keinerlei Kapazitäten.

Die in Bundesland B befragten Schulleiterinnen und Schulleiter wunderten sich zum Teil über die Frage nach der Rolle des Schulamts beim Umgang mit den Vergleichsarbeiten. „Kontakt mit dem Schulamt gibt es diesbezüglich nicht“ (11SL1A). Das war „kein Hauptthema“ bisher (13SL1A). Die Schulaufsicht war aus Sicht eines Schulleiters „in den letzten Jahren vor allem mit sich selbst beschäftigt“ (11SL1A). Dies wird allerdings positiv gewertet, weil dann die Schulen selbstständiger agieren können. Wenn die Schulaufsicht wahrnimmt, dass im Bereich Schulentwicklung „viel läuft“, hält sie sich heraus. Eine Schulleitungsperson sieht die Trennung von Landesinstitut und Schulaufsicht als Problem an. Bei einer Schulleitertagung war ein Referent des Landesinstituts anwesend, und es wurden organisatorische Probleme bei der Durchführung der Vergleichsarbeiten angesprochen. Es entstand jedoch der Eindruck, dass dies nicht gehört wird. Die Schulaufsicht wird deshalb in Bezug auf Vergleichsarbeiten gegenüber dem Landesinstitut als machtlos angesehen.

Bundesland C

In Bundesland C wurden Schulaufsichtspersonen aus zwei verschiedenen Bezirken sowie zusätzlich in einem Bezirk die Leitung für schulische Qualitätssicherung interviewt. Über alle Fälle hinweg besteht ein unklares Rollenverständnis in den Schulaufsichten, das zwischen Kontrolle und Unterstützung variiert und als Grund für die mangelnde Unterstützung aufseiten der Schulaufsichten gesehen werden kann. Damit geht größtenteils eine Übertragung der Verantwortung an die Schulen einher. Die Schulaufsichtspersonen externalisieren ihre Verantwortung für die Vergleichsarbeiten in beiden Bezirken an die Schulen – mit der Folge, dass diese auf Einzelschulebene i. d. R. wenig Beachtung finden und genutzt werden. In Einzelfällen werden VERA-Ergebnisse als Grundlage für das Beschließen von Zielvereinbarungen verwendet (2SA1A, 14; 2SA1B, 49; 2SA2A, 11).

Insgesamt nimmt die Schulaufsicht des Bundeslandes C auch ihre Unterstützungsfunktion kaum wahr bzw. gelingt die Unterstützung durch die Schulaufsicht nur bedingt (2SA2A, 19). Die Schulaufsicht eines Bezirks im Bundesland C räumt hierbei selbst ein, dass die Schulen im Umgang mit VERA noch mehr Hilfen erhal-

ten müssten, „wie man dann mit den Ergebnissen sinnvoll umgehen kann. Ja, also da fühlen sich die Schulen dann schon manchmal auch ein bisschen alleingelassen, auch vom Landesinstitut alleingelassen“ (2SA2A, 105). Dies wird jedoch von den anderen schulischen Akteuren unterschiedlich bewertet. Diese erwarten in vielen Fällen auch keine konkreten Unterstützungssysteme oder sehen keinen Handlungsbedarf seitens der Schulaufsicht aufgrund guter VERA-Ergebnisse (2SA1A, 16; 2SA2A, 3). In den Fällen, in denen Unterstützung der Schulen von der Schulaufsicht gewünscht wird, werden Fortbildungen zum Umgang mit VERA oder zusätzliche Ressourcen der schulischen Akteure genannt.

Als Gründe, die dazu führen, dass VERA 8 insgesamt vernachlässigt wird, wurden auf Schulaufsichtsebene „wenig Zeit für VERA“ (2SA2A, 63; 2SA2B 76–77), eine „zu späte Rückmeldung der Ergebnisse“, „stabile Ergebnisse“ (2SA2A, 27) und mangelndes Wissen über datengestützte Unterrichtsentwicklung (2SA2A, 65) genannt. Folgende Faktoren hemmen aus Sicht der Schulaufsicht die instrumentelle Datennutzung auf Einzelschulebene (2SA2B): Der lange Rückmeldezeitraum, das schwere Level der VERA-Tests, zu gute Ergebnisse, kein Nutzen der Vergleichsarbeiten für die Schulentwicklung sowie wechselnde Schwerpunkte innerhalb der Tests.

Bundesland D

In Bundesland D wurden Interviews mit zwei Schulaufsichtspersonen geführt. VERA wird seitens der Schulaufsicht des Bundeslandes D einhellig in erster Linie als ein Instrument für die Schulen gesehen: „Das ist ein Instrument für die Schulen vordergründig und natürlich gegebenenfalls dann in den Diskussionen innerhalb Schulaufsicht und Schule und natürlich auch innerhalb von Schulaufsicht zu gucken, wo dann gegebenenfalls dann noch mal Einflussnotwendigkeiten sich ergeben“ (3SA1B, 6). Dies korrespondiert mit der Einschätzung der innerschulischen Akteure. Diese berichten in zwei Fällen, dass die Schulaufsicht „andere Sorgen“ habe, als sich um eine Kommunikation mit den Schulen zu VERA-Ergebnissen zu bemühen (Fälle 34 und 35). Der geringe Stellenwert von VERA zeigt sich auch darin, dass diese insgesamt nicht als Kontrollinstrument in Bundesland D auf Schulaufsichtsebene wahrgenommen werden. Bei den unterrichts- und schulbezogenen Prozessen, die auf Grundlage von VERA 8 resultieren, erfahren drei der fünf untersuchten Schulen in Bundesland D Unterstützung durch die Schulaufsicht. Diese war in den jeweiligen Fällen unterschiedlich stark ausgeprägt.

In Bundesland D werden hierbei in den meisten Fällen Fortbildungen wahrgenommen (wobei auch VERA-Informationsveranstaltungen dazu gezählt werden) und z. T. auch Fortbildungsmaßnahmen seitens der Schulaufsicht angeboten bzw. (mit-)entwickelt (3SA1B, 15, 21, 44). In diesem Zusammenhang erklärt eine Schulaufsicht ihre Funktion folgendermaßen: „Also, es geht eher weniger darum, Konsequenzen anzudrohen, als dann eher den pädagogischen Prozess in der Schule zu begleiten und zu steuern, um zu sagen, dass die Schule hier, in den einzelnen Fächern, wo ggf. negative Ergebnisse erreicht worden sind, in den Fachkonferenzen mit dem einzelnen Lehrer/-innen darüber zu diskutieren, was wie verändert werden kann. Also, ob nun inhaltlich, methodisch, didaktisch, fachlich, hängt dann immer vom konkreten Einzelfall ab“ (3SA1A, 20).

Die Schulaufsicht schätzt in einigen Fällen (34, 35) VERA 8 als einen Teil der Selbstevaluation ein (3SA2B, 63). Hierbei kann der schulische Output im Vergleich zu anderen Schulen und zum mittleren Landesergebnis in Bezug gesetzt werden. In einem Fall (31) soll ein Vertreter der Schulverwaltung eingeladen werden, um Lehrkräften die Möglichkeit eines Meinungsaustauschs über VERA zu bieten. Generell lassen sich jedoch keinerlei Routinen in Bezug auf eine Rezeption und Nutzung von VERA-Ergebnissen seitens der Schulaufsicht mit den Schulen feststellen.

Die Kommunikation der Schulaufsicht aus Bundesland D mit den Schulen ist in allen Fällen noch immer stark administrativer Natur. Generell gibt es keinen systematischen Austausch seitens der Schulaufsicht mit den Schulen über VERA (SL1B, 61–63; 3SA4A, 14; 3SA2B, 22). VERA-Ergebnisse werden lediglich in einigen Fällen bei datengestützten Qualitätsgesprächen und in den Schulleiterdienstberatungen mit der Schulleitung thematisiert (3SA1A, 6).

In fast allen Fällen des Bundeslandes D werden auf Grundlage von VERA keine Zielvereinbarungen mit der Schulaufsicht beschlossen; diese können aber grundsätzlich seitens der Schulaufsicht bei schlechten Ergebnissen in VERA vereinbart werden.

5. Diskussion der Befunde

Limitation der Studie

In dieser Studie wurden die instrumentelle Datennutzung sowie damit zusammenhängende Aspekte auf Ebene der Einzelschule und der Schulaufsicht mit qualitativen Interviewdaten beschrieben. Die Aussagekraft der Befunde ist aufgrund der Stichprobe und der Datenerhebung als begrenzt zu sehen. Es können lediglich Aussagen zu einer Schulart (Gymnasium) gemacht werden. Ebenso wurden die Interviews in nur vier Bundesländern durchgeführt, sodass auch nur begrenzt etwas über die Bedeutung länderspezifischer Regelungen für die schulinterne Datennutzung ausgesagt werden kann. Andererseits war es möglich, aufgrund der Befragung möglichst vieler Akteure innerhalb einer Schule ein zuverlässiges Bild der Datennutzung zu zeichnen und auch zu validieren. Ebenso können aufgrund der qualitativen Aussagen das Niveau und Varianten der instrumentellen Datennutzung recht gut abgeschätzt werden.

Einordnung der Befunde

Forschungsfrage 1: Die Studie zeigt, dass auf Ebene der Lehrkräfte und der Fachkonferenzen eine instrumentelle Nutzung stattfinden kann. Es gibt jedoch keine Hinweise auf eine Ableitung von Maßnahmen, die sich an den in den didaktischen Handreichungen beschriebenen Kompetenzmodellen und Kompetenzstufen orientieren. Die Nutzungsformen und die Begründungsmuster sind überwiegend an Lehrplänen orientiert (nicht kompetenzorientiert) und folgen in der Regel sehr einfachen Begründungsschemata. Eine Anbindung an aktuelle fachdidaktische Diskussionen ist nicht erkennbar.

Dieser Befund stellt zumindest die in einigen Studien berichtete „kriteriale Interpretation“ der Testdaten infrage (vgl. Kühle & Peek 2007; Kühle 2010; Groß Ophoff

2013b). Dies könnte daran liegen, dass in diesen quantitativen Fragebogenstudien Lehrkräfte bei Aussagen zur kriterialen Bezugsnorm aufgrund sozialer Erwünschtheit ankreuzen, da die entsprechenden Items an die Reformsemantik der Vergleichsarbeiten anknüpfen. Andererseits muss man natürlich auch die Validität der Interviewdaten dieser Studie hinterfragen. Aufgrund von Erinnerungseffekten könnte es durchaus sein, dass Lehrkräfte im Interview sehr pauschal über ihre Rezeption und Nutzung der Daten berichten, eventuell aber bei der Lektüre des Ergebnisberichts doch die eine oder andere Kompetenzstufenbeschreibung gelesen und in Verbindung mit den Schülerdaten gebracht haben. Allerdings haben wir bei mehreren Hundert Lehrerinterviews keinen einzigen Hinweis erhalten, der sich auf die Kompetenzmodellierung und die Vorschläge in den didaktischen Handreichungen bezogen hätte.

Forschungsfrage 2: In der Studie konnten auf Einzelschulebene bestimmte Muster der erfolgreichen und weniger erfolgreichen Nutzung von Testdaten herausgearbeitet werden. Dabei zeigte sich, dass eine systematisch geförderte Kommunikation über VERA nicht unbedingt eine zwingende Voraussetzung für die instrumentelle Datennutzung auf Ebene der Lehrkräfte sein muss, wie dies in Modellen häufig suggeriert wird (vgl. Visscher & Coe 2003; Helmke 2004; Groß Ophoff 2013a). Auch Lehrkräfte an Gymnasien mit einer bewussten Verweigerungshaltung gegenüber VERA auf Ebene der Fachkonferenzen und der Schulleitung berichten über Maßnahmen im Anschluss an VERA. In einigen wenigen Fällen konnten wir jedoch eine strukturierte und von allen Akteuren getragene Form der Kommunikation über Vergleichsarbeiten beschreiben. Der Fall 11 mit der Besetzung einer Funktionsstelle für die datenbasierte Schulentwicklung muss jedoch als Einzelfall betrachtet werden. In den Positivbeispielen konnten wir allerdings auch kein besonders hohes Niveau der instrumentellen Datennutzung feststellen.

Forschungsfrage 3: Bei den Schulaufsichtspersonen gab es deutliche Bundesländerunterschiede in Bezug auf die Sichtweise der Datennutzung an Schulen. Die in den Bundesländern B, C und D geführten Interviews zeigen eine zurückhaltende und die Autonomie der Schule betonende Schulaufsicht. In Bundesland A kam dagegen der Eindruck auf, dass die Schulaufsicht die Datenhoheit der Schulen nicht befürwortet und man eine Unterstützung der Schulen nur durch verstärkte Kontrolle auch mit Testdaten erzielen kann. Diese Differenz lässt sich aber nur teilweise mit den verschiedenen Regelungsstrukturen zu Vergleichsarbeiten erklären. In Bundesland B ist die Haltung plausibel, weil die Schulaufsicht allenfalls über die externe Evaluation etwas über den Umgang mit Testdaten erfahren könnte. In Bundesland C und D liegen der Schulaufsicht allerdings auf Schulebene aggregierte Daten vor, was aber nicht zu gesteigerten Kontroll- oder Unterstützungsbemühungen führt. In Bundesland A scheint dagegen eine bestimmte Auffassung vom Amt der Schulaufsicht dafür verantwortlich zu sein, dass man gern verstärkten Zugriff auf die Daten hat, um so die Schulen besser voranbringen zu können.

Die Befunde zu Forschungsfrage 3 lassen sich zumindest ansatzweise auch in die internationale Forschung zu datenbasierter Schulentwicklung einordnen. Die diffuse Rolle der Schulaufsicht in Deutschland in Bezug auf Dateneinsicht und Unterstützung sowie die damit korrespondierende Verunsicherung der Schulen sind keine günstigen Voraussetzungen für eine systematische Etablierung einer datenbasierten

Schulentwicklung an Einzelschulen. Gerade die in internationalen Studien genannten Aspekte einer Erfolg versprechenden Kooperation zwischen Schulaufsicht und Schulen (vgl. Yang et al. 1999; Hayes & Rutt 1999; Demie 2003; Peng et al. 2006; Opfer, Henry & Masburn 2008) wurden in den Interviews immer wieder als Problem angesprochen. Ein Beispiel ist Vertrauen als Basis für die Zusammenarbeit zwischen Schulen und Schulaufsicht. In den von uns untersuchten Fällen herrschte dagegen eher Misstrauen, Desinformiertheit oder gegenseitiges Desinteresse. Ein weiteres Beispiel ist die Bildung von regionalen Netzwerken der datenbasierten Schulentwicklung. In den Interviews wurde zwar immer wieder von bestimmten Fortbildungsinitiativen oder Schulleiterdienstbesprechungen berichtet, jedoch hatte keine dieser Maßnahmen Bestand bzw. wurde als unterstützend wahrgenommen.

Bildungspraktische Implikationen

Aus den empirischen Ergebnissen der Untersuchung können nicht unmittelbar handlungspraktische Implikationen für die einzelnen Akteursgruppen oder die Bildungspolitik abgeleitet werden. Bei der positiv selektierten Stichprobe der Schulen ist zudem Vorsicht bei verallgemeinernden Schlussfolgerungen geboten. Die im Folgenden skizzierten praktischen Implikationen sind als Hinweise auf Gelingensbedingungen und Risikofaktoren bezüglich der Nutzung von Vergleichsarbeiten zu verstehen.

Die innerschulische Akzeptanz und Nutzung von Vergleichsarbeiten kann durch die Schulleitung gefördert werden, indem sie vor allem ausreichende Ressourcen für den intensiven Austausch über die VERA-Ergebnisse sowie deren schul- und unterrichtsbezogene Verarbeitung bereitstellt. Dabei ist die Stärkung der Autonomie der Fachkonferenzen und anderen wichtigen Organisationseinheiten (Gremien) erforderlich, da vor allem innerhalb der Fachkonferenzen VERA-Ergebnisse i. d. R. diskutiert und in diesem Rahmen Maßnahmen abgeleitet werden. Die regelmäßige Teilnahme der Schulleitung an diesen Gremien oder das Einfordern der Berichtslegung über den Verlauf der Konferenzsitzungen erhöht die Verbindlichkeit der Auseinandersetzung mit VERA 8. Darüber hinaus können eine etablierte (positive) Feedbackkultur in der Schule sowie das Zurückgreifen auf bzw. Einfordern von externen Unterstützungssystemen (intensivere Kooperation mit der Schulaufsicht oder den Landesinstituten oder das Heranziehen einer Evaluationsberaterin oder eines Evaluationsberaters) die praxiswirksame Nutzung von VERA auf Einzelschulebene befördern (vgl. Muslic 2014).

Auf Ebene der Schulaufsicht sollte vor allem eine Zielklärung bezüglich VERA erfolgen. Die Schulaufsicht sollte ihren Umgang mit den Daten gegenüber den Schulen sehr klar kommunizieren. Wenn Maßnahmen zur Unterstützung der Datennutzung angeboten werden, sollten diese mit bestehenden Fortbildungen bzw. Personalentwicklungsmaßnahmen sinnvoll gekoppelt sein. Beispielsweise könnte die Nutzung von VERA-Daten ein festes Thema in Schulleiterfortbildungen sein. Ebenso wäre zu prüfen, inwiefern regionale Lehrerfortbildungen zu fachdidaktischen Themen auf VERA-Kompetenzmodellierungen zurückgreifen.

6. Weiterführende Literatur zu den Projektergebnissen

- Kuper, H. & Muslic, B. (2012). Evidenzbasierte Steuerung. Der Umgang mit Rückmeldungen aus Lernstandserhebungen als Instrument einer evidenzbasierten Schulentwicklung. In Hornberg, S. & Parreira do Amaral (Hrsg.), *Deregulierung im Bildungssystem*, S. 151–164. Münster: Waxmann.
- Maier, U., Ramsteck, C. & Frühwacht, A. (2013). Lehr-lerntheoretische Argumentationsmuster bei der Interpretation und Nutzung von Vergleichsarbeitsrückmeldungen durch Gymnasiallehrkräfte. *Die Deutsche Schule – Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis*, Beiheft 12, S. 74–96.
- Muslic, B. (2014). Wie können Schulleiterinnen und Schulleiter die Akzeptanz und Nutzung von Lernstandserhebungen in den Schulen erhöhen? *Zeitschrift für Schulverwaltung*. Carl Link Verlag, S. 115–117.
- Muslic, B., Ramsteck, C. & Kuper, H. (2013). Das Verhältnis von Schulleitung und Schulaufsicht im Kontext testbasierter Schulreform. *Die Deutsche Schule – Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis*, Beiheft 12, S. 97–120.
- Ramsteck, C., Maier, U., Graf, T., Muslic, B. & Kuper, H. (2013). *Die Realisierung testbasierter Schulreform in der Mehrebenenstruktur des Schulsystems – Kategoriale Grundauswertung, Fallbeschreibungen und Fallanalysen*. Abgerufen unter http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/weiterbildung_bildungsmanagement/media/downloads/Die-Realisierung-testbasierter-Schulreform-in-der-Mehrebenenstruktur-des-Schulsystems_-Kategoriale-Grundauswertung_-Fallbeschreibungen-und-Fallanalysen_.pdf am 23.03.2015.
- Ramsteck C. & Maier U. (2015). Testdatenbasierte Schul- und Unterrichtsentwicklung – Analyse von Handlungsmustern bei der Rezeption und Nutzung von Vergleichsarbeitsdaten. In Schrader, J., Schmid, J., Amos, K. & Thiel, A. (Hrsg.), *Governance von Bildung im Wandel. Interdisziplinäre Zugänge*, S. 119–144. Wiesbaden: Springer Verlag für Sozialwissenschaften.
- Ramsteck, C., Muslic, B., Graf, T., Maier, U. & Kuper, H. (2015). Data-Based School Improvement: The Role of Principals and School Supervisory Authorities within the Context of Low-Stakes Mandatory Proficiency Testing in four German states. In *International Journal of Educational Management (Special Issue): School leadership in Germany between low stakes testing and high expectations* 29, 6.

7. Literaturverzeichnis

- Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher* 36, H. 5, S. 258–267.
- Bach, A., Wurster, S., Thillmann, K., Pant, H. A. & Thiel, F. (2014). Vergleichsarbeiten und schulische Personalentwicklung – Ausmaß und Voraussetzungen der Datennutzung. *Zeitschrift für Erziehungswissenschaft* 17, 1, S. 61–84.
- Bonsen, M., Büchter, A. & Peek, R. (2006). Datengestützte Schul- und Unterrichtsentwicklung: Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen

- und Lehrer. In Bos, W., Holtappels, H. G., Pfeiffer, H., Rolff, H.-G. & Schulz-Zander, R. (Hrsg.), *Jahrbuch der Schulentwicklung* 14, S. 125–148. Weinheim: Beltz.
- Butzke-Rudzynski, M. & Bachorz, K. (2013). Wie kann Schulaufsicht die Nutzung von Lernstandserhebungen unterstützen? Kooperation und Professionalisierung in moderierten Netzwerken von Fachkonferenzen. *Pädagogik* 65, H. 12, S. 44–47.
- Carlson, D., Borman, G. D. & Robinson, M. (2011). A Multistate District-Level Cluster Randomized Trial of the Impact of Data-Driven Reform on Reading and Mathematics Achievement. *Educational Evaluation and Policy Analysis* 33, H. 3, S. 378–398.
- Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study. *Journal of Classroom Interaction* 38, H. 1, S. 1–10.
- Cheng, L. & Curtis, A. (2004). Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In Cheng, L., Watanabe, Y. & Curtis, A. (Hrsg.), *Washback in Language Testing. Research Contexts and Methods* 3, H. 17. Mahwah/London: Lawrence Erlbaum.
- Demie, F. (2003). Using Value-added Data for School Self-evaluation: A case study of practice in inner-city schools. *School Leadership & Management* 23, H. 4, S. 445–467.
- Diemer, T. (2013). *Innerschulische Wirklichkeiten neuer Steuerung. Zur Nutzung zentraler Lernstandserhebungen*. Berlin: Springer VS.
- Fiege, C., Reuther, F. & Nachtigall, C. (2011). Faire Vergleiche? – Berücksichtigung von Kontextbedingungen des Lernens beim Vergleich von Testergebnissen aus deutschen Vergleichsarbeiten. *Zeitschrift für Bildungsforschung*, 1, H. 2, S. 133–149.
- Firestone, W. A., Winter, J. & Fitz, J. (2000). Different assessments, common practice? Mathematics testing and teaching in the USA and England and Wales. *Assessment in Education* 7, H. 1, S. 13–37.
- Frühwacht, A. (2012). *Bildungsstandards in der Grundschule. Bildungsstandards und Vergleichsarbeiten aus der Sicht von deutschen und finnischen Lehrkräften*. Bad Heilbrunn: Klinkhardt.
- Graf, T., Emmrich, R., Harych, P. & Brunner, M. (2013). Durchführungseffekte bei Vergleichsarbeiten in Jahrgangsstufe 8. *Empirische Pädagogik*, 27, H. 4, S. 459–473.
- Groß Ophoff, J. (2013a). *Lernstandserhebungen: Reflexion und Nutzung*, Waxmann, Münster.
- Groß Ophoff, J. (2013b). Der Effekt der Bezugsnormorientierung auf die Reflexion und Nutzung von Rückmeldungen aus Vergleichsarbeiten. *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. *Empirische Pädagogik* 21, H. 4, S. 411–427.
- Groß Ophoff, J., Koch, U., Hosenfeld, I. & Helmke, A. (2006). Ergebnisrückmeldung und ihre Rezeption im Projekt VERA. In Kuper, H. & Schneewind, J. (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen*, S. 19–40. Münster: Waxmann.
- Harris, A., Jamison, I. & Russ, J. (1997). A study of „effective“ departments in secondary schools. In Harris, A., Bennett, N. & Preedy, M. (Hrsg.), *Organizational Effectiveness and Management in Schools*, S. 147–161. Buckingham: Open University Press.

- Hayes, S. G. & Rutt, S. (1999). Primary analysis for secondary schools: An LEA research officer's perspective on helping secondary schools interpret assessment data for school improvement purposes. *Improving Schools* 2, H. 2, S. 44–52.
- Helmke, A. (2004). Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule, *Seminar 10*, H. 2, S. 90–112.
- Herman, J. L. (2004). The Effects of Testing on Instruction. In Fuhrman, S. H. & Elmore, R.F. (Hrsg.), *Redesigning Accountability Systems for Education*, S. 141–166. New York, London: Teachers College Press.
- Jäger, S. (2012). *Rezeption und Nutzung von Vergleichsarbeiten durch Lehrkräfte – eine qualitative Interviewstudie*. Dissertation. Schwäbisch Gmünd: Pädagogische Hochschule Schwäbisch Gmünd.
- Koch, U. (2011). *Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnismeldungen aus Vergleichsarbeiten*. Münster: Waxmann.
- Koch, U. & Hosenfeld, I. (2013). Wie objektiv werden Leseverstehensaufgaben im Rahmen der Vergleichsarbeiten in der Grundschule ausgewertet? *Empirische Pädagogik* 27, H. 4, S. 474–496.
- Kühle, B. (2010). *Zentrale Lernstandserhebungen - Ergebnisorientierte Unterrichtsentwicklung? Schulische Strategien beim Umgang mit Ergebnissen aus den Schulrückmeldungen im Kontext der ersten Lernstandserhebungen 2004/2005 in Nordrhein-Westfalen*. Berlin: Köster.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen. *Empirische Pädagogik* 21, H. 4, S. 428–447.
- Kultusministerkonferenz (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Link Luchterhand.
- Kultusministerkonferenz (2012). *Beschluss zur Weiterentwicklung von VERA. Beschluss der Kultusministerkonferenz vom 8.3.2012*.
- Kuper, H. & Hartung, V. (2007). Überzeugungen zur Verwendung des Wissens aus Lernstandserhebungen. *Zeitschrift für Erziehungswissenschaft* 10, H. 2, S. 214–229.
- Lane, S., Parke, C. S. & Stone, C. A. (2002). The Impact of a State Performance-Based Assessment and Accountability Program on Mathematics Instruction and Student Learning: Evidence from Survey Data and School Performance. *Educational Assessment* 8, H. 4, S. 279–315.
- Lee, M., Louis, K. S. & Anderson, S. (2012). Local Education Authorities and Student Learning: The effects of policies and practices. *School Effectiveness and School Improvement* 23, H. 2, S. 133–158.
- Levin, J. A. & Datnow, A. (2012). The Principal Role in Data-Driven Decision Making: Using case-study data to develop multi-mediator models of educational reform. *School Effectiveness and School Improvement* 23, H. 2, S. 179–201.
- Louis, K. S., Febey, K. & Schroeder, R. (2005). State-Mandated Accountability in High Schools: Teachers' Interpretations of a New Era. *Educational Evaluation and Policy Analysis* 27, H. 2, S. 177–204.
- Luxia, Q. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes writing test in China. *Assessment in Education* 14, H. 1, S. 51–74.

- Maier, U. (2009). *Wie gehen Lehrerinnen und Lehrer mit Vergleichsarbeiten um? Eine Studie zu testbasierten Schulreformen in Baden-Württemberg und Thüringen*. Hohengehren: Schneider-Verlag.
- Maier, U. (2010). Effekte testbasierter Rechenschaftslegung auf Schule und Unterricht: Ist die internationale Befundlage auf Vergleichsarbeiten im deutschsprachigen Raum übertragbar? *Zeitschrift für Pädagogik* 56, H. 1, S. 112–128.
- Maier, U. & Kuper, H. (2012). Vergleichsarbeiten als Instrument der Qualitätsentwicklung an Schulen. Überblick zum Forschungsstand, *Die Deutsche Schule* 104, H. 1, S. 88–99.
- Maier, U., Metz, K., Bohl, T., Kleinknecht, M. & Schymala, 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*, S. 197–224. Wiesbaden: Springer VS.
- Maier, U., Ramsteck, C. & Frühwacht, A. (2013). Lehr-lerntheoretische Argumentationsmuster bei der Interpretation und Nutzung von Vergleichsarbeitsrückmeldungen durch Gymnasiallehrkräfte. In Ackeren, I. v., Heinrich, M. & Thiel, F. (Hrsg.): *Lehr-lerntheoretische Argumentationsmuster bei der Interpretation und Nutzung von Vergleichsarbeitsrückmeldungen durch Gymnasiallehrkräfte*. Münster, *Die Deutsche Schule*, Beiheft 12/2013, S. 74–96.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- Muslic, B. (2014). Wie können Schulleiterinnen und Schulleiter die Akzeptanz und Nutzung von Lernstandserhebungen in den Schulen erhöhen? *Zeitschrift für Schulverwaltung*. Carl Link Verlag, S. 115–117.
- Nachtigall, C. (Hrsg.) (2005). *Landesbericht – Thüringer Kompetenztest 2005*. Jena: Friedrich-Schiller-Universität.
- Nachtigall, C. & Jantowski, A. (2007). Die Thüringer Kompetenztests unter besonderer Berücksichtigung der Evaluationsergebnisse zum Rezeptionsverhalten. *Empirische Pädagogik* 21, H. 4, S. 401–410.
- Nachtigall, C. & Hellrung, K. (2013). Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. *Empirische Pädagogik* 27, H. 4, S. 423–441.
- Opfer, V. D., Henry, G. T. & Mashburn A. J. (2008). The district effect: Systemic responses to highstakes accountability policies in six southern states. *American Journal of Education* 114, H. 2, S. 299–332.
- Parke, C. S. Lane, S. & Stone, C. A. (2006). Impact of a State Performance Assessment Program in Reading and Writing. *Educational Research and Evaluation* 12, H. 3, S. 239–269.
- Peng, W. J., Thomas, S. M., Yang, X. & Li, J. (2006). Developing school evaluation methods to improve the quality of schooling in China: a pilot 'value added' study. *Assessment in Education* 13, H. 2, S. 135–154.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, 68, S. 679–682.
- Ramsteck, C., Muslic, B., Graf, T., Maier, U. & Kuper, H. (2015). Data-Based School Improvement: The Role of Principals and School Supervisory Authorities within the Context of Low-Stakes Mandatory Proficiency Testing in four German states.

- International Journal of Educational Management (Special Issue): School leadership in Germany between low stakes testing and high expectations*, volume 29, Issue 6.
- Ramsteck, C., Maier, U., Graf, T., Muslic, B. & Kuper, H. (2013). *Die Realisierung testbasierter Schulreform in der Mehrebenenstruktur des Schulsystems – Kategoriale Grundausswertung, Fallbeschreibungen und Fallanalysen*. Abgerufen unter http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/weiterbildung_bildungsmanagement/media/downloads/Die-Realisierung-testbasierter-Schulreform-in-der-Mehrebenenstruktur-des-Schulsystems_-Kategoriale-Grundausswertung_-Fallbeschreibungen-und-Fallanalysen_.pdf am 23.03.2015
- Rudd, P. & Davies, D. (2002). *A revolution in the Use of Data? The LEA Role in Data Collection Analysis and Use and its Impact on Pupil Performance*. Slough: NFER.
- Saunders, L. (2000). Understanding schools' use of 'value added' data: the psychology and sociology of numbers. *Research Papers in Education* 15, H. 3, S. 241–258.
- Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education* 26, H. 3, S. 482–496.
- Schildkamp, K. & Visscher, A. (2010). The use of performance feedback in school improvement in Louisiana. *Teaching and Teacher Education* 26, H. 7, S. 1389–1403.
- Spoden, C., Fleischer, J. & Leutner, D. (2014). Niedrige Testmodellpassung als Resultat mangelnder Auswertungsobjektivität bei der Kodierung landesweiter Vergleichsarbeiten durch Lehrkräfte. *Journal für Mathematik-Didaktik* 35, H. 1, S. 79–99.
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In Hamilton, L. S., Brian M., Stecher, S. & Klein, P. (Hrsg.), *Making sense of test-based accountability in education*, S. 79–100. Santa Monica: RAND Education.
- Van Petegem, P. & Vanhoof, J. (2007). Towards a model of effective school feedback: School heads' points of view. *Educational Research and Evaluation* 13, H. 4, S. 311–325.
- Vanhoof, J., Verhaeghe, G., Van Petegem, P. & Valcke, M. (2012). Flemish Primary Teachers' Use of School Performance Feedback and the Relationship with School Characteristics. *Educational Research* 54, 4, S. 431–449.
- Verhaeghe, G., Vanhoof, J., Valcke, M. & Van Petegem, P. (2010). Using school performance feedback: perceptions of primary school principals. *School Effectiveness and School Improvement* 21, H. 2, S. 167–188.
- Visscher, A. J. & Coe, R. (2003). School performance feedback systems: Conceptualisation, Analysis, and Reflection. *School effectiveness and school improvement* 14, H. 3, S. 321–349.
- Watanabe, Y. (2004). Methodology in Washback Studies. In Cheng, L., Watanabe, Y. & Curtis, A. (Hrsg.), *Washback in Language Testing. Research Contexts and Methods*, S. 19–36. Mahwah, London: Lawrence Erlbaum.
- Wikeley, F., Stoll, L. & Lodge, C. (2002). Effective School Improvement: English Case Studies. *Educational Research and Evaluation. School effectiveness and school improvement in a European context* 8, H. 4, S. 363–385.
- Wurster, S., Richert, D., Schliesing, A. & Pant, H. A. (2013). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. *Rezeption und Nut-*

zung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. *Die Deutsche Schule*, Beiheft 12/2013, S. 19–50.

Yang, M., Goldstein, H., Rath, T. & Hill, N. (1999). The Use of Assessment Data for School Improvement Purposes. *Oxford Review of Education* 25, H. 4, S. 469–483.

Zeitler, S. (2010). *Bildungsstandards in der Schule. Eine rekonstruktive Studie zur Implementation der Bildungsstandards*. Dissertation. Berlin: Humboldt-Universität.

Marc Piopiunik, Guido Schwerdt, Ludger Wößmann¹

Zentrale Abschlussprüfungen als Steuerungsinstrument im Schulsystem: Arbeitsmarkteffekte und Wirkungsmechanismen²

1. Einleitung

Zentrale Abschlussprüfungen sind ein wichtiges Element der Qualitätssicherung im Bildungssystem. Mit nur einer Ausnahme (Rheinland-Pfalz) nutzen inzwischen alle Bundesländer dieses Steuerungsinstrument. Nationale wie internationale Studien belegen, dass zentrale Prüfungssysteme mit besseren Schülerleistungen einhergehen (z. B. Bishop 1997; Wößmann 2003, 2008; Jürges u. a. 2005b). Kritiker warnen jedoch, dass dieser Zusammenhang teilweise auf eine Verengung der Lehrinhalte und des Lernverhaltens auf die zentralen Prüfungen zurückgehen könnte. Dies wirft die Frage auf, ob sich positive Auswirkungen auch langfristig jenseits verbesserter Testleistungen etwa auf dem Arbeitsmarkt zeigen. Diese Frage ist bisher nicht beantwortet.

Ziel des BMBF-geförderten Forschungsprojektes „Zentrale Abschlussprüfungen als Steuerungsinstrument im Schulsystem: Arbeitsmarkteffekte, Wirkungsmechanismen und Effektheterogenitäten“ war es daher, das bestehende Wissen über Wirkungsmechanismen und Arbeitsmarkteffekte zentraler Prüfungen zu erweitern. Konkret wurde zum einen untersucht, ob zentrale Abschlussprüfungen mit dem langfristigen Arbeitsmarkterfolg zusammenhängen. Zum anderen wurde als Wirkungsmechanismus untersucht, ob zentrale Abschlussprüfungen einen Informationswert am Arbeitsmarkt haben, indem Noten stärker mit dem Einkommen zusammenhängen und so durch höhere extrinsische Belohnungen langfristig die Lernanreize für Schüler erhöhen können.³

1 Dr. Marc Piopiunik: ifo Institut für Wirtschaftsforschung an der Universität München, piopiunik@ifo.de; Prof. Dr. Guido Schwerdt: Universität Konstanz, CESifo und IZA, guido.schwerdt@uni-konstanz.de; Prof. Dr. Ludger Wößmann: Lehrstuhl für Volkswirtschaftslehre, insb. Bildungsökonomik, Ludwig-Maximilians-Universität München und Leiter ifo Zentrum für Bildungsökonomik, CESifo und IZA, woessmann@ifo.de.

2 Der vorliegende Beitrag ist eine gekürzte und abgewandelte Version von Piopiunik, Schwerdt und Wößmann (2014).

3 Eine dritte vom Forschungsprojekt bearbeitete Fragestellung behandelte Effektheterogenitäten. Es wurde untersucht, ob sich Arbeitsmarkt- und Kompetenzeffekte zentraler Abschlussprüfungen nach Personengruppen, Fächern, Prüfungseigenschaften und sonstigen Rahmenbedingungen unterscheiden. Generell zeigen sich die Effekte zumeist in verschiedenen Bevölkerungsgruppen und Zusammenhängen. Eine wichtige Interaktion besteht darin, dass sich Autonomiereformen positiver auf Schülerleistungen auswirken, wenn es gleichzeitig zentrale Abschlussprüfungen gibt (Hanushek, Link und Wößmann 2013). Auf Details der Heterogenitätsuntersuchungen wird im Folgenden nicht näher eingegangen.

Um diese Fragen zu beantworten, nutzte das Projekt die Tatsache aus, dass Absolventen mit und ohne zentrale Abschlussprüfungen in Deutschland gemeinsam auf einem nationalen Arbeitsmarkt tätig sind. Je nach Sekundarschulabschluss hatte bis Anfang der 2000er-Jahre lediglich knapp die Hälfte der Bundesländer zentrale Abschlussprüfungen. Anhand der repräsentativen Datensätze des Sozio-ökonomischen Panels (SOEP) und der Absolventenbefragungen des Hochschul-Informationen-Systems (HIS) lassen sich so Zusammenhänge zwischen Zentralität des Prüfungssystems und Arbeitsmarkterfolg untersuchen. Beide Datensätze ermöglichen insbesondere eine Verknüpfung von Informationen über das Bundesland (und damit die Zentralität) des Sekundarschulabschlusses mit dem späteren Arbeitsmarkterfolg.

Die bisher einzige Studie zu den Arbeitsmarktwirkungen zentraler Abschlussprüfungen von Backes-Gellner und Veen (2008) findet keinen Zusammenhang zwischen zentralen Prüfungen und späterem Einkommen von Abiturienten. Allerdings sollten sich Effekte zentraler Abschlussprüfungen aus theoretischer Sicht besonders bei solchen Schulabsolventen finden, die direkt nach dem Schulabschluss in den Arbeitsmarkt eintreten, was in Deutschland insbesondere für Hauptschulabsolventen der Fall ist. Denn in diesem Fall ist das Schulabschlusszeugnis das einzige (erst kürzlich) erworbene Signal über die Produktivität eines Bewerbers, das Unternehmen von diesen Absolventen erhalten. Darüber hinaus ist es auf Arbeitsmärkten mit rigider Lohnbildung möglich, dass sich bessere Bildungsleistungen weniger im Arbeitseinkommen als vielmehr in Form von Beschäftigung bzw. Arbeitslosigkeit auswirken.

In der bildungsökonomischen Literatur herrscht als möglicher Erklärungsansatz für Leistungseffekte zentraler Abschlussprüfungen die bessere Signalwirkung externer Prüfungsnoten vor. Demnach liefern zentrale Prüfungen potenziellen Arbeitgebern und weiterführenden Bildungsinstitutionen ein besseres Signal über die erzielten Prüfungsleistungen, wodurch sich die extrinsische Belohnung für bessere Leistungen und damit die Lernanreize für Schüler erhöhen (z. B. Bishop und Wößmann 2004). Durch die verbesserte Signalwirkung von Prüfungsleistungen wird die Bereitschaft potenzieller Arbeitgeber erhöht, bessere Prüfungsleistungen stärker zu honorieren. Dieser Wirkungsmechanismus hat zur Folge, dass die Schulabschlussnote stärker mit dem späteren Arbeitsmarkterfolg zusammenhängen sollte, wenn sie in einem Schulsystem mit zentralen Abschlussprüfungen erworben wurde. Über den empirischen Gehalt dieses Erklärungsansatzes ist bisher allerdings wenig bekannt.

Vor diesem Hintergrund liefert das Forschungsprojekt neue Erkenntnisse, die die bestehende Evidenz zum Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitsmarkterfolg in dreifacher Hinsicht erweitern. Erstens wird die Variation in der Zentralität der Abschlussprüfungen zwischen den Bundesländern ausgenutzt, um den Einfluss zentraler Abschlussprüfungen auf das Arbeitseinkommen nicht nur von Abiturienten, sondern auch von Personen mit Hauptschul- und Realschulabschluss zu untersuchen. Zweitens wird neben dem Arbeitseinkommen die Arbeitslosigkeit als weiterer Indikator des Arbeitsmarkterfolges betrachtet. Drittens wird der Zusammenhang von Abiturnoten und Arbeitseinkommen nach der Zentralität des Prüfungssystems im Rahmen einer ökonometrischen Differenz-in-Differenzen-Schätzung untersucht, um empirische Evidenz über die Signalwirkung als möglichen Wirkungsmechanismus zu erlangen. Die empirischen Analysen sind in

vertiefter Form in den Studien von Piopiunik, Schwerdt und Wößmann (2013, 2014) und Schwerdt und Wößmann (2014) dargestellt.

2. Literaturüberblick und Ableitung testbarer Hypothesen

Wir beginnen mit einer Diskussion der relevanten bildungsökonomischen Literatur.

2.1 Zentrale Abschlussprüfungen und Schülerleistungen

Zahlreiche internationale Studien zeigen, dass Schüler in Schulsystemen mit zentralen Abschlussprüfungen deutlich besser abschneiden als Schüler in Schulsystemen ohne zentrale Abschlussprüfungen (vgl. Bishop 1997, 2006 mit aggregierten Daten; Wößmann 2003 und Wößmann u.a. 2009 mit individuellen TIMSS- bzw. PISA-Daten; Bishop 1997 für kanadische Provinzen; Graham und Husted 1993 sowie Bishop u. a. 2000 für US-Staaten; Wößmann 2008 für einen Überblick).⁴

Für die deutschen Bundesländer existieren Vergleichsstudien, die zumindest teilweise einen positiven Effekt von zentralen Prüfungen auf Schülerleistungen nahelegen. Baumert und Watermann (2000) zeigen anhand der TIMSS-Daten, dass ein zentral gestelltes Abitur insbesondere im unteren Leistungsbereich mit besseren Mathematik-, nicht aber Physikleistungen einhergeht und gleichzeitig die Motivation und das Lernverhalten der Schüler nicht beeinträchtigt. Jürges u. a. (2005b) verwenden ebenfalls die TIMSS-Daten im Bundesländervergleich und finden einen positiven Effekt zentraler Abschlussprüfungen auf Schülerleistungen. Jürges u. a. (2005a) verwenden Daten der PISA-2000-E-Studie und finden ebenfalls einen positiven Zusammenhang zwischen zentralen Abschlussprüfungen und Schülerleistungen in Deutschland. Wößmann (2010) bestätigt den Zusammenhang mit aggregierten Bundeslanddaten der drei PISA-E-Wellen 2000, 2003 und 2006 und findet zudem, dass sich der positive Zusammenhang zwischen zentralen Abschlussprüfungen und Schülerleistungen in seiner Größenordnung nicht signifikant zwischen den Bundesländerschätzungen und den Schätzungen im internationalen Vergleich der OECD-Staaten unterscheidet.

Die Einführung zentraler Abschlussprüfungen in vielen Bundesländern seit Beginn der 2000er-Jahre ermöglicht zudem Analysen im Längsschnitt. In einem Differenz-in-Differenzen-Ansatz anhand der Mikrodaten der PISA-E-Wellen 2000, 2003 und 2006 findet Lüdemann (2011), dass die Einführung zentraler Abschlussprüfungen in der Haupt- und Realschule mit einem signifikanten Anstieg der Schülerleistungen insbesondere in Mathematik einherging. Anhand einer längsschnittlichen Beobachtung von 37 Schulen in Bremen und Hessen untersucht die Forschungsgruppe um Maag Merki die Wirkung der Einführung zentraler Abiturprüfungen auf Schulklima, Einschätzungen und Verhaltensweisen von Lehrern und Schülern, Un-

⁴ Die Befundlage zum Zusammenhang zwischen zentralen Abschlussprüfungen und Schülerleistungen ist jedoch durchaus gemischt, und es gibt auch Studien, die keinen nachweisbaren Zusammenhang finden können (z. B. Block et al. 2011; Holme et al. 2010, Klein et al. 2014).

terrichtsprozesse und Vergleichbarkeit der Benotung (Maag Merki 2012; Holmeier 2013). Es zeigt sich u. a., dass sich die Mathematikleistungen von Bremer Grundkurschülern verbessert haben, wobei allerdings mögliche Verzerrungen durch Selektion und fächerübergreifende Auswirkungen nicht ausgeschlossen werden können.

2.2 Schülerleistungen und Arbeitsmarkterfolg

Zahlreiche Studien zeigen, dass kognitive Fähigkeiten, wie sie beispielsweise in Schülerleistungstests gemessen werden, positiv mit dem längerfristigen Erfolg am Arbeitsmarkt zusammenhängen (vgl. Murnane u. a. 2000; Altonji und Pierret 2001; Chetty u. a. 2011 sowie Hanushek und Wößmann 2008 und Hanushek und Rivkin 2012 für einen Überblick). Auch für Deutschland zeigen sich signifikante Zusammenhänge von (allerdings im Erwachsenenalter gemessenen) Testleistungen und Einkommen am Arbeitsmarkt (Leuven u. a. 2004; Hanushek und Zhang 2009; Antoni und Heineck 2012; Hanushek u. a. 2014).

Zusammengenommen legen diese Studien also nahe, dass Schüler, die ihren Schulabschluss in einem Bildungssystem mit zentralen Prüfungen erworben haben, später auf dem Arbeitsmarkt erfolgreicher sein könnten als Schüler, die ihren Schulabschluss in einem System ohne zentrale Prüfungen erworben haben.

2.3 Arbeitsmarkteffekte zentraler Abschlussprüfungen

Allerdings lässt diese Evidenz nicht zwangsläufig auf einen kausalen Zusammenhang von zentralen Prüfungen und volkswirtschaftlicher Produktivität durch verbesserte Schülerleistungen schließen. Kritiker warnen insbesondere, dass der positive Effekt von zentralen Prüfungen auf Schülerleistungen lediglich durch veränderte Unterrichtsstrategien wie „Teaching to the Test“ der Lehrkräfte zustande gekommen sein könnte. Hamilton u. a. (2007) berichten beispielsweise, dass das Lehrpersonal auf die Einführung von Accountability- und Monitoring-Maßnahmen mit einem eingeschränkten und zielgerichteten Unterricht im Hinblick auf die Testinhalte reagiert.

Ob zentrale Abschlussprüfungen auch längerfristige Effekte auf dem Arbeitsmarkt haben, ist aufgrund dieser möglicherweise auftretenden unerwünschten Mechanismen also eine offene Frage. Die bisher einzige Studie zu dieser Forschungsfrage stammt von Backes-Gellner und Veen (2008).⁵ Zur Identifikation eines Zentralprüfungseffektes auf den Arbeitsmarkterfolg bei Personen mit Hochschulreife verwenden die Autoren Unterschiede in den Prüfungssystemen zwischen den deutschen Bundesländern. Backes-Gellner und Veen finden keinen signifikanten Unterschied in den Arbeitseinkommen von Abiturienten mit und ohne Zentralabitur.

Dieses Ergebnis könnte aber auch durch den speziellen Fokus der Studie bedingt sein. So beziehen sich Backes-Gellner und Veen (2008) zum einen nur auf Schulab-

5 Außerhalb Deutschlands ist die Evidenz aufgrund der begrenzten Variation der Prüfungssysteme innerhalb der meisten Länder sowie des Fehlens von Datensätzen, die Prüfungsarten mit individuellen Arbeitsergebnissen verbinden, sehr beschränkt.

gänger mit Hochschulreife. Effekte zentraler Prüfungen auf den Arbeitsmarkterfolg dürften aus theoretischer Sicht jedoch besonders bei Schulabschlüssen ausgeprägt sein, die unmittelbar auf den Arbeitsmarkt führen. Dies trifft in Deutschland insbesondere auf Personen mit Hauptschulabschluss zu. Personen, die nach dem Schulabschluss noch weitere Bildungsgänge absolvieren wie die meisten Abiturienten (aber auch – wie weiter unten beschrieben – rund die Hälfte der Realschulabsolventen), erwerben noch ein weiteres Produktivitätssignal (beispielsweise die Hochschulabschlussnote), wodurch die Schulabschlussnote deutlich weniger bedeutsam werden könnte. Bei Schulabgängern, die direkt in den Arbeitsmarkt eintreten, ist die Schulabschlussnote hingegen das einzige Produktivitätssignal für den Arbeitgeber. Daraus ergibt sich folgende Hypothese:

Hypothese 1: Arbeitsmarkteffekte zentraler Abschlussprüfungen sind umso ausgeprägter, je schneller der Eintritt in den Arbeitsmarkt auf den Schulabschluss erfolgt.

Im dritten Abschnitt testen wir diese Hypothese, indem wir neben Personen mit Hochschulreife auch Personen mit Haupt- bzw. Realschulabschluss in die empirische Analyse einbeziehen.

Zum anderen untersuchen Backes-Gellner und Veen (2008) ausschließlich Effekte auf das Einkommen. Es ist jedoch zu vermuten, dass sich die Effekte von zentralen Prüfungen auch in Form von Arbeitslosigkeit zeigen. Beschäftigungseffekte dürften umso eher eintreten, je rigider die Lohnbildung auf dem Arbeitsmarkt ist, je weniger also die individuellen Einkommen der individuellen Produktivität angepasst werden können. Daher ist es insbesondere in Deutschland, einem Land mit historisch eher rigider Lohnbildung (Cahuc und Zylberberg 2004), gut möglich, dass zentrale Prüfungen langfristige Beschäftigungseffekte haben. Daher folgt:

Hypothese 2: Bei rigider Lohnbildung zeigen sich Arbeitsmarkteffekte zentraler Abschlussprüfungen in Form von geringerer Arbeitslosigkeit.

Wir testen diese Hypothese, indem die Arbeitslosigkeit als weiterer Indikator des Arbeitsmarkterfolgs in die Untersuchung einbezogen wird.

2.4 Wirkungsmechanismen

Die Unsicherheit bezüglich der Existenz längerfristiger Zentralprüfungseffekte auf den Arbeitsmarkterfolg verdeutlicht, dass ein besseres Verständnis des Wirkungszusammenhangs zwischen zentralen Prüfungen und Schülerleistungen notwendig ist. Erklärungsansätze, die längerfristige Effekte implizieren, beruhen im Wesentlichen auf einer Verbesserung der Anreizstrukturen der Akteure im Schulsystem, die durch die Besonderheiten von zentralen Prüfungssystemen bedingt werden.

Die ökonomische Literatur betont dabei vor allem die verbesserte Signalwirkung der Abschlussprüfungen für potenzielle Arbeitgeber und weiterführende Bildungsinstitutionen. Die Signaltheorie der Bildungsabschlüsse geht auf die Arbeiten des Nobelpreisträgers Michael Spence zurück. Spence (1973) argumentiert,

dass sich Individuen mit hohen (aber unbeobachtbaren) Fähigkeiten von ansonsten vergleichbaren Individuen mit geringeren Fähigkeiten durch höhere Bildungsabschlüsse differenzieren können, da die individuellen Kosten des Erreichens der höheren Abschlüsse aufgrund der höheren (unbeobachtbaren) Fähigkeiten geringer sind. Da Bildungsabschlüsse beobachtbar sind, dienen diese den potenziellen Arbeitgebern als Signal für die unbeobachtbaren produktiven Fähigkeiten. Becker (1982) und Becker und Rosen (1992) erweitern die theoretischen Modelle zur Signaltheorie durch die explizite Einbeziehung von Prüfungsleistungen. Eine in diesem Zusammenhang wichtige theoretische Implikation ist, dass auch eine akkuratere und vergleichbarere Benotung der Leistungsniveaus von Schülern zu Effizienzgewinnen führen kann.

Bishop und Wößmann (2004) lassen diese Überlegungen in ein Modell der Bildungsproduktion einfließen. Unter der Annahme, dass zentrale Prüfungen die Signalwirkung von Prüfungsleistungen auf dem Arbeitsmarkt erhöhen, zeigen sie theoretisch auf, dass Arbeitgeber bereit sind, Prüfungsleistungen, die verlässlichere Informationen über das Humankapital eines Arbeitnehmers beinhalten, höher zu entlohnen. Dies bedingt, dass bessere Prüfungsergebnisse mit einem höheren Ertrag – höhere Beschäftigungswahrscheinlichkeiten, höhere Löhne oder erleichterte Aufnahme in Institutionen der höheren Bildung – einhergehen, was langfristig die Lernanreize von Schülern in zentralen Prüfungssystemen erhöht. Schüler reagieren auf diese höheren extrinsischen Anreize mit höheren Lernanstrengungen, was wiederum zu besseren Schülerleistungen führt.

Dieser Erklärungsansatz beruht allerdings auf der zentralen Annahme, dass Prüfungsleistungen in Systemen mit zentralen Abschlussprüfungen eine stärkere Signalwirkung auf dem Arbeitsmarkt besitzen.⁶ Diese Annahme ist jedoch nicht unmittelbar verifizierbar. Sollte die Annahme aber korrekt sein, so lässt sich theoretisch ableiten, dass der Zusammenhang zwischen Produktivität auf dem Arbeitsmarkt und Prüfungsleistungen stärker sein sollte, falls die Prüfungsleistungen in zentralen Prüfungen erbracht wurden. Auf Basis dieser Implikation formulieren wir folgende Hypothese, die wir im vierten Abschnitt empirisch überprüfen:

Hypothese 3: Der Zusammenhang zwischen Prüfungsleistung und Arbeitsmarkterfolg ist in Systemen mit zentralen Abschlussprüfungen stärker als in Systemen ohne zentrale Abschlussprüfungen.

6 Dies wird auch durch die Ergebnisse von Büchel u. a. (2004) nahegelegt, die zeigen, dass Noten stärker mit Leistungen korrespondieren, wenn die Lernstandards zentral festgelegt werden. Dies lässt vermuten, dass Prüfungsleistungen in Bundesländern mit zentralen Abschlussprüfungen für potenzielle Arbeitgeber ein besseres Signal sind als Prüfungsleistungen in Bundesländern ohne zentrale Abschlussprüfungen.

3. Zentrale Abschlussprüfungen und Arbeitsmarkterfolg

Zunächst untersuchen wir den Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitsmarkterfolg. Die Analysen basieren auf Piopiunik u. a. (2013).

3.1 Empirisches Schätzmodell

Zur Identifikation der Zentralprüfungseffekte wird die Tatsache ausgenutzt, dass bis zu Beginn der 2000er-Jahre knapp die Hälfte der Bundesländer zentrale Abschlussprüfungen eingesetzt hat, die andere Hälfte jedoch nicht. Diese Variation der Prüfungssysteme wird für die drei Abschlussarten Hauptschul-, Realschulabschluss und Abitur separat betrachtet.

Um zunächst den durchschnittlichen Effekt von Zentralprüfungen auf das individuelle Arbeitseinkommen für alle Abschlussarten zu schätzen, wird der Zusammenhang zwischen dem logarithmierten monatlichen Bruttoarbeitseinkommen und einem binären Indikator analysiert, der angibt, ob das Individuum seinen Schulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben hat. Dabei wird für den Typ des Sekundarschulabschlusses sowie für weitere Eigenschaften der Schulabsolventen – Alter, Geschlecht, Migrationsstatus, Bildungsstand der Eltern und Indikator für Ostdeutschland – kontrolliert (siehe Piopiunik u. a. 2014 für Details). Zusätzlich werden mit Jahresindikatoren makroökonomische Entwicklungen berücksichtigt, da das Modell Einkommensbeobachtungen über mehrere Jahre zusammen betrachtet. In zusätzlichen Robustheitsanalysen kontrollieren wir ebenfalls für diverse andere Unterschiede im Bildungssystem zwischen den Bundesländern.⁷

Um Hypothese 1 zu testen, wird das Regressionsmodell separat für die unterschiedlichen Sekundarschulabschlüsse geschätzt. Die schulabschlusspezifischen Schätzungen ermöglichen es zu überprüfen, ob der langfristige Zentralprüfungseffekt tatsächlich für Personen mit Hauptschulabschluss am stärksten ist. Um Hypothese 2 zu testen, wird in vergleichbarer Weise vorgegangen, wobei nicht lineare Probit-Modelle verwendet werden.

3.2 Daten

Zur Identifikation langfristiger Arbeitsmarkteffekte zentraler Prüfungen werden Unterschiede in der Zentralität der Abschlussprüfungen zwischen den Bundesländern ausgenutzt. Im Jahr 2001 haben vier der sechzehn Bundesländer zentrale Abschlussprüfungen beim Hauptschulabschluss verwendet (Baden-Württemberg,

⁷ Insofern diese Kontrollvariablen nicht alle arbeitsmarktrelevanten Unterschiede der Bildungssysteme vollständig abgreifen, können die geschätzten Koeffizienten nicht vollständig als kausaler Zentralprüfungseffekt interpretiert werden. Beispielsweise könnten lokale Arbeitsmarkteffekte die Effekte der Schulprüfungssysteme konfundieren, falls die Unterschiede zwischen lokalen Arbeitsmärkten nicht – auch nicht langfristig – durch Unterschiede in den Schulprüfungssystemen hervorgerufen worden sind.

Bayern, Sachsen und Thüringen), sechs Bundesländer beim Realschulabschluss (zusätzlich Mecklenburg-Vorpommern und Sachsen-Anhalt) und sieben Bundesländer im Abitur (zusätzlich Saarland).⁸

Die Schätzungen der Arbeitsmarkteffekte basieren auf den Mikrodaten des Sozio-oekonomischen Panels (SOEP). Das SOEP ist eine repräsentative Wiederholungsbefragung privater Haushalte in Deutschland, die seit 1984 (Ostdeutschland seit 1990) jährlich durchgeführt wird. Das SOEP enthält ausführliche Bildungs- und Arbeitsmarktinformationen über die befragten Individuen sowie eine Vielzahl relevanter Hintergrundinformationen.

Der Vorteil des SOEP liegt darin, dass auf Individualebene Informationen über die Zentralität der Abschlussprüfungen mit dem späteren Arbeitsmarkterfolg verknüpft werden können. Das SOEP enthält Informationen über das Bundesland und die Art des Schulabschlusses. Anhand dieser beiden Informationen kann für jedes Individuum im Datensatz bestimmt werden, ob der Sekundarschulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben wurde. Da das Bundesland des Schulabschlusses vom SOEP erst ab 2001 abgefragt wurde, umfasst die Stichprobe den Zeitraum 2001 bis 2010.

Um den Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitsmarkterfolg zu untersuchen, umfasst die Stichprobe Individuen im Alter von 18 bis 65 Jahren. Individuen, die ihren Schulabschluss in der ehemaligen DDR oder in Berlin erhalten haben, bleiben unberücksichtigt. Ebenfalls mussten Abiturienten aus Rheinland-Pfalz (keine zentralen Prüfungen) und aus dem Saarland (zentrale Prüfungen) ausgeschlossen werden, da diese beiden Bundesländer im SOEP-Datensatz zusammen kodiert sind. Insgesamt beinhaltet die SOEP-Einkommensstichprobe, welche ausschließlich Individuen beinhaltet, die entweder vollzeit- oder teilzeitbeschäftigt sind, fast 5.500 Personen und mehr als 27.000 Personen-Jahr-Beobachtungen.

Die Beschäftigungsstichprobe für die Analyse der Arbeitslosigkeitseffekte umfasst zusätzlich auch arbeitslos gemeldete Personen. Diese Stichprobe beinhaltet etwa 5.700 Personen und mehr als 29.000 Personen-Jahr-Beobachtungen.

3.3 Zentrale Abschlussprüfungen und Arbeitseinkommen

Die Regressionsergebnisse zeigen, dass Personen, die ihren Schulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben haben, im Durchschnitt 6,4 Prozent mehr verdienen als Personen, die ihren Schulabschluss in einem Bundesland ohne zentrale Abschlussprüfungen erworben haben (siehe Piopiunik u. a. 2014 für Details). Der Effekt ist statistisch hochsignifikant. Dieser Durchschnittsef-

⁸ In den 2000er-Jahren haben die meisten Bundesländer, die vorher keine zentralen Abschlussprüfungen hatten, solche eingeführt. Es ist zu früh, die Auswirkungen der Einführung dieser zentralen Abschlussprüfungen auf den Arbeitsmarkterfolg zu untersuchen (Lüdemann 2011 untersucht Auswirkungen der Reformen auf Schülerleistungen). Daher bezieht sich unsere Analyse auf den Zustand im Jahr 2001 vor diesen Reformen, der für die Individuen in unseren Stichproben relevant ist.

fekt bezieht sich auf alle Personen und verdeckt daher möglicherweise bedeutende Unterschiede zwischen den unterschiedlichen Schulabschlussarten.

Die schulabschlussspezifischen Ergebnisse deuten in der Tat auf deutliche Unterschiede zwischen den verschiedenen Schulabschlüssen hin. Personen mit Hauptschulabschluss – die Absolventengruppe mit dem unmittelbarsten Arbeitsmarkteintritt – verdienen durchschnittlich 10 Prozent mehr, wenn sie ihren Schulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben haben. Für Personen mit Realschulabschluss und Abitur lässt sich hingegen kein statistisch signifikanter Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitseinkommen feststellen. Das Abiturientenergebnis bestätigt somit die Ergebnisse von Backes-Gellner und Veen (2008). Allerdings sind die ermittelten Einkommenseffekte zentraler Prüfungen sowohl bei Personen mit Realschulabschluss (3,7 Prozent) als auch bei Abiturienten (4,4 Prozent) positiv – und weisen damit in die erwartete Richtung – und quantitativ bedeutsam, aber nicht statistisch signifikant.

Die starken Einkommenseffekte für Personen mit Hauptschulabschluss sowie die nicht signifikanten Ergebnisse für Personen mit Realschulabschluss bzw. Abitur liefern Evidenz für Hypothese 1:⁹ Die Arbeitsmarkteffekte zentraler Abschlussprüfungen sind umso ausgeprägter, je schneller der Eintritt in den Arbeitsmarkt auf den Schulabschluss erfolgt. In unserer Stichprobe tritt die überwiegende Mehrheit (79 Prozent) der Personen mit Hauptschulabschluss in der Tat direkt nach Schulabschluss, also ohne vorher einen weiteren Bildungsabschluss zu erwerben, in den Arbeitsmarkt (68 Prozent machen eine duale Ausbildung, und 11 Prozent erwerben keinen weiteren Abschluss). Im Gegensatz dazu betreten nur 21 Prozent der Personen mit Hochschulreife den Arbeitsmarkt direkt nach Schulabschluss. Personen mit Realschulabschluss liegen dazwischen: Etwa die Hälfte (ca. 52 Prozent) tritt direkt nach dem Schulabschluss in den Arbeitsmarkt, während die andere Hälfte (ca. 48 Prozent) nach der Schule zunächst einen weiteren Bildungsabschluss erwirbt (an Vollzeit-Berufsschulen oder in Institutionen tertiärer Bildung). Diese Unterschiede im Arbeitsmarkteintritt stehen in Einklang mit stärkeren Effekten zentraler Abschlussprüfungen für Personen mit Hauptschulabschluss.

3.4 Zentrale Abschlussprüfungen und Arbeitslosigkeit

Die Analysen zum Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitslosigkeit zeigen, dass Personen, die ihren Schulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben haben, im Durchschnitt ein um 2,1 Prozentpunkte geringeres Arbeitslosigkeitsrisiko haben als Personen mit einem Schulabschluss aus einer nicht zentralen Prüfung. Dieser statistisch hochsignifikante Effekt könnte wiederum Unterschiede zwischen den Schulabschlussarten verdecken.

In der Tat ist der Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitslosigkeit für Personen mit Hauptschulabschluss am stärksten ausgeprägt. Per-

⁹ Das Ergebnismuster von Effekten bei Hauptschulabsolventen, aber nicht bei Abiturienten, steht mit den Ergebnissen für Schülerleistungen von Jürges u. a. (2012) in Einklang.

sonen mit Hauptschulabschluss haben eine 4,5 Prozentpunkte geringere Wahrscheinlichkeit, arbeitslos zu sein, wenn sie ihren Schulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben haben. Bei Personen mit Realschulabschluss findet sich kein Zusammenhang zwischen zentralen Abschlussprüfungen und Arbeitslosigkeit. Während es bei Abiturienten keinen Effekt auf das Arbeitseinkommen gab, zeigt sich, dass Abiturienten eine um 1,9 Prozentpunkte geringere Arbeitslosigkeitswahrscheinlichkeit haben, wenn sie ihr Abitur in einem Bundesland mit zentralen Abschlussprüfungen erworben haben. Dieses Ergebnis liefert Evidenz für Hypothese 2, dass sich auf einem Arbeitsmarkt mit rigider Lohnbildung Arbeitsmarkteffekte zentraler Abschlussprüfungen auch in Form von Arbeitslosigkeit zeigen können.

4. Der Informationswert zentraler Abiturnoten als Wirkungsmechanismus

In diesem Abschnitt testen wir die Implikation der Signalwirkung als möglichen Wirkungsmechanismus, dass die Noten zentral gestellter Abschlussprüfungen einen höheren Informationswert auf dem Arbeitsmarkt haben. Die methodische Vorgehensweise folgt Schwerdt und Wößmann (2014).

4.1 Empirisches Schätzmodell

Den Kern der empirischen Strategie bildet die Tatsache, dass es auf dem deutschen Arbeitsmarkt möglich ist, sowohl für Schulabgänger aus Zentralprüfungssystemen als auch für Schulabgänger aus Systemen ohne zentrale Prüfungselemente vergleichbare Schätzergebnisse für den Zusammenhang zwischen Abschlussnoten und Arbeitsmarkterträgen zu erhalten. Dieser Zusammenhang wird also getrennt für Schulabgänger aus Zentralprüfungssystemen und Schulabgänger aus Systemen ohne zentrale Prüfungselemente geschätzt. Dazu wird das logarithmierte monatliche Bruttoarbeitseinkommen auf die Abschlussnote und relevante Kontrollvariablen (Alter, Geschlecht, Indikatoren für Fachabitur und Gymnasialbesuch und Bildungsstand der Eltern) regressiert. Ein signifikanter Schätzwert für den Koeffizienten der Note impliziert, dass bessere Abschlussnoten mit besseren Arbeitsmarkterträgen einhergehen.

Um die Gültigkeit von Hypothese 3 zu überprüfen, wird zudem ein Gesamtmodell mit einem Interaktionseffekt zwischen einem Indikator für die Art des Prüfungssystems und der Abschlussnote geschätzt. Zusätzlich kann in der Schätzung für das Bundesland des Schulabschlusses kontrolliert werden. Somit wird eine Verzerrung der Ergebnisse durch unterschiedliche Arbeitsmarktbedingungen oder sonstige bundeslandspezifische Einflussfaktoren ausgeschlossen. Die Identifikation des interessierenden Effektes basiert somit nicht auf dem Vergleich des allgemeinen Lohnniveaus, sondern auf dem Vergleich von Lohndifferenzen zwischen Schulabgängern mit guten und schlechten Abschlussnoten aus Bundesländern mit und ohne zentrale Abschlussprüfungen. Die Identifikationsstrategie entspricht damit der eines Differenz-in-Differenzen-Ansatzes.

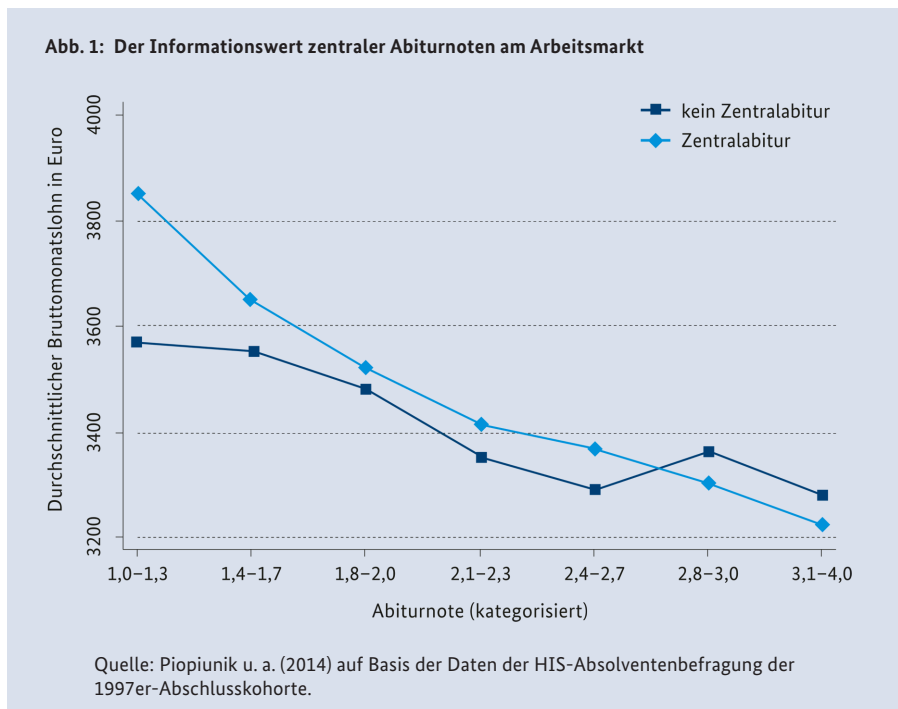
4.2 Daten

Für die Schätzungen benötigt man Informationen sowohl über den Arbeitsmarkterfolg als auch über die Note und das Bundesland des Schulabschlusses. Leider stehen solche Daten nicht für alle Schulabgänger zur Verfügung. Aber die Absolventenbefragungen des Hochschul-Informations-Systems (HIS) beinhalten alle notwendigen Informationen. Dies sind Längsschnittuntersuchungen von Hochschulabsolventen. Für diese Studie werden die Daten der Absolventenkohorte von 1997 verwendet, die als Scientific Use File (ZA 4272) über das GESIS-ZA – Zentralarchiv für Empirische Wirtschaftsforschung verfügbar gemacht wurden.

Die HIS-Daten beinhalten neben umfangreichen Informationen zum Arbeitsmarkterfolg und relevanten Hintergrundvariablen der Absolventen auch die Abschlussnote und das Bundesland der Hochschulzugangsberechtigung. Die finale Stichprobe beinhaltet 4.701 Hochschulabsolventen, die im Jahr 2003 beschäftigt sind.

4.3 Ergebnisse

Abbildung 1 zeigt die Variation in den Daten, die das beschriebene Modell mit Interaktionseffekt ausnutzt, und verdeutlicht, weshalb die Identifikation in diesem Ansatz einer Differenz-in-Differenzen-Schätzung gleichkommt. Auf der horizontalen Achse ist die Abiturnote in kategorisierter Form abgetragen. Für jede Notenkatgorie wird der durchschnittliche Bruttomonatslohn im Jahre 2003 auf der vertikalen Achse



abgetragen. Die Abbildung verdeutlicht, dass bessere Noten erwartungsgemäß mit einem höheren Lohn einhergehen. Insbesondere legt die Abbildung nahe, dass der Zusammenhang zwischen Abschlussnote und Arbeitsmarkterfolg in Systemen mit zentralen Abschlussprüfungen stärker ausgeprägt ist als in Systemen ohne zentrale Abschlussprüfungen und damit eine in zentralen Prüfungen erzielte Abiturnote einen höheren Informationswert für den Arbeitsmarkterfolg aufweist.

In multivariaten Regressionsanalysen lässt sich untersuchen, ob dieser Unterschied statistisch signifikant ist und auch nach Kontrolle für weitere Einflussfaktoren Bestand hat. Die Ergebnisse bestätigen die grafische Darstellung, dass bessere Noten in beiden Prüfungssystemen statistisch signifikant mit einem höheren Lohn einhergehen. Dies bestätigt, dass Abiturnoten im Hinblick auf spätere Arbeitsmarkterträge informativ sind.

Allerdings sind die Zusammenhänge unterschiedlich stark. Unter Absolventen aus Bundesländern ohne zentrale Abschlussprüfungen geht eine Notenverbesserung um eine Notenstufe (beispielsweise von 2,5 auf 1,5) mit einem Lohnanstieg um 2,5 Prozent einher. Hingegen erhalten Absolventen aus Bundesländern mit zentralen Abschlussprüfungen im Durchschnitt ein um 9,3 Prozent höheres Einkommen pro Notenstufe. Ob der resultierende Unterschied von knapp 6 Prozentpunkten allerdings statistisch signifikant ist, ist aus den separaten Schätzungen nicht ersichtlich.

Die Antwort auf diese Frage liefern die Ergebnisse des Modells mit Interaktionseffekt. Diese bestätigen zunächst, dass ohne zentrale Abschlussprüfungen eine Abiturnotenverbesserung um eine Notenstufe mit einem Lohnanstieg um 2,5 bis 2,9 Prozent einhergeht (einmal mit und einmal ohne fixe Effekte für das Bundesland der Hochschulreife). Der Koeffizient des Interaktionsterms zwischen Prüfungszentralität und Abiturnote zeigt, dass dieser Zusammenhang bei Absolventen aus Bundesländern mit zentralen Abschlussprüfungen um 5,5 bis 6,9 Prozentpunkte stärker ist. Die Schätzung des Interaktionsterms zeigt, dass der Unterschied auch statistisch hochsignifikant ist.

In weiteren Analysen zeigt sich, dass sich der Zusammenhang kaum ändert, wenn zusätzliche Kontrollvariablen für Merkmale der Hochschulausbildung wie Dauer des Studiums, Hochschulabschlussnote, Studienfach und Bundesland der Hochschule aufgenommen werden (Schwerdt und Wößmann 2014). Der Koeffizient verringert sich um höchstens ein Drittel, wenn Merkmale des ausgeübten Berufs wie Berufsgruppe, Industrie, öffentlicher Sektor und Bundesland der Berufstätigkeit als Kontrollvariablen aufgenommen werden. Daher scheint der höhere Signalwert zentraler Abiturnoten für das spätere Einkommen nicht in erster Linie durch eine selektive Wahl des Hochschulstudiums oder des Berufs zustande zu kommen.

Insgesamt belegen die Ergebnisse, dass der Informationswert von Abiturnoten aus Bundesländern mit zentralen Prüfungen deutlich stärker ausgeprägt ist als der Informationswert von Abiturnoten aus Bundesländern ohne zentrale Prüfungen. Dieses Ergebnis ist nicht nur relevant im Hinblick auf das bessere Verständnis der Wirkungsmechanismen von zentralen Prüfungen, sondern deutet auch auf eine potenzielle Verbesserung des Matching-Prozesses von Stellensuchenden und offenen

Stellen auf dem Arbeitsmarkt durch zentrale Prüfungen hin.¹⁰ Einschränkend muss festgehalten werden, dass die Ergebnisse zwar einen höheren Informationswert von Prüfungsleistungen im Hinblick auf die spätere Produktivität am Arbeitsmarkt belegen, es jedoch im Rahmen der vorliegenden Daten nicht möglich ist zu bestimmen, ob und inwieweit Unternehmen tatsächlich diese Information bei der Einstellungsentscheidung und Lohnsetzung berücksichtigen.

5. Schlussbemerkungen

Das Forschungsprojekt hat Evidenz geliefert, dass zentrale Abschlussprüfungen auch langfristige Auswirkungen auf dem Arbeitsmarkt haben können. Die Ergebnisse erweitern die bisherige empirische Evidenz, die sich ausschließlich auf Schulabgänger mit Hochschulreife sowie nur auf Einkommenseffekte bezieht und deshalb keine längerfristigen Effekte von zentralen Abschlussprüfungen auf den Arbeitsmarkterfolg findet. Unsere Ergebnisse deuten zum einen darauf hin, dass die Arbeitsmarkteffekte zentraler Prüfungen am ausgeprägtesten sind für Schulabsolventen, die unmittelbar nach Schulabschluss in den Arbeitsmarkt treten. Konkret finden wir, dass Personen mit Hauptschulabschluss rund 10 Prozent mehr verdienen und ein 4,5 Prozentpunkte geringeres Arbeitslosigkeitsrisiko haben, wenn sie ihren Schulabschluss in einem Bundesland mit zentralen Abschlussprüfungen erworben haben.

Zum anderen legen die Ergebnisse nahe, dass sich Arbeitsmarkteffekte bei rigider Lohnbildung auch in Form von Beschäftigung zeigen: So finden sich für Abiturienten zwar keine Einkommenseffekte, aber positive Beschäftigungseffekte zentraler Prüfungen in der Größenordnung von einem knapp 2 Prozentpunkte geringeren Arbeitslosigkeitsrisiko. Obgleich es im Rahmen des Studiendesigns nicht möglich ist, für alle Einflussfaktoren des individuellen Arbeitsmarkterfolgs zu kontrollieren und die Ergebnisse daher nicht zwangsläufig kausal interpretiert werden können, legen diese Befunde zumindest nahe, dass zentrale Abschlussprüfungen mit langfristigen Auswirkungen auf dem Arbeitsmarkt einhergehen.

Des Weiteren hat das Forschungsprojekt gezeigt, dass der Zusammenhang zwischen Abiturnote und Arbeitsmarkterfolg in Systemen mit zentralen Abschlussprüfungen stärker ist als in Systemen ohne zentrale Abschlussprüfungen. Konkret zeigt sich, dass ohne zentrale Abschlussprüfungen eine Notenverbesserung um eine Notenstufe mit einem Lohnanstieg um knapp 3 Prozent einhergeht. Mit zentralen Abschlussprüfungen ist dieser Zusammenhang um 6 bis 7 Prozentpunkte stärker.

Dies liefert Evidenz für einen möglichen Wirkungsmechanismus, weshalb zentrale Prüfungen zu besseren Schülerleistungen führen können: Da bessere Schülerleistungen in zentralen Prüfungssystemen stärker belohnt werden, erhöht dies den Anreiz für Schüler, sich besonders anzustrengen. Dieser empirische Befund ist zwar

¹⁰ In weiteren Analysen zeigt sich etwa, dass Schulabgänger mit Hochschulreife aus einem Bundesland mit zentralen Abschlussprüfungen deutlich mehr Einladungen zu Bewerbungsgesprächen und mehr Jobangebote pro Bewerbung bekommen als Schulabgänger, die ihre Hochschulreife in Bundesländern ohne zentrale Abschlussprüfungen erworben haben (Schwerdt und Wößmann 2014).

kein unmittelbarer Beleg dafür, dass zentrale Prüfungen durch höhere extrinsische Anreize zu höheren Lernanstrengungen führen. Allerdings ist ein höherer Informationswert von zentralen Prüfungsleistungen für das Einkommen und damit eine höhere Signalwirkung im Hinblick auf die spätere Produktivität am Arbeitsmarkt eine notwendige Voraussetzung für diese Wirkungskette.

Um die Mechanismen, die zum beobachteten Zusammenhang zwischen Zentralität des Prüfungssystems und Arbeitsmarkterfolg führen, besser zu verstehen, ist eine umfangreiche, repräsentative Befragung von mehr als 1.000 Personalleitern in deutschen Unternehmen im Rahmen des Anschlussprojektes „Abschlussprüfungen als Steuerungsinstrument im Schulsystem: Die Bedeutung von Schulabschlussnoten bei der Einstellungsentscheidung von Unternehmen“ geplant. Diese Erhebung wird es ermöglichen, die Bedeutung, die Schulnoten tatsächlich für die Einstellungsentscheidung in deutschen Unternehmen haben, zu untersuchen.¹¹

Literaturverzeichnis

- Altonji, J. G., C. R. Pierret (2001), Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics*, 116 (1), S. 313–350.
- Antoni, M., G. Heineck (2012), *Do Literacy and Numeracy Pay off? On the Relationship between Basic Skills and Earnings*. IAB Discussion Paper 21/2012, Nürnberg: Institute for Employment Research.
- Backes-Gellner, U., S. Veen (2008), The Consequences of Central Examinations on Educational Quality Standards and Labour Market Outcomes. *Oxford Review of Education*, 34 (5), S. 569–588.
- Baumert, J., Watermann, R. (2000), Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. In J. Baumert, W. Bos, R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2*, S. 317–372. Opladen: Leske + Budrich.
- Becker, W. E. (1982), The Educational Process and Student Achievement Given Uncertainty in Measurement. *American Economic Review*, 72 (1), S. 229–236.
- Becker, W. E., S. Rosen (1992), The Learning Effect of Assessment and Evaluation in High School. *Economics of Education Review*, 11 (2), S. 107–118.
- Bishop, J. H. (1997), The Effect of National Standards and Curriculum-Based Examinations on Achievement. *American Economic Review*, 87 (2), S. 260–264.
- Bishop, J. H. (2006), Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems, and Peer Pressure. In A. Hanushek, E. A., Welch, F. (Hrsg.), *Handbook of the Economics of Education*, S. 909–944. Amsterdam: North-Holland.

11 Weitere Informationen zum Anschlussprojekt sind auf der SteBis-Website http://www.stebis.de/forschungsprojekte/woessmann_2tefoerderphase/index.html erhältlich.

- Bishop, J. H., J. Moriarty, F. Mane (2000), Diplomas for Learning, not Seat Time: The Impact of New York Regents Examinations. *Economics of Education Review*, 19 (4), S. 333–349.
- Bishop, J. H., L. Wößmann (2004), Institutional Effects in a Simple Model of Educational Production. *Education Economics* 12 (1), S. 17–38.
- Block, R., E. D. Klein, I. van Ackeren, S. M. Kühn (2011), Leistungseffekte des Zentralabiturs? Eine kritische Auseinandersetzung mit bildungsökonomischen Interpretationen zu den Effekten der Prüfungsorganisation auf der Basis von PISA E 2003-Daten. *Bildungsforschung*, 8 (1), S. 215–238.
- Büchel, F., H. Jürges, K. Schneider (2004), Leistungs- und Signaleffekte zentraler Abschlussprüfungen: Eine TIMSS-Auswertung bei deutschen Haupt- und Realschülern. In U. Backes-Gellner, P. Moog (Hrsg.), *Ökonomie der Evaluation von Schulen und Hochschulen. Schriften des Vereins für Socialpolitik*, Band 302, S. 53–73.
- Cahuc, P., A. Zylberberg (2004), *Labor Economics*. Cambridge: MIT Press.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. Whitmore Schanzenbach, D. Yagan (2011), How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, 126 (4), S. 1593–1660.
- Graham, A., T. Husted (1993), Understanding State Variation in SAT Scores. *Economics of Education Review*, 12 (3), S. 197–202.
- Hamilton, L. S., B. M. Stecher, J. A. Marsh, J. S. McCombs, A. Robyn, J. Russell, S. Naftel, H. Barney (2007), Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States. Santa Monica: RAND.
- Hanushek, E. A., S. Link, L. Wößmann (2013), Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA. *Journal of Development Economics*, 104, S. 212–232.
- Hanushek, E. A., S. G. Rivkin (2012), The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics*, 4, S. 131–157.
- Hanushek, E. A., G. Schwerdt, S. Wiederhold, L. Wößmann (2014). Returns to Skills around the World. *European Economic Review*, im Erscheinen.
- Hanushek, E. A., L. Wößmann (2008), The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46 (3), S. 607–668.
- Hanushek, E. A., L. Zhang (2009), Quality-consistent Estimates of International Schooling and Skill Gradients, *Journal of Human Capital*, 3 (2), S. 107–143.
- Holme, J. J., M. P. Richards, J. B. Jimerson, R. W. Cohen (2010), Assessing the Effects of High School Exit Examinations. *Review of Educational Research*, 80 (4), S. 476–526.
- Holmeier, M. (2013), *Leistungsbeurteilung im Zentralabitur*, Wiesbaden: Springer VS.
- Jürges, H., W. F. Richter, K. Schneider (2005a), Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality. *FinanzArchiv*, 61 (3), S. 298–326.
- Jürges, H., K. Schneider, F. Büchel (2005b), The Effect of Central Exit Examinations on Student Achievement: Quasi-experimental Evidence from TIMSS Germany. *Journal of the European Economic Association*, 3 (5), S. 1134–1155.
- Jürges, H., K. Schneider, M. Senkbeil, C. H. Carstensen (2012), Assessment Drives Learning: The Effect of Central Exit Exams on Curricular Knowledge and Mathematical Literacy. *Economics of Education Review*, 31 (1), S. 56–65.

- Klein, E. D., M. Krüger, S. M. Kühn, I. van Ackeren (2014), Wirkungen zentraler Abschlussprüfungen im Mehrebenenensystem Schule. Eine Zwischenbilanz internationaler und nationaler Befunde und Forschungsdesiderata. *Zeitschrift für Erziehungswissenschaft*, 17 (1), S. 7–33.
- Leuven, E., H. Oosterbeek, H. van Ophem (2004), Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skills. *Economic Journal*, 114 (495), S. 466–486.
- Lüdemann, E. (2011), Intended and Unintended Short-run Effects of the Introduction of Central Exit Exams. *Schooling and the Formation of Cognitive and Non-cognitive Outcomes*, Chapter 4. ifo Beiträge zur Wirtschaftsforschung 39. München: ifo Institut.
- Maag Merki, K. (Hrsg.) (2012), *Zentralabitur: Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*, Wiesbaden: Springer VS.
- Murnane, R. J., J. B. Willett, Y. Duhaldeborde, J. H. Tyler (2000), How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings? *Journal of Policy Analysis and Management*, 19 (4), S. 547–568.
- Piopiunik, M., G. Schwerdt, L. Wößmann (2013), Central School Exit Exams and Labor-Market Outcomes. *European Journal of Political Economy* 31, S. 93–108.
- Piopiunik, M., G. Schwerdt, L. Wößmann (2014), Zentrale Abschlussprüfungen, Signalwirkung von Abiturnoten und Arbeitsmarkterfolg in Deutschland. *Zeitschrift für Erziehungswissenschaft* 17 (1), S. 35–60.
- Schwerdt, G., L. Wößmann (2014), *The Information Value of Central School Exams*, Mimeo, ifo Institut.
- Spence, A. M. (1973), Job Market Signaling. *Quarterly Journal of Economics*, 87 (3), S. 355–374.
- Wößmann, L. (2003), Central Exit Exams and Student Achievement: International Evidence. In P. E. Peterson, M. R. West (Hrsg.), *No Child Left Behind? The Politics and Practice of School Accountability*, S. 292–323, Washington, DC: Brookings Institution Press.
- Wößmann, L. (2008), Zentrale Abschlussprüfungen und Schülerleistungen: Individualanalysen anhand von vier internationalen Tests, *Zeitschrift für Pädagogik*, 54 (6), S. 810–826.
- Wößmann, L. (2010), Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries. *Jahrbücher für Nationalökonomie und Statistik*, 230 (2), S. 234–270.
- Wößmann, L., E. Lüdemann, G. Schütz, M. R. West (2009), *School Accountability, Autonomy and Choice around the World*. Cheltenham: Edward Elgar.

2. Externe Evaluation durch Schulinspektion: Steuerungsimpulse und Datenrezeption

Kathrin Dederling

Der Umgang von Schulen mit Schulinspektionen – Zum Projekt „Externe Evaluation/Schulinspektion und Qualitätsentwicklung von Schulen“

Zusammenfassung

Der Beitrag richtet seinen Blick auf das Instrument der Schulinspektion, dem in Deutschland im Zuge der Neuen Steuerung eine große Bedeutung zukommt. Er geht der Frage nach, wie Schulen mit den Ergebnissen der Schulinspektion umgehen, und basiert dabei auf der Annahme, dass die spezifische Situation einer Schule einen Einfluss darauf hat, ob und in welcher Weise die Ergebnisse der Schulinspektion in Schulen rezipiert und verarbeitet werden. Um die Annahme zu untermauern, werden Ergebnisse aus dem Projekt „Externe Evaluation/Schulinspektion und schulische Qualitätsentwicklung“ des BMBF-Förderschwerpunktes „Steuerung im Bildungssystem (SteBis)“ präsentiert, die im Rahmen von Fallstudien auf der Basis von Leitfadeninterviews mit Schulleitungen, Lehrkräften, Schulinspektoren und Schulaufsichtsbeamten sowie standardisierten Befragungen mit Schulleitungen und Lehrkräften in den Jahren 2010 bis 2013 gewonnen wurden. Vor dem Hintergrund des von Ehren und Visscher (2006) vorgelegten und von Landwehr (2011) erweiterten heuristischen Rahmenmodells zur Wirksamkeit von Schulinspektion können aus den Beschreibungen der Verarbeitungsprozesse an den Schulen einige innerschulische Merkmale sowie externe Impulse als relevante Faktoren abgeleitet werden. Die Ergebnisse deuten darauf hin, dass vor allem das Ausmaß, in dem eine etablierte Organisationsstruktur der Entwicklungsarbeit (festgelegte Zuständigkeiten und Orte für erforderliche Arbeitsschritte) in den Schulen vorhanden ist, wichtig ist. Es wird daher dafür plädiert, dass – sofern auch weiterhin am Einsatz eines einheitlichen Inspektionsverfahrens für alle allgemeinbildenden Schulen der Bundesländer festgehalten werden soll – zum einen die mit der Schulinspektion verbundenen Wirkungshoffnungen der Bildungspolitik und -administration an das jeweilige Niveau der Schulentwicklungsarbeit in den einzelnen Schulen angepasst werden sollten. Zum anderen sollte die potenzielle Unterstützung der Schulen bei der Ergebnisverarbeitung noch besser als bisher an deren Bedarfslage ausgerichtet werden.

1. Einleitung

Während die evidenzbasierte Steuerung des Bildungssystems in den USA und einigen europäischen Ländern bereits eine lange Tradition aufweist, finden sich Aktivitäten zur Implementierung einer derartigen Steuerungsstrategie in Deutschland erst seit Ende der 1990er-Jahre. Verfahren der Schulinspektion stellen in diesem Zusam-

menhang seit etwa zehn Jahren in allen Bundesländern ein wesentliches Element dar (vgl. Döbert et al. 2008). Über die Bundesländergrenzen hinweg zeigt sich dabei ein Sachverhalt: Alle allgemeinbildenden Schulen eines Landes werden anhand eines einheitlichen Verfahrens extern begutachtet, und für alle Schulen wird in identischer Weise die Erwartung formuliert, dass sie aus den Ergebnissen der Begutachtung Impulse für ihre pädagogische und organisatorische Weiterentwicklung ziehen und Entwicklungsaktivitäten auf den Weg bringen – ungeachtet des spezifischen Entwicklungsstandes, den die Schulen im Bereich der Qualitätsentwicklungsarbeit vorweisen. Mit anderen Worten: Ob Schulen bereits seit vielen Jahren aus eigener Kraft heraus oder im Rahmen organisierter Projekte engagiert an Entwicklungsvorhaben arbeiten und dabei bereits bestimmte Strukturen herausgebildet haben, die für eine erfolgreiche Entwicklungsarbeit erforderlich sind, oder ob Schulen bisher keine Entwicklungsarbeit geleistet haben (möglicherweise auch, weil sie erst vor einigen Jahren aus bestehenden Einrichtungen neu gegründet worden sind), wird nicht in Rechnung gestellt.

Hier setzt der vorliegende Beitrag an: Er basiert auf der Annahme, dass die spezifische Situation einer Schule Auswirkungen darauf hat, ob und in welcher Weise die Ergebnisse der Schulinspektion in Schulen rezipiert und verarbeitet werden. Vor allem das Ausmaß, in dem eine etablierte Organisationsstruktur der Entwicklungsarbeit – festgelegte Zuständigkeiten und etablierte Orte für erforderliche Arbeitsschritte – vorhanden ist, ist wichtig.

Um die Annahme zu untermauern, werden die Befunde des SteBis-Projekts „Externe Evaluation/Schulinspektion und schulische Qualitätsentwicklung“ bemüht, die exemplarisch in zwei deutschen Bundesländern (Hamburg und Thüringen) mittels standardisierter Befragungen und qualitativer Fallstudien gewonnen worden sind.

Im Folgenden wird zunächst Bezug auf das Verfahren der Schulinspektion in Deutschland genommen (Kapitel 2). Anschließend wird der derzeitige Forschungsstand zum Zusammenhang von Schulinspektion und Schulentwicklung in Deutschland expliziert (Kapitel 3), bevor die eigene Studie (Kapitel 4) sowie deren Ergebnisse in Bezug auf die eingangs formulierte Annahme präsentiert werden (Kapitel 5). Der Beitrag mündet in einem Fazit, in dem Überlegungen für eine angemessene Berücksichtigung der spezifischen Schulsituation angestellt werden (Kapitel 6).

2. Schulinspektion in Deutschland

Die Schulinspektion ist als Instrument der Qualitätsentwicklung in den Jahren 2004 bis 2008 in allen 16 Bundesländern implementiert worden (vgl. Döbert et al. 2008). Länderübergreifend finden sich bestimmte Elemente, die das Verfahren der externen Begutachtung auszeichnen und eine wiederkehrende Grundstruktur darstellen: Schulinspektionen liegt stets ein normativer Rahmen mit einheitlichen Qualitätskriterien zugrunde. Inhalte derartiger Qualitätsrahmen sind neben den Ergebnissen (Output) einer Schule größtenteils Prozessqualitäten wie z. B. die Qualität des Unterrichts, das Schulmanagement sowie Maßnahmen der Qualitätsentwicklung. Die Schulinspektionen erfolgen – nach Ankündigung in den Schulen – mittels ein-

heitlicher Verfahren bzw. Instrumente durch ein ausgebildetes Inspektorenteam. Bestandteil ist immer ein Schulbesuch mit Interviews, Unterrichtsbeobachtungen und ggf. einer Schulbegehung nach einem festgelegten Ablauf. Ihm vorgeschaltet sind zumeist Dokumentenanalysen und Befragungen von Lehrkräften, Eltern und Schülerinnen und Schülern. Erfasst werden sollen die Sichtweisen aller schulischen Akteure (Schulleitungen, Lehrkräfte, Schüler und Schülerinnen, Eltern, nicht unterrichtendes Personal). Die so gewonnenen Ergebnisse werden aggregiert und zumeist in ein Stärken-Schwächen-Profil überführt, das in einem Evaluationsbericht an die Schulen (und die Schulaufsicht) zurückgemeldet wird (vgl. Dederling 2012). Intendiert wird damit, mittels der Standardisierung der Verfahren, der Pluralität der Instrumente und der Vielfalt der Perspektiven zu möglichst validen Einschätzungen der Schulqualität zu gelangen. Vornehmlich soll durch die mündliche wie schriftliche Rückmeldung der Ergebnisse Wissen generiert werden, das dann erstens von der Schule selbst genutzt werden kann, um eigenständig Entwicklungsprozesse anzustoßen (Schulentwicklungsfunktion). Es kann zweitens von der zuständigen Schulaufsicht genutzt werden, um einen Überblick über den Qualitätsstand der Schulen zu erhalten und ggf. (bei schwachen Schulen) Entwicklungsmaßnahmen anzuordnen (Kontrollfunktion). Während der Vorbereitung und Durchführung der Schulinspektion können drittens die normativen Rahmen in den Blick der Schulen geraten und Gegenstand einer Auseinandersetzung werden (Funktion der Normendurchsetzung) (vgl. Landwehr 2011). Konsequenzen werden aus den Ergebnissen vonseiten der Bildungspolitik bzw. -administration nicht abgeleitet; Schulen haben demnach weder Belohnungen noch Sanktionen zu erwarten. Jenseits dieser Grundmerkmale, die sich über die Ländergrenzen hinweg finden, zeigen sich allerdings länderspezifische Ausgestaltungen (vgl. Döbert et al. 2008), etwa hinsichtlich des Standardisierungsgrades der Instrumente, der Weitergabe von Handlungsempfehlungen im Ergebnisbericht und der Identifikation von Schulen, die einen Mindeststandard von Qualität nicht erreichen.

In den meisten Bundesländern sind alle Schulen inzwischen einmal inspiziert und die Verfahren mehr oder weniger umfangreichen Modifikationen unterzogen worden (vgl. Dederling et al. 2014). In drei Bundesländern (Berlin, Brandenburg und Hamburg) werden die schulspezifischen Ergebnisse nun verpflichtend veröffentlicht. In den anderen Ländern wird hierauf nach wie vor verzichtet und an der Idee einer Steuerung über Einsicht (vgl. Böttger-Beer & Koch 2008) festgehalten. Wieder abgeschafft worden ist die Schulinspektion bislang nur in einem Bundesland: in Schleswig Holstein.

3. Schulinspektion als Instrument der schulischen Qualitätsentwicklung – Stand der Forschung

Zu der Frage der Auswirkungen von Schulinspektion auf betroffene Schulen sind im internationalen Kontext inzwischen einige Studien vorgelegt worden. Mit Blick auf die innerschulischen Entwicklungsaktivitäten infolge der Schulinspektion stellt sich die Erkenntnislage uneinheitlich dar (vgl. Dederling 2012; Ehren et al. 2013; Ehren et al. 2014; de Wolf & Janssens 2007; Husfeldt 2011). Dies gilt auch für die (wenigen)

Studien, die den Zusammenhang von Schulinspektion und Schülerleistungen in den Blick nehmen. Jene Studien, die sich auf Effekte der englischen Schulinspektion OFSTED (The Office for Standards in Education) beziehen, zeigen für das Gesamtsystem keine oder leicht negative Wirkungen auf Schülerleistungen als Folge von Schulinspektion (vgl. Cullingford & Daniels 1999; Rosenthal 2004; Shaw et al. 2003; Wilcox & Gray 1996). Eine neuere Studie aus den Niederlanden belegt hingegen, dass die Schulinspektion auf der Systemebene ggf. zu leichten Verbesserungen von Schülerleistungen führen kann (vgl. Luginbuhl et al. 2009). Es muss allerdings erwähnt werden, dass die vorliegenden Studien wegen methodischer Probleme (Stichprobenfehler) zumeist wenig belastbar sind (vgl. Pietsch et al. 2014).

Die Übertragbarkeit der internationalen Befunde auf den deutschen Kontext ist nun aufgrund von Unterschieden in der Steuerungsphilosophie der Länder ebenso wie in der Ausgestaltung der Verfahren nicht ohne Weiteres möglich. So sind beispielsweise in England weitreichende Konsequenzen für die Schulen an das schlechte Abschneiden bei der Schulinspektion geknüpft, wohingegen dies in Deutschland nicht der Fall ist (vgl. Dederling 2012). Aus diesem Grund wird der Forschungsstand nachfolgend nur für Deutschland dargestellt. Wichtig für die Einschätzung der Befunde ist der Sachverhalt, dass die Ergebnisse der entsprechenden Untersuchungen auf Selbstauskünften der beteiligten schulischen Akteure (insbesondere der Schulleitungen und Lehrkräfte) basieren, also nicht extern validiert wurden.

In den deutschen Untersuchungen wird überwiegend auf das Potenzial der Schulinspektion für die Qualitätsentwicklung der Schulen abgehoben. Die vorliegenden Studien untersuchen einerseits den *inerschulischen Verarbeitungsprozess* der zurückgemeldeten Inspektionsergebnisse und andererseits *inhaltliche Effektbereiche* – Bereiche, in denen Auswirkungen der zurückgemeldeten Inspektionsergebnisse wahrgenommen werden. Unter einem solchen Untersuchungsfokus entsteht dann bezüglich der intendierten Wirkungen der Schulinspektion ein auch gegenüber dem internationalen Forschungsstand zunächst doch recht positives Bild: Hinsichtlich der *inerschulischen Verarbeitungsprozesse* – und hier bezüglich der *Weitergabe der Inspektionsergebnisse* – attestieren alle vorliegenden Studien den meisten Schulen umfassende Aktivitäten der Kommunikation. Die wichtigsten am Schulleben beteiligten Personengruppen – Lehrkräfte, Schülerinnen und Schüler und Eltern – sind umfassend über die Inspektionsergebnisse informiert worden (vgl. Böttcher et al. 2010; Gärtner & Wurster 2009; Gärtner et al. 2009; Huber 2008).

Auch in Bezug auf die *Auseinandersetzung mit den Ergebnissen* sprechen die Befunde der empirischen Studien dafür, dass in der Mehrheit der Schulen Aktivitäten stattgefunden haben (vgl. Gärtner & Wurster 2009; Gärtner et al. 2009; Huber 2008; Kotthoff & Böttcher 2010; Wurster et al. 2013). Mit der Auswertung der Daten waren zumeist die Schulleitung und die Lehrerschaft betraut, wobei die Untersuchungen Hinweise auf unterschiedliche Verarbeitungs-konstellationen geben (u. a. Bildung von Arbeitsgruppen, Organisation Pädagogischer Tage, Auswertungsgespräche mit der Schulaufsicht bzw. dem Schulträger).

Hinsichtlich der *Maßnahmenplanung bzw. Maßnahmenumsetzung* attestieren die ersten vorgelegten Untersuchungen der Mehrheit der Schulen, aus der Analyse der Inspektionsergebnisse Schlussfolgerungen abgeleitet und Maßnahmen auf den

Weg gebracht zu haben (vgl. Böttcher et al. 2010; Dederig & Müller 2011; Gärtner & Wurster 2009; Gärtner et al. 2009; Huber 2008; Pietsch 2011; Sommer 2011) – und dies, obwohl ein Neuheitswert der Inspektionsergebnisse sowohl von Schulleitungen als auch von Lehrkräften überwiegend verneint wird (vgl. Böttcher et al. 2010; Gärtner & Wurster 2009). Derartige Maßnahmenplanungen brauchen allerdings in der Regel viel Zeit (vgl. Lambrecht et al. 2008); entsprechend ist das an den Schulen zum Zeitpunkt der Untersuchungsdurchführung vorfindbare Entwicklungsstadium – darauf weisen Gärtner und Wurster (2009) hin – abhängig vom jeweiligen Inspektionszeitraum: Schulen, deren Inspektion länger zurückliegt, sind im Planungs- und Umsetzungsstadium weiter vorangeschritten. Neuere Studien – die auch aus dem SteBis-Kontext stammen – kommen allerdings zu weniger positiven Ergebnissen in Bezug auf die Auswirkungen der Schulinspektion auf die Schul- und Unterrichtsentwicklung: So resümieren Böhm-Kasper und Selders (2013, siehe Beitrag in diesem Heft), dass die von den Schulleitungen wahrgenommenen Effekte der Schulinspektion eher marginal seien und sich im Zeitverlauf kaum veränderten. Auch Wurster et al. (2013) konstatieren in ihrer Untersuchung, dass ein beachtlicher Teil der in die Untersuchung einbezogenen Schulen (zwischen 33 % und 50 %) trotz eines diagnostizierten Entwicklungsbedarfs keine Maßnahmen ergreifen. Gärtner et al. (2014) weisen in ihrer als Längsschnitt auf Schulebene angelegten Studie mit inspizierten und nicht inspizierten Schulen (Kontrollgruppendesign) nach, dass die Schulinspektion einen vergleichsweise geringen Einfluss auf die Wahrnehmungen von Schulleitungen und Lehrkräften bezüglich einer Verbesserung von Schulqualität ausübt.

Wurster und Gärtner (2013) identifizieren fünf Typen von Schulen, in denen im Zuge der Schulinspektion jeweils unterschiedliche Wahrnehmungs- und Reaktionsprozesse beschrieben werden können. Sie differenzieren zwischen aktiven, reaktiven, (selbst-)zufriedenen, aktiven unzufriedenen und positiven unzufriedenen Schulen und tragen damit zur genaueren Beschreibung innerschulischer Verarbeitungsprozesse bei. Dabei nutzen nur zwei Typen (insgesamt 41 % der Schulen) die Inspektionsergebnisse: Die *aktiven Schulen*, die gute Inspektionsergebnisse erzielt haben, kommunizieren und reflektieren über diese in sehr umfassendem Maße und zeigen sowohl vor als auch nach der Schulinspektion viele Aktivitäten, etwa die Planung von Maßnahmen. Die *aktiven unzufriedenen Schulen*, die unterdurchschnittliche Inspektionsergebnisse zurückgemeldet bekommen haben, zeigen vor wie nach der Schulinspektion eine hohe Aktivität, äußern viele Veränderungswünsche und sind in der Maßnahmenplanung weit fortgeschritten. Bei den anderen drei Typen bleibt eine weiter gehende Ergebnismutzung hingegen aus.

Die Befunde zu den *inhaltlichen Effektbereichen* deuten darauf hin, dass es eine große Übereinstimmung zwischen den in den Evaluationsberichten formulierten Schwächen und Empfehlungen und den von den Schulen ergriffenen Maßnahmen zu geben scheint (vgl. Huber 2008). Doch Schulen – so die Befunde einer neueren Studie – nutzen die Inspektionsergebnisse auch in Bereichen, in denen ihnen kein akuter Handlungsbedarf attestiert worden ist (vgl. Wurster et al. 2013).

Es werden zudem prioritäre Bereiche der Planung und Umsetzung von Maßnahmen herausgestellt: Aktivitäten im Bereich der Schulprogrammarbeit, im Bereich der schulinternen Evaluation, und viele Schulleitungen berichten auch von Auswirkungen

auf den Bereich des Unterrichts in ihrer Schule – beispielsweise auf die Binnendifferenzierung (vgl. Dederich & Müller 2011; Gärtner et al. 2009; Wurster et al. 2013).

Erst in der jüngsten Vergangenheit rückt vereinzelt der Zusammenhang von Schulinspektion und Leistungsentwicklung der Schüler in den Fokus der Aufmerksamkeit (vgl. Pietsch et al. 2013; Pietsch et al. 2014). Für die Schulinspektion in Hamburg wurde mithilfe von Trenddaten des Hamburger Zentralabiturs und Längsschnittdaten der Studie „Kompetenzen und Einstellungen von Schülerinnen und Schülern“ (KESS) (zwei Messzeitpunkte) und auf der Basis kausalanalytischer Verfahren gezeigt, dass Effekte auf Lernzuwächse sowie Leistungstrends von Schülern – insbesondere im Fach Deutsch (Leseverständnis), aber auch im Fach Mathematik – nachgewiesen werden können (vgl. Pietsch et al. 2014).

Zum Zusammenhang der spezifischen Situation von Schulen und der Art der Rezeption und Verarbeitung der Ergebnisse der Schulinspektion in Schulen liegen bisher keine Erkenntnisse vor. In dieser Hinsicht intendiert der Aufsatz, einen Beitrag zur Schließung dieser Forschungslücke zu leisten.

4. Die eigene Studie

In den Jahren 2010 bis 2013 ist – gefördert vom Bundesministerium für Bildung und Forschung (BMBF) und angesiedelt im Förderschwerpunkt *SteBis* im Rahmenprogramm zur Förderung der *Empirischen Bildungsforschung* – eine Studie¹ durchgeführt worden, die an den theoretischen und empirischen (Forschungs-)Stand von vor vier Jahren anknüpfte. Diese Studie wird im Folgenden hinsichtlich ihrer Fragestellung und ihres theoretischen Hintergrundes (Kapitel 4.1) sowie ihres methodischen Ansatzes und ihrer Durchführung (Kapitel 4.2) dargestellt.

4.1 Fragestellung und theoretischer Hintergrund

Im Mittelpunkt der Untersuchung stand das Verfahren der Schulinspektion mit seinen Auswirkungen auf die innerschulische Entwicklungsarbeit. Dementsprechend ging die Studie der übergeordneten Frage nach, inwiefern die Schulinspektion einen Einfluss auf Aktivitäten hat, die der pädagogischen und organisatorischen Weiterentwicklung von Schulen dienen. Schulentwicklung wurde dabei als dauerhafte, bewusste und systematische Weiterentwicklung der Einzelschule mit dem Ziel einer Verbesserung ihrer Qualität verstanden, die sich inhaltlich auf die Bereiche der Organisation (Organisationsentwicklung), der Lernkultur (Unterrichtsentwicklung) und des Personals (Personalentwicklung) beziehen kann.

Das Ziel der Untersuchung bestand darin, Erkenntnisse über die Prozesse in Schulen zu generieren, die durch die Schulinspektion ausgelöst werden, und darüber zu einer Einschätzung zu gelangen, inwiefern diese Verfahren über die Förderung

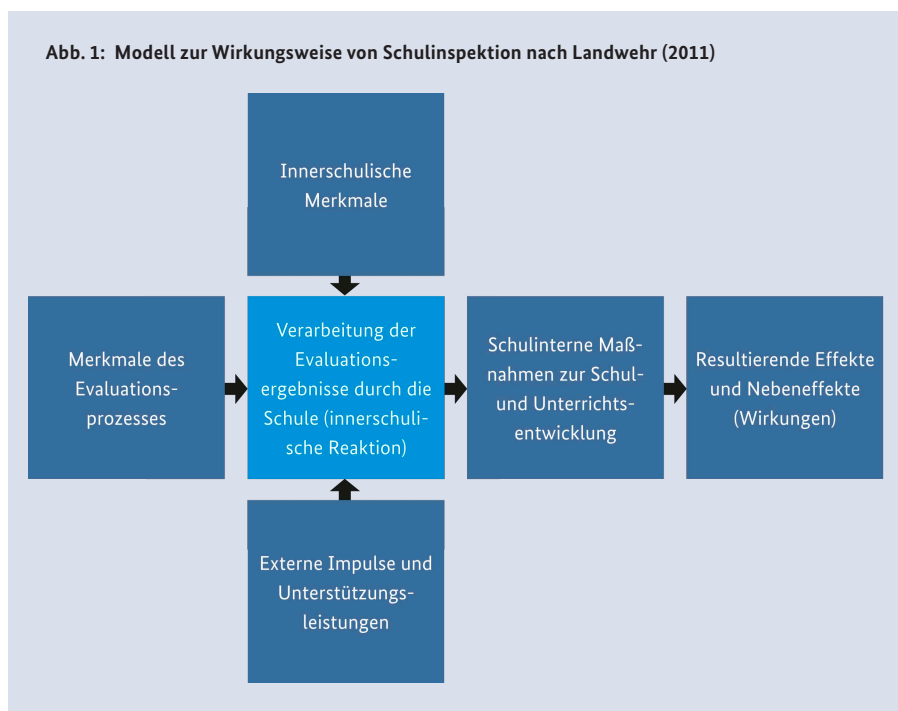
1 An der Studie mit dem Titel „Externe Evaluation/Schulinspektion und schulische Qualitätsentwicklung“ waren Kathrin Dederich (Leitung), Nina Fritsch und Christian Weyer beteiligt.

innerschulischer Diskussions- und Entwicklungsaktivitäten zur Qualitätsentwicklung beitragen. Zudem wurde mit der Studie intendiert, Informationen über die Einflusskraft verfahrensspezifischer, schulischer und außerschulischer Faktoren auf die innerschulische Entwicklung bereitzustellen.

Vor dem Hintergrund einer Vorstellung des Schulsystems als Mehrebenensystem (vgl. Altrichter & Maag Merki 2010) lag der Fokus der Betrachtung auf der Ebene der Einzelschule (Mesoebene). Im Mittelpunkt standen demnach die Aktivitäten und Handlungsweisen der Schulleitungen, Lehrkräfte, nicht pädagogischen Mitarbeiter sowie ggf. Schüler und Eltern – wenngleich diese in ihrer Verbindung zu der Ebene der Schulministerien und Schulaufsicht (Makroebene) und deren Aktivitäten zur Schaffung von Rahmenbedingungen für die schulische Arbeit und der Ebene der Unterstützungssysteme, also Berater, andere Schulen etc. (intermediäre Ebene), gesehen wurden.

Dabei wurde mit einem theoretischen Rahmenmodell gearbeitet, das auf dem von Ehren und Visscher (2006) konzipierten Modell zu der Wirkungsweise von Schulinspektion basiert. Dieses Modell bildet das Beziehungsgeflecht ab, das – im Sinne von Kausalketten – zwischen *Merkmale des Evaluationsprozesses, Reaktionen der Schulen auf die Ergebnisse der externen Evaluation* und daraus resultierenden *Effekten und Nebeneffekten* bestehen. Es nimmt mit *innerschulischen Merkmalen* sowie den *externen Impulsen und Unterstützungsleistungen* zwei Gruppen von Faktoren auf, die die innerschulischen Aktivitäten infolge von Schulinspektion beeinflussen (zur Beschreibung dieses Modells siehe den Beitrag von Böhm-Kasper & Selders in diesem Band). Mit ihm ist insbesondere deshalb ein weiterführender Beitrag zur Thematik geleistet worden, weil erstmals Variablen zur Beschreibung der Wirksamkeitsannahmen der Schulinspektion zusammengestellt worden sind (vgl. Pietsch et al. 2014). Nichtsdestotrotz ist es einer kritischen Betrachtung unterzogen worden: Husfeldt (2011) merkt an, dass der Zusammenhang zwischen den Variablen unklar sei – etwa zwischen den ausgelösten Maßnahmen und den Schülerleistungen. Sie moniert zudem, dass das Modell keine Informationen über eine sinnvolle Weiterarbeit der Schulen mit den Ergebnissen enthalte. Von ihr kritisiert wird auch die Beschränkung des Modells auf die Funktion der Schulentwicklung, die nur eine von mehreren Funktionen der Schulinspektion darstelle. Weitere Anmerkungen in der Literatur beziehen sich auf die fehlende Berücksichtigung des Mehrebenencharakters des Schulsystems, der eine Verknüpfung der individuellen Wahrnehmungs- und Handlungserträge der verschiedenen Akteure erforderlich mache, sowie auf fehlende Informationen zu den Handlungsbeiträgen der einzelnen Akteure im Prozess der Rezeption und Verarbeitung der Inspektionsdaten (vgl. Altrichter 2010; Pietsch et al. 2014)². Auch eine gewisse Unterkomplexität bei der Beschreibung des innerschulischen Verarbeitungsprozesses wurde ihm unterstellt. Diesen letztgenannten Punkt aufgreifend, wurde der eigenen Untersuchung ein von Landwehr (2011) vorgelegtes Modell zugrunde gelegt (vgl. Abbildung 1). In ihm wird davon ausgegangen, dass

2 Jüngere Entwicklungen im theoretisch-konzeptionellen Bereich greifen diese Kritik auf (vgl. Ehren et al. 2013) und beinhalten zudem neue Perspektiven (vgl. u.a. Ehren et al. 2014; Grek et al. 2013). Sie konnten im Rahmen unseres Forschungsprojekts nicht berücksichtigt werden, sollen hier jedoch zumindest Erwähnung finden.



der schulinterne Prozess zur Verarbeitung der Schulinspektionsergebnisse durch die Schulleitung und das Kollegium einen entscheidenden Einfluss auf die Wirksamkeit der Schulinspektion hat und gewissermaßen die „Nahtstelle zwischen Evaluation und Schulentwicklung“ (Landwehr 2011, S. 50) darstellt. Der Akzeptanz der Ergebnisse kommt hier eine intermittierende Funktion zu; sie hat Auswirkungen auf die innerschulische Rezeption und Reflexion der Evaluationsdaten und die Ableitung von Entwicklungsaktivitäten aus ihnen. Dementsprechend unterteilt Landwehr (2011) die Phase der *Reaktionen der Schulen auf die Ergebnisse der externen Evaluation* des Modells von Ehren und Visscher (2006) in seinem Modell in die Phase der *Verarbeitung und Akzeptanz der Inspektionsergebnisse durch die Schule* und die Phase der *Planung und Umsetzung von Maßnahmen zur Schul- und Unterrichtsentwicklung*. Darüber hinaus wurde das Modell im Projekt um ein innerschulisches Merkmal – das Schulleitungshandeln – ergänzt. Damit wird den Befunden einer zunehmenden Zahl empirischer Studien entsprochen, denen zufolge der Schulleitung bei der Qualitätssicherung und Qualitätsentwicklung eine Schlüsselrolle zukommt (vgl. Wissinger 2011). Für das Schulleitungshandeln wird ein messbarer, meist direkter Einfluss auf die Lehr- und Lernqualität nachgewiesen (vgl. z. B. Bonsen et al. 2002; Louis et al. 2010).

Auch ein externer Impuls – bildungspolitische Vorgaben zur Umsetzung aktueller Reformvorhaben, die Schulen bei ihrer Arbeit berücksichtigen müssen – wurde in der eigenen Studie in das Modell eingefügt. Hier wurde vermutet, dass die Zeit und die Ressourcen, die zur Umsetzung der Vorgaben erforderlich sind, für die Verarbeitung der Inspektionsergebnisse nicht zur Verfügung stehen und dass die Vorgaben somit einen (negativen) Einfluss auf den Prozess ausüben.

4.2 Methodischer Ansatz und Durchführung der Studie

Zunächst werden nun der methodische Ansatz und die Durchführung der Studie knapp skizziert. Anschließend werden ausgewählte Ergebnisse präsentiert.

4.2.1 Darstellung im Überblick

Die Untersuchung wurde exemplarisch in zwei Bundesländern – Hamburg³ und Thüringen – durchgeführt, die sich in Bezug auf ihre Verfahren der Schulinspektion in einigen Punkten unterschieden. So werden z. B. die *Inspektionsteams* in Hamburg aus hauptamtlich als Inspektoren tätigen wissenschaftlichen Bildungsexperten, in Thüringen hingegen aus nebenamtlich als Evaluatoren arbeitenden Schulleitungen und Lehrkräften gebildet. Die Dauer des Inspektionszyklus umfasst in Hamburg 6 Monate, während sie sich in Thüringen auf 24 Monate beläuft. Dabei ist der unterschiedliche zeitliche Umfang auch mit inhaltlichen Unterschieden verbunden: In Thüringen werden die Schulen bis April eines Jahres vom Staatlichen Schulamt über den anstehenden Schulbesuch informiert. Bevor ab August – im neuen Schuljahr also – der Schulbesuch stattfindet, sind die Schulen angehalten, auf der Basis von Checklisten eine schulinterne Diskussion zur Selbsteinschätzung vorzunehmen. Dieses Selbstbild wird mit der Einschätzung des zuständigen Referenten des Staatlichen Schulamtes – und damit mit einem Fremdbild – in Verbindung gebracht, beide Perspektiven werden im Evaluationsbericht berücksichtigt. In Hamburg findet ein derartiger Perspektivenabgleich im Vorfeld des Schulbesuchs nicht statt.

Die Rückmeldeformate beinhalten in Hamburg ein standardisiertes Berichtswesen ohne Handlungsempfehlungen für die Schulen, wohingegen in Thüringen auf eine Standardisierung der Inspektionsberichte verzichtet wird, Handlungsempfehlungen jedoch vorgesehen sind (vgl. Dederig et al. 2014). Diese Unterschiede führten zur Formulierung einer Reihe von Hypothesen – etwa jener, dass Rückmeldeformate mit Handlungsempfehlungen mehr Qualitätsentwicklungsaktivitäten in den Schulen bedingen als Rückmeldeformate ohne Handlungsempfehlungen. Aufgrund der spezifischen thematischen Ausrichtung des vorliegenden Beitrags wird hier nicht weiter auf sie eingegangen.

Um die Aktivitäten, die im Zuge der Schulinspektion in Schulen stattfinden, in der Breite zu erfassen und somit einen Überblick zu erhalten, wurden alle allgemeinbildenden Schulen dieser beiden Länder einbezogen, die im ersten Halbjahr des Jahres 2011 einer externen Begutachtung unterzogen wurden (n = 84). Entwickelt wurde ein Quasi-Längsschnitt-Design (Schulebene), bei dem eine Erhebung von Daten zu drei unterschiedlichen Zeitpunkten stattfand: zwei Wochen vor dem obligatorischen Schulbesuch (Erhebungszeitpunkt [EZP] 1), vier Monate nach der Schulinspektion (EZP 2) und zwölf Monate nach dem Schulbesuch (EZP 3). De facto handelte es sich bei den Erhebungszeitpunkten um Erhebungszeiträume, die – den individuellen

3 Die Untersuchung bezieht sich auf den ersten Zyklus der Schulinspektion in Hamburg. Seit 2013 werden alle Schulen des Bundeslandes bereits zum zweiten Mal inspiziert – und zwar auf der Basis eines überarbeiteten Verfahrens.

Terminen der Schulbesuche im ersten Halbjahr entsprechend – jeweils sechs Monate umfassten (vgl. Abbildung 2). Prozessbegleitend wurden die Schulleitungen und Lehrkräfte mit besonderen Funktionen im Bereich der Schulentwicklungsarbeit befragt.

Die Inhalte der ersten standardisierten Befragung bezogen sich (1) auf innerschulische Merkmale (Erfahrungen im Bereich der Schulentwicklung, Kommunikations-, Kooperations- und Partizipationsstrukturen etc.), (2) auf Merkmale des außerschulischen Umfeldes (z. B. Legitimationsdruck durch die Öffentlichkeit) und (3) auf Informationen zur Vorbereitung des Schulbesuchs und zur Bewertung des Verfahrens. Befragt wurden Schulleitungen und Lehrkräfte mit besonderen Funktionen in der Schulentwicklungsarbeit. Die zweite Befragung umfasste inhaltlich (1) die Durchführung der Schulinspektion und die Ergebnisrückmeldung, (2) die Art des Inspektionsergebnisses, (3) die Rezeption und Verarbeitung der Inspektionsergebnisse und (4) das außerschulische Umfeld (z. B. externe Unterstützungsangebote). Sie richtete sich in einem ersten Teil an alle Lehrkräfte, in einem zweiten Teil nur an die oben genannten Akteure. Die dritte Befragung schließlich bezog sich auf die Schwerpunkte der innerschulischen Entwicklungsarbeit zwölf Monate nach der Schulinspektion und berücksichtigte wiederum nur die Schulleitungen und Lehrkräfte auf Funktionsstellen. Zwar wäre es wünschenswert gewesen, bei allen Befragungen alle Lehrkräfte und darüber hinaus weitere Akteure wie Eltern und Schüler einzubeziehen, im Rahmen des Projekts war jedoch eine Fokussierung auf ausgewählte Akteure erforderlich. Die Befragten wurden jeweils um Angaben auf einer vierstufigen Antwortskala (stimme zu, stimme eher zu, stimme eher nicht zu, stimme nicht zu bzw. trifft zu, trifft eher zu, trifft eher nicht zu, trifft nicht zu) gebeten.

Pro Schule wurden bis zu zehn Personen einbezogen, wobei die (auf die inspizierten Schulen bezogenen) insgesamt geringen Rücklaufquoten über die Erhebungszeitpunkte hinweg noch abnahmen. Tabelle 1 gibt Auskunft über den Rücklauf der einzelnen Erhebungszeitpunkte nach Ländern.

Land	Stichprobe (Schulen)	Rücklauf EZP 1 (Schulen/Lehrkräfte)	Rücklauf EZP 2 (Schulen/Lehrkräfte)	Rücklauf EZP 3 (Schulen/Lehrkräfte)
Hamburg	38	27/151	23/246	19/72
Thüringen	46	18/104	12/104	14/58
Gesamt	84	45/255	35/350	33/130

Es wird deutlich, dass dieser insgesamt von 53,6 % bei EZP 1 auf 39,1 % bei EZP 3 zurückging. Ein systematischer Ländervergleich war dadurch – insbesondere auf der Ebene einzelner Fragebogenitems – erschwert bis nicht zu realisieren.



Um die Prozesse der Auseinandersetzung mit der Schulinspektion und deren Ergebnissen genauer zu analysieren, bestand für Hamburg die Möglichkeit, zumindest in geringem Umfang zusätzlich retrospektiv angelegte Fallstudien auf Basis qualitativer Interviews an ausgewählten Schulen vorzunehmen (vgl. Abbildung 2). Auf diese Fallstudien und ihre Ergebnisse wird nachfolgend rekurriert. Befunde der standardisierten Befragungen werden für die Fallschulen unterstützend einbezogen.

4.2.2 Im Fokus: Fallstudien an ausgewählten Schulen

In unserer Studie stellte der *innerschulische Verarbeitungsprozess der Schulinspektion* den zu analysierenden „Fall“ dar. Dieser sollte möglichst umfassend in die Analyse einbezogen werden (vgl. Fatke 2010). Das Vorgehen in der eigenen Untersuchung gliederte sich dabei in zwei Schritte: In einem *ersten* Schritt wurden die spezifischen Bedingungen und Abläufe an den Einzelschulen chronologisch nachgezeichnet. Auf der Basis der fallspezifischen Erkenntnisse wurde in einem zweiten Schritt ein Vergleich der Einzelfälle vorgenommen. Das Ziel bestand darin, Gründe für die spezifische Ausgestaltung von Verarbeitungsprozessen zu identifizieren. Aus der Gruppe jener Schulen, die an der standardisierten Befragung vier Monate nach dem obligatorischen Schulbesuch (EZP 2) teilgenommen haben, wurden vier Schulen ausgewählt. Der konzeptionelle Ansatz der Schulauswahl sah dabei zunächst zwei Kriterien vor, nach denen systematisch unterschieden werden sollte: 1. die Art des Inspektionsergebnisses und 2. die Schulstufe. Mit dem Kriterium der *Art des Inspektionsergebnisses* war die Annahme verbunden, dass überwiegend positive Bewertungen der schulischen Qualitätskriterien an den Schulen zu geringer ausgeprägten innerschulischen Entwicklungsaktivitäten führen als überwiegend mittelmäßige und negative Bewertungen. Neuere empirische Befunde weisen nun darauf hin, dass diese Annahme nicht für alle Schulen gilt (vgl. Wurster & Gärtner 2013; Wurster et al. 2013; siehe hierzu Kapitel 3): So werden auch gute Ergebnisse mitunter von Schulen intensiv genutzt (aktive Schulen), andererseits führen weniger gute Ergebnisse in einigen Schulen auch nicht zu Aktivitäten – etwa dann, wenn der Nutzen als gering betrachtet wird (reaktive Schulen) oder das Ergebnis als so schlecht wahrgenommen wird, dass

Resignation und Passivität eintreten (passive unzufriedene Schulen). Die explizierte Annahme stellt nach Ansicht des Forschungsteams aber dennoch einen angemessenen Ausgangspunkt dar. Die Wahl des Kriteriums der *Schulstufe* begründete sich durch die Vermutung, dass die Schulstufe oftmals mit der Schulgröße konfundiert. Es wurde also davon ausgegangen, dass es sich bei Primarschulen eher um kleinere Einrichtungen handelt als bei Sekundarschulen und dass dadurch an den Schulen jeweils unterschiedliche Kontexte für die Verarbeitung der Ergebnisse der Schulinspektion bestehen.

In der standardisierten Befragung vier Monate nach dem Schulbesuch waren die Schulleitungen und Lehrkräfte auf Funktionsstellen in der Schulentwicklung um Angaben zu den Inspektionsergebnissen ihrer Schule gebeten worden. Für die zwölf Qualitätsbereiche, die im Inspektionsbericht bewertet wurden – z. B. *Führung wahrnehmen, Beratungsangebote gestalten, Die Schulgemeinschaft beteiligen* – sollten die Befragten jeweils die Bewertungsstufe der Inspektion angeben (Stark [4] – Eher stark als schwach [3] – Eher schwach als stark [2] – Schwach [1]). Da die überwiegende Mehrheit der Schulen ihre Inspektionsberichte nicht für eine Analyse zur Verfügung stellte, war diese Art der Abfrage die einzige Möglichkeit, Informationen über die Inspektionsergebnisse zu erhalten. Aus den Qualitätsbereichen wurde eine Skala gebildet, deren Mittelwert als Orientierungspunkt für die Einschätzung der Inspektionsergebnisse als gut bzw. weniger gut diente. Jene Schulen, die entsprechende Angaben gemacht haben, hatten insgesamt gute bis mittelmäßige Ergebnisse erzielt. Daher wurden für die Fallstudien zwei Schulen mit eher guten und zwei Schulen mit mittelmäßigen Ergebnissen ausgewählt. Da sich nur sehr wenige Schulen überhaupt zu einer Teilnahme an der Untersuchung bereit erklärten, konnte das Kriterium der *Schulstufe* nicht in der zuvor beschriebenen Systematik berücksichtigt werden. In die Untersuchung einbezogen werden konnten eine Primarschule und drei Sekundarschulen. Bei allen vier Schulen handelte es sich aber um Einrichtungen von (mindestens) mittlerer Größe (mit einer Schülerzahl von 500 und mehr) mit (mindestens) mittelgroßen Lehrerkollegien (mit einer Lehrerzahl von 50 und mehr). Damit fand sich in ihnen ein ähnlicher Kontext für die Verarbeitung der Ergebnisse der Schulinspektion, und unsere Schulauswahl weist – wenn auch anders als geplant – eine Systematik auf.

Im Rahmen der Fallstudien wurden qualitative Interviews durchgeführt, und es wurde auf die individuellen Ergebnisse der vier Schulen aus den standardisierten Befragungen zurückgegriffen, um die Fallbeschreibungen an einigen Stellen zu ergänzen. Als Interviewpartner fungierten 1. Mitglieder der Leitungen und Steuergruppen der Schulen, 2. die für die Schulen jeweils zuständigen Schulinspektoren und 3. die für die Schulen jeweils zuständigen Schulaufsichtsvertreterinnen und -vertreter. Durch die Einbeziehung dieser Personen, die aus ihren unterschiedlichen Positionen und Verantwortungen im System Schule heraus über die Verarbeitungsprozesse berichteten, konnte der Prozess jeweils mehrperspektivisch nachgezeichnet werden. Pro Fall haben wir drei bis fünf Interviews auf Basis von Leitfäden durchgeführt, die zwischen 45 und 150 Minuten umfassten.

Die Auswertung der transkribierten Interviewpassagen erfolgte computergestützt anhand des Textverarbeitungsprogramms MAXQDA. Mit dem Ziel einer Strukturie-

zung der Datenmenge wurde – der Basistechnik des Textverarbeitungsprogramms entsprechend – ein hierarchisch aufgebautes Kategoriensystem erstellt, dessen Elementen (Codes) Textsegmente zugeordnet wurden. Das Vorgehen bei der Kategorienbildung war deduktiv wie induktiv. In einem ersten Schritt wurden Kategorien aus Aspekten des Interviewleitfadens, die auf der Ausgangsfrage basierten, hergeleitet. In einem zweiten Schritt erfolgte die Definition von Kategorien aus den Daten selbst. Bei der Auswertung des codierten Materials wurde eine Technik verwendet, die sich an das Verfahren der zusammenfassenden Interpretation der qualitativen Inhaltsanalyse anlehnt (vgl. Lamnek 2010; Mayring 2010).

5. Befunde zum schulischen Umgang mit Ergebnissen der Schulinspektion

An den vier näher betrachteten Schulen sind die Ergebnisse der Schulinspektion verbreitet und – zumindest von den Schulleitungen und Lehrkräften mit besonderen Funktionen im Bereich der Schulentwicklung – rezipiert worden. Für die Schulentwicklung nutzbar gemacht worden sind sie aber nur in begrenztem Maße. Aus den Prozessbeschreibungen – die hier nicht allesamt nachgezeichnet werden können (vgl. hierzu Dederig et al. 2014) – lassen sich einige Faktoren ableiten, die sich begünstigend bzw. hemmend auf die Ergebnisverarbeitung auswirken: Vor dem Hintergrund der in Kapitel 4.1 explizierten theoretischen Modelle, die der eigenen Studie zugrunde gelegt wurden, kommt erstens zwei Faktoren eine Bedeutung zu, die dem Bereich *Verarbeitung* zugeordnet werden können: 1. die Akzeptanz des Verfahrens der Schulinspektion insgesamt bzw. seiner Komponenten (z. B. des Handlungsrahmens Schulqualität) durch die Schulleitung und die Lehrkräfte sowie 2. deren Akzeptanz der schulspezifischen Inspektionsergebnisse. Zweitens lassen sich einige Faktoren identifizieren, die *inerschulische Merkmale* darstellen: Hier sind 1. die Organisationsstruktur der Schulentwicklungsarbeit, 2. die innovative Orientierung der Lehrer und 3. das Schulleitungshandeln anzuführen. Schließlich sind als *externe Impulse* bildungspolitische Vorgaben zu nennen. Konkret geht es um die an Schulen gerichtete Forderung, die Idee der inklusiven Schule umzusetzen und die Schulorganisation auf einen Ganztagsbetrieb umzustellen.

Die Befunde deuten dabei darauf hin, dass einem innerschulischen Merkmal – der *Organisationsstruktur der Schulentwicklungsarbeit* – eine besondere Bedeutung zukommt. Um diesen Sachverhalt zu veranschaulichen, seien die Verarbeitungsprozesse an zwei Schulen detaillierter beschrieben. Tabelle 2 gibt zunächst einen Überblick über die Ausprägungen relevanter Merkmale an allen vier Fallschulen.

Tabelle 2: Relevante Kriterien des hier vorgenommenen Fallvergleichs

Kriterien	Schule Alpha	Schule Delta	Schule Sigma	Schule Omega
Ausmaß der Aktivitäten der Qualitätsentwicklung	hoch	hoch	gering	gering
Grad der Organisation der Aktivitäten der Qualitätsentwicklung	hoch	weniger hoch	weniger hoch	weniger hoch
Existenz eines Konzepts für Schulentwicklung	ja	nein	nein	nein
Existenz einer Steuergruppe	ja	nein	nein	nein
Ausmaß der Innovationsorientierung der Lehrkräfte	hoch	hoch	hoch	gering
Ausmaß der Einbeziehung des Kollegiums in Aktivitäten der Qualitätsentwicklung	hoch	hoch	weniger hoch	weniger hoch
Teilnahme an systematischen Entwicklungsprojekten	ja	nein	nein	nein
Systematische/Strukturierte Verarbeitung der Ergebnisse der Schulinspektion	ja	nein	nein	nein

Für den nachfolgenden Fallvergleich werden die Schulen *Alpha* und *Delta* gewählt, weil sie bei einem hohen Grad der Aktivität und Aufgeschlossenheit im Bereich der Qualitätsentwicklung deutliche Unterschiede hinsichtlich der Organisation dieses Bereichs ebenso wie bezüglich der Verarbeitung der Ergebnisse der Schulinspektion aufweisen.

Dass dabei ausschließlich Aussagen der schulischen Akteure, nicht aber der Schulinspektoren und Schulaufsichtsvertreter ins Feld geführt werden, ist darauf zurückzuführen, dass Letztere keine Angaben zu den hier beschriebenen innerschulischen Prozessen gemacht haben.

Schule Alpha

Bei Schule *Alpha* handelt es sich um eine Sekundarschule mittlerer Größe (500 bis 1000 Schüler) mit einem mittelgroßen Kollegium (50 bis 75 Lehrkräfte). Die derzeitige Schulleitung ist seit mehr als zehn Jahren an der Schule. Die Schülerschaft stammt überwiegend aus bildungsnahen und sozial gut gestellten Kontexten.

Bereits seit mehreren Jahren betreibt die Lehrerschaft vielfältige Aktivitäten im Bereich der Qualitätsentwicklung und -verbesserung. Den Ergebnissen unserer standardisierten Befragung zum ersten Erhebungszeitpunkt zufolge – an der von dieser Schule neun Mitglieder der Schulleitungen und Lehrkräfte mit besonderen Funktionen in der Schulentwicklung teilgenommen haben – finden diese Aktivitäten

nicht überwiegend als unverbundene Einzelmaßnahmen statt; es existiert ein Konzept mit konkreten Maßnahmen und Entwicklungsschwerpunkten sowie eindeutig geregelten Zuständigkeiten für Aufgaben der Schulentwicklung. Die hierin zum Ausdruck kommende Organisation der Schulentwicklungsaktivitäten ist nach Aussage der Befragten insgesamt überdurchschnittlich gut ($M = 3.90$; $p < .05$)⁴. Auch die Organisation der Arbeit generell ist an dieser Schule zufriedenstellend ($M = 3.10$). Die Lehrkräfte gestalten ihre Schule also nach gemeinsamen pädagogischen Zielen, und alle Befragten geben an, dass das Schulprofil gemeinsam erarbeitet wird. Auch dass der Informationsfluss innerhalb des Kollegiums gut ist, bestätigen alle Befragten.

Die Schule hat in den vergangenen Jahren an systematischen Schulentwicklungsprojekten teilgenommen, insbesondere an einem Netzwerkprojekt, in dem die Schulleitungen und Funktionsträger im Bereich der Schulentwicklung hinsichtlich des Schulmanagements geschult worden sind. Vor einigen Jahren hat die Schulleitung zudem einen externen Berater hinzugezogen, um mit dessen Unterstützung ein Leitbild für die schulische Arbeit zu entwickeln. Gleichzeitig wurde ein Organisationskreislauf entwickelt, der der Schulentwicklungsarbeit seither eine Struktur gibt (s. u.).

Leitung und Lehrerschaft an dieser Schule sind Fragen und Themen der Schulentwicklung und Evaluation gegenüber sehr aufgeschlossen ($M = 3.11$; $p < .05$). Vorbehalte gegenüber Veränderungen bestehen nicht, und auch eine mangelnde Bereitschaft, die eigenen pädagogischen Ansätze zu überprüfen, wird nicht konstatiert.

Die Schulentwicklungs- und Evaluationsaktivitäten werden von einer Schulleitung initiiert bzw. unterstützt, die von sich selbst behauptet, keine Leitung zu sein, die Verordnungen „von oben nach unten weitergibt“ (I-SL_Alpha). Vielmehr bildet diese mit der stellvertretenden Schulleitung ein Team, das sich im ständigen Gespräch mit den Kollegen befindet. Dass die Lehrkräfte an ihrer Schule in die Planung von Maßnahmen der Schulentwicklung aktiv einbezogen werden, bestätigt dementsprechend die Mehrheit der Befragten. Sechs von neun Befragten bekräftigen, dass die Vorstellungen der Lehrkräfte bei schulischen Entwicklungsprozessen ausreichend einbezogen werden. Zudem existiert zur Koordination der Aktivitäten an der Schule eine Steuergruppe.

Es wurde bereits erwähnt, dass sich die Schule bei ihrer Entwicklungsarbeit an einen Organisationskreislauf anlehnt, der am Schuljahresrhythmus orientiert ist: Jeweils zum Schuljahresende (April/Mai) werden in den Gremien (Schülerrat, Elternrat, Lehrerkonferenz, Schulkonferenz) Themen auf einem von der Steuergruppe durchgeführten Jour fixe festgesetzt, die als relevant für die Schulentwicklung des nächsten Jahres erachtet werden. Auf dieser Basis erarbeitet die Steuergruppe vier bis fünf Themen, die dem Kollegium im Mai auf einer Bilanzkonferenz vorgestellt werden. Die Ergebnisse der Diskussion über diese Themen werden festgehalten und in der Steuergruppe aufgearbeitet, bevor auf einer Lehrerkonferenz noch einmal über die Themen abgestimmt wird. Schließlich muss die Schulkonferenz die Themen akzeptieren:

4 Die angegebenen Mittelwerte basieren auf den vorab beschriebenen vierstufigen Antwortskalen. Es wurde ein t-Test durchgeführt.

„Und dadurch, dass wir in den Gremien Elternrat, Schülerrat und Lehrerkonferenz die Transparenz des Entwicklungsprozesses haben, von April bis Mai, haben wir es bisher so gehabt, dass die Schulkonferenz immer einstimmig zu den Entwicklungsthemen zugestimmt hat“ (I-SL_Alpha).

Zu Beginn des neuen Schuljahres setzt dann die Arbeit an verschiedenen Projekten ein.

Die Ergebnisse der Schulinspektion wurden der Schule acht Wochen nach dem Schulbesuch im Juli zurückgemeldet. Dadurch bestand die Möglichkeit, sie in die laufende Schulentwicklungsarbeit einzubeziehen. In drei Qualitätsbereichen ist die externe Bewertung dabei nicht so gut ausgefallen, vor allem die Leistungsbeurteilung wurde als wenig transparent und uneinheitlich eingestuft. Dass in diesem Bereich ein erhöhter Handlungsbedarf insofern besteht, dass an einheitlichen Formen der Leistungsrückmeldung gearbeitet werden muss, wurde vom Eltern- und vom Schülerrat ebenso gesehen wie von der Lehrerkonferenz. Die Verbesserung der Leistungsbewertung wurde deshalb – zusätzlich zu den im Mai vereinbarten Themen – als prioritäres Thema der Schulentwicklungsarbeit der kommenden drei Jahre festgesetzt. Wie bei den anderen Themen auch wurde eine Arbeitsgruppe gebildet, in der verbindlich festgelegt wurde, welche Personen sich in regelmäßigen Abständen zur Entwicklung eines Rückmeldekonzepes mit welchen Zwischenzielen treffen. Da die Strukturen, die an der Schule in der Entwicklungsarbeit schon existierten, ein schnelles und unkompliziertes Vorgehen bei der Bearbeitung des erkannten Defizits ermöglichten, kann hier davon gesprochen werden, dass die Ergebnisse an Schule Alpha in die laufende Schulentwicklungsarbeit eingespeist worden sind.

Schule Delta

Bei Schule Delta handelt es sich um eine Primarschule mittlerer Größe (500 bis 750 Schüler) mit einem mittelgroßen Kollegium (50 bis 75 Lehrkräfte), der die derzeitige Schulleitung seit mehr als zehn Jahren vorsteht. Die Schülerschaft stammt überwiegend aus bildungsfernen und sozial weniger gut gestellten bis prekären Kontexten.

Auch an dieser Schule treffen wir auf eine Lehrerschaft, die im Bereich der Schulentwicklung und Evaluation schon seit mehreren Jahren sehr engagiert ist. Anders als an der zuvor dargestellten Sekundarschule Alpha laufen diese Aktivitäten aber weniger systematisiert ab. So merkt die Schulleitung an:

„Wir sind eine Schule, die sehr viel Schulentwicklung macht, tendenziell zu viel. Also, jeder, der ein tolles Projekt hat, setzt das hier auch um, das ist so ein bisschen [die] Schulkultur. Es ist ganz schwer, mal ‚Nein‘ zu sagen, und das so ein bisschen einzugrenzen, das ist eher unsere Aufgabe“ (I-SL_Delta).

Die Ergebnisse unserer ersten standardisierten Befragung zwei Wochen vor dem Schulbesuch – an der von dieser Schule zehn Mitglieder der Schulleitung und Lehrkräfte auf Funktionsstellen teilgenommen haben – bestätigen dies. Betrachtet man die Einschätzungen der Lehrkräfte von Schule Delta zur Organisation der Schulentwicklung im Vergleich zu denen aller Lehrerinnen und Lehrer, deren Schulen im ersten Halbjahr 2011 extern begutachtet worden sind, so zeigt sich, dass sie weniger positiv sind ($M = 3.19$; $p < .05$). Auch die allgemeine Organisation der Arbeit verläuft – den Angaben der Befragten entsprechend – weniger strukturiert ($M = 2.93$; $p < .05$).

So sind einige Befragten der Meinung, dass keine gemeinsame pädagogische Zielsetzung an der Schule vorgenommen und nicht gemeinsam am Profil der Schule gearbeitet wird. Dass der Informationsfluss innerhalb des Kollegiums dennoch gut ist, bestätigten bis auf eine Person alle Befragten. Ebenso fällt das Ergebnis bei der Einbindung von Lehrkräften in die Planung von Maßnahmen der Schulentwicklung und der Berücksichtigung ihrer Vorstellungen bei schulischen Entwicklungsprozessen aus: Mit einer Ausnahme sind sich alle Befragten einig, dass diese an der Schule beobachtbar sei. Dementsprechend werden die Lehrerinnen und Lehrer an Schule *Delta* stark in die Schulentwicklungsarbeit einbezogen ($M = 3.45$; $p < .05$).

Laut Angabe der Befragten hat die Schule in den vergangenen sechs Jahren nicht an systematischen Schulentwicklungsprojekten, Modellversuchen oder einer weiteren flächendeckenden Maßnahme zur Förderung der Schulentwicklung teilgenommen.

Die Schulentwicklungs- und Evaluationsaktivitäten werden von einer Schulleitung begleitet, die sich neuen Projekten gegenüber als sehr offen charakterisiert und sich als Manager/in dieser Projekte verstanden wissen will (vgl. I-SL_Delta).

Dass im Kollegium Vorbehalte gegenüber Veränderungen bestehen, meint ein geringer Teil der Befragten. Fünf der zehn Lehrkräfte, die an unserer Datenerhebung teilgenommen haben, konstatieren ihrem Kollegium allerdings einen Mangel an Bereitschaft, die eigenen pädagogischen Ansätze zu überprüfen. Schule *Delta* steht den Themen Schulentwicklung und Evaluation aber dennoch positiv und aufgeschlossen gegenüber ($M = 3.81$; $p < .05$).

Die Ergebnisse der Schulinspektion sind an Schule *Delta* im Schulentwicklungsausschuss thematisiert worden. Dabei wurden zwei Schwächen herausgestellt, bei denen die Lehrkräfte einen deutlichen Handlungsbedarf sahen: die Leistungsbewertung und die Curriculumarbeit. Es wurde an der Schule dann aber kein Vorgehen vereinbart, mit dem die erkannten Defizite bearbeitet werden konnten. Es wurden also weder Personen bestimmt, die zuständig sein sollten, noch ein Arbeitsplan mit festen Zeitangaben für die Entwicklung von Maßnahmen oder Konzepten erstellt. Die Themen sind an der Schule von einzelnen Lehrkräften hin und wieder ins Bewusstsein gerufen worden und in dieser Hinsicht – z. B. durch den Besuch einer Fortbildung – in „Wellenbewegungen“ bearbeitet worden; eine strukturierte Vorgehensweise findet sich allerdings nicht. Schulleitung und Schulaufsicht betonen gleichermaßen, dass die schulischen Gegebenheiten und die alltäglichen Erfordernisse die Bedeutung der Schulinspektion und deren Ergebnisse – zumal sie insgesamt so sehr positiv ausgefallen sind – in den Hintergrund rücken lassen. Die Notwendigkeit, die innerschulische Entwicklungsarbeit insgesamt zu systematisieren, wird von der Schulleitung gesehen und die Umsetzung als zukünftige Aufgabe bestimmt. So führt sie aus:

„Wir haben uns jetzt vorgenommen – oder es ist auch ein Anliegen von mir – mal unsere ganze Projektlandschaft zu sortieren jetzt nach diesen sehr stürmischen Jahren“ (I-SL_Delta).

Inwieweit die Schulinspektionsergebnisse in diese Neustrukturierung integriert werden oder auch für die laufende Schulentwicklungsarbeit von Nutzen sein können, wird nicht konkretisiert.

6. Fazit

Der vorliegende Beitrag ging von der Annahme aus, dass die spezifische Situation einer Schule Auswirkungen darauf hat, ob und in welcher Weise die Ergebnisse der Schulinspektion in Schulen rezipiert und verarbeitet werden. Er orientierte sich an theoretischen Rahmenmodellen, die mit *innerschulischen Merkmalen* und *externen Impulsen und Unterstützungsleistungen* zwei Gruppen von Faktoren aufnehmen, die die innerschulischen Aktivitäten infolge der Schulinspektion beeinflussen. Die spezifische Ausgestaltung von Merkmalen dieser beiden Gruppen bedingt dabei die spezifische Situation einer Schule.

Vor dem Hintergrund von Fallbeschreibungen, die im Rahmen des Projekts *Externe Evaluation/Schulinspektion und schulische Qualitätsentwicklung* erstellt und hier in Auszügen präsentiert worden sind, lässt sich zusammenfassend festhalten: Schule *Alpha* zeichnete sich durch langjährige und vielfältige Schulentwicklungsaktivitäten aus, die sehr gut organisiert und in den vergangenen Jahren verstärkt in systematische Kontexte eingebunden waren (Projekte, externe Beratung). Die innerschulische Entwicklungsarbeit folgte einem Organisationskreislauf, der am Schuljahresrhythmus orientiert war und alle am Schulleben beteiligten Gruppen einbezog. Zudem wurde die Organisation der Arbeit generell (u. a. der Informationsfluss innerhalb des Kollegiums) als durchschnittlich gut eingeschätzt. Auch Schule *Delta* zeichnete sich durch vielfältige Schulentwicklungsaktivitäten aus, die jedoch als unterdurchschnittlich gut organisiert eingeschätzt wurden. Sie fanden eher in Einzelprojekten und auf Engagement einzelner Lehrkräfte hin statt, ein umfassendes Konzept existierte nicht. Die Aktivitäten sind nach Aussage der Befragten in den vergangenen Jahren nicht in systematische Kontexte – wie ein Schulentwicklungsprojekt – eingebunden gewesen. Zudem war die Organisation der Arbeit generell (u. a. der Informationsfluss innerhalb des Kollegiums) unterdurchschnittlich gut.

Betrachtet man nun die Prozesse, die in beiden Schulen nach der Rückmeldung der Ergebnisse stattgefunden haben, so zeigt sich, dass an Schule *Alpha* jene Ergebnisse, die auf einen Verbesserungsbedarf hinwiesen, in die bestehende Arbeitsstruktur eingespeist wurden. An Schule *Delta* bestand eine derartige Arbeitsstruktur nicht; die Ergebnisse der Schulinspektion hatten hier nicht zu Verbesserungsaktivitäten geführt; auch ein Jahr nach der Ergebnisrückmeldung wurde vonseiten der schulischen Akteure in den Interviews erwähnt, dass zunächst einmal eine Strukturierung der Vielzahl an Entwicklungsaktivitäten vorgenommen und entschieden werden müsse, in welchen Bereichen bzw. zu welchen Themen die Schule in Zukunft gebündelt Schulentwicklung betreiben wolle. Die Ergebnisse der Schulinspektion könnten hier eine Rolle spielen.

An beiden Schulen finden sich eine hohe innovative Orientierung der Lehrkräfte und ein Schulleitungshandeln, das als partizipativ charakterisiert werden kann.

Diese Faktoren, die prinzipiell förderlich für den Auswertungsprozess sind, werden an Schule *Delta* ganz offensichtlich vom Fehlen einer Organisationsstruktur für die Schulentwicklungsarbeit überlagert bzw. außer Kraft gesetzt. Das Potenzial, das von den rückgemeldeten Daten ausgehen könnte, verpufft hier also gewissermaßen trotz förderlicher schulkultureller Bedingungen.

In den Fallbeschreibungen konnten damit einige Faktoren identifiziert werden, die in den eingangs erwähnten Rahmenmodellen als Einflussfaktoren im innerschulischen Bereich angeführt werden. Die Erkenntnisse der eigenen Studie können nun mit den Befunden anderer Untersuchungen in Zusammenhang gebracht werden. So können sie herangezogen werden, um zu erklären, warum ein nicht unbeträchtlicher Teil von Schulen keine Entwicklungsmaßnahmen auf den Weg bringen, obwohl sie Defizite in bestimmten Bereichen aufweisen (vgl. Wurster et al. 2013).

Bei der Interpretation der Befunde sind nun allerdings einige methodische Aspekte zu berücksichtigen, die auf Grenzen der Untersuchung hinweisen. In Bezug auf die *Anlage der Untersuchung* sind dabei zunächst die einbezogenen Akteure zu nennen: Befragt wurden die Schulleitungen und Lehrkräfte der inspizierten Schulen, nicht aber weitere pädagogische Mitarbeiterinnen und Mitarbeiter, Eltern und Schülerinnen und Schüler. Diese Auswahl von Akteuren bedingt eine Sicht auf die angelaufenen Verarbeitungsprozesse, die – trotz eines großen Überblickswissens – eine zweifelsohne spezifische ist. Sie müsste idealerweise durch die Sichtweisen der oben genannten Akteure ergänzt oder relativiert werden. Die Auswahl der Akteure war bereits bei der Antragstellung aus forschungsökonomischen Gründen erfolgt und von den Gutachtern befürwortet worden, muss nachträglich aber skeptisch betrachtet werden.

Die Auswahl der Schulen für die Fallstudien wurde anhand zweier Kriterien vorgenommen: der *Art des Inspektionsergebnisses* und der *Schulstufe*. Diese erschienen angesichts der damals vorliegenden empirischen Befunde sowie theoretischer Überlegungen plausibel. Inzwischen liegen – wie in Kapitel 3 skizziert – neuere empirische Befunde vor, die belegen, dass die innerschulischen Aktivitäten infolge der Schulinspektion auch unabhängig von den zurückgemeldeten Ergebnissen stattfinden (vgl. Gärtner & Wurster 2013; Gärtner et al. 2014). Und auch in unseren Fallstudien konnten wir keine systematischen Unterschiede zwischen jenen Schulen, die ein insgesamt eher gutes Inspektionsergebnis erreicht haben, und jenen Schulen, die ein eher mittelmäßiges Ergebnis erzielt haben, identifizieren. Nachfolgenden Untersuchungen sollten deshalb andere Kriterien für eine Auswahl von Fallschulen zugrunde gelegt werden.

Hinsichtlich der *Durchführung der Untersuchung* stellen insbesondere die geringen Rücklaufquoten bei den standardisierten Befragungen ein Problem dar. Wegen ihnen konnte die Bedeutung der Organisationsstruktur der Schulentwicklungsarbeit für die Verarbeitung der Ergebnisse der Schulinspektion nur in den Fallstudien, nicht aber in den standardisierten Befragungen über drei Erhebungszeitpunkte hinweg nachgewiesen werden. Die präsentierten Befunde sollten daher als Arbeitshypothese verstanden werden, die es in größer angelegten Untersuchungen weiter zu prüfen gilt. Nichtsdestotrotz weisen sie darauf hin, dass Schulen unterschiedlich gut auf die Erwartungen vorbereitet sind, die im Zuge der evidenzbasierten Steuerung vonsei-

ten der Bildungspolitik und Bildungsadministration an sie herangetragen werden. Soll mit Blick auf die Schulinspektion auch zukünftig am Einsatz eines einheitlichen Verfahrens für alle allgemeinbildenden Schulen der Bundesländer festgehalten werden, so sollte eine differenziertere Betrachtung von Schulen in zweifacher Weise vorgenommen werden: Zum einen sollte in Bezug auf die *Wirkungshoffnungen*, die sich auf die innerschulische Verarbeitung der Schulinspektionsergebnisse beziehen, das spezifische Niveau der Schulentwicklungsarbeit der Dreh- und Angelpunkt sein: Jene Maßstäbe, die an Schulen mit weiter reichenden Erfahrungen und ausgeprägten Kompetenzen sowie einer ausgebildeten Organisations- und Kommunikationsstruktur im Bereich der innerschulischen Entwicklungsarbeit angelegt werden, sollten dabei andere sein als jene, die für weniger schulentwicklungsaffine Schulen gelten. Zum anderen sollte im Hinblick auf die *Unterstützung*, durch die die innerschulische Verarbeitung der Schulinspektionsergebnisse flankiert wird, die jeweilige Bedarfslage der Schule im Fokus der Bemühungen stehen und zur Entwicklung passgenauer Beratungs- und Begleitungsprogramme führen. Hier sind Bildungspolitik und Bildungsadministration noch stärker als bisher gefordert, entsprechende Bedarfe ausfindig zu machen, auf sie ausgerichtete Programme zu konzipieren und die inzwischen in nahezu allen Bundesländern auch von staatlicher Seite angebotenen Programme zu optimieren. Dabei kann davon ausgegangen werden, dass nicht alle Schulen eine Unterstützung benötigen. Für jene Schulen, die einen Bedarf haben, sind verschiedene Varianten der Beratung und Begleitung denkbar: So könnten für einzelne Schulen Programme konzipiert werden, die die individuellen Problemlagen aufgreifen und gewissermaßen maßgeschneidert entwickelt werden. Es könnten aber auch feststehende Qualifizierungsmodule konzipiert werden, an denen mehrere Schulen teilnehmen und bei denen die erworbenen Kenntnisse – unter Begleitung von Moderatoren – dann für die individuelle Problemlage nutzbar gemacht bzw. auf diese angewendet werden. Exemplarisch ließe sich hier das Qualitätsentwicklungsprojekt *Selbstständige Schule* der Landesregierung Nordrhein-Westfalen und der Bertelsmann Stiftung nennen, das in den vergangenen Jahren ebendiese Programmvarianten realisiert hat (vgl. Projektgruppe „Selbstständige Schule“ 2008).

So oder so ist hier aber impliziert, dass Schulen ggf. zunächst einmal bei der Einrichtung von Strukturen für die Schulentwicklungsarbeit, beim Aufbau von Know-how im Bereich des Schulmanagements etc. extern unterstützt werden. Nur so werden langfristig positivere Effekte der Schulinspektion zu erwarten sein – um die Worte Slavins (2005, S. 277) zu bemühen:

„... we must learn how to identify the different needs of schools and provide for these needs, so that the efforts of dedicated reformers are exerted where they will do the greatest good for children.“

Literatur

- Altrichter, H. (2010). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 219–254. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Altrichter, H. & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 15–39. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Böhm-Kasper, O. & Selders, O. (2013). „Schulinspektionen sollten regelmäßig durchgeführt werden“? Ländervergleichende Analyse der Wahrnehmung und Akzeptanz von Schulinspektionsverfahren. In I. van Ackeren, M. Heinrich & F. Thiel (Hrsg.), *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund*. 12. Beiheft der Zeitschrift Die Deutsche Schule. Münster u. a.: Waxmann, S. 121–153.
- Böttcher, W., Keune, M. & Neiwert, P. (2010). *Evaluationsbericht zum Projekt „Schulinspektion in Hessen – Wirkungen auf die Qualitätsentwicklung von Schulen und die Arbeit der Schulaufsicht“*. Westfälische Wilhelms-Universität Münster.
- Böttger-Beer, M. & Koch, E. (2008). Externe Schulevaluation in Sachsen. Ein Dialog zwischen Wissenschaft und Praxis. In W. Böttcher, W. Bos, H. Döbert, H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB)*, S. 253–264. Münster u. a.: Waxmann.
- Bonsen, M., Gathen, J. von der, Iglhaut, C. & Pfeiffer, H. (2002). *Die Wirksamkeit von Schulleitung. Empirische Annäherungen an ein Gesamtmodell schulischen Leitungshandelns*. Weinheim: Juventa.
- Cullingford, C. & Daniels, S. (1999). Effects of OFSTED inspections on school performance. In C. Cullingford (Hrsg.), *An inspector calls. OFSTED and its effects on school standards*, S. 59–69. London: Kogan.
- Dedering, K. (2012). Schulinspektion als wirksamer Weg der Systemsteuerung? *Zeitschrift für Pädagogik*, 58 (1), S. 69–88.
- Dedering, K. & Müller, S. (2011). School Improvement through Inspections? International research findings and first insights from Germany. *Journal of Educational Change*, 12 (3), S. 301–322.
- Dedering, K., Fritsch, N., Nierhoff, C., Richter, K. & Weyer, C. (2014). *Schulinspektion als Instrument der Qualitätsentwicklung von Schulen. Abschlussbericht zum Projekt „Externe Evaluation/Schulinspektion und Qualitätsentwicklung in Schulen“ im Förderschwerpunkt „Steuerung im Bildungssystem (SteBis)“ des Bundesministeriums für Bildung und Forschung (BMBF)*. Universität Vechta.
- Döbert, H., Rürup, M. & Dedering, K. (2008). Externe Evaluation von Schulen in Deutschland – die Konzepte der Bundesländer, ihre Gemeinsamkeiten und Unterschiede. In H. Döbert & K. Dedering (Hrsg.), *Externe Evaluation von Schulen – Historische, rechtliche und vergleichende Aspekte*, S. 63–151. Münster u. a.: Waxmann.
- Ehren, M. C. M. & Visscher, A.J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54 (1), S. 51–72.

- Ehren, M. C. M., Altrichter, H., McNamara, G. & O'Hara, J. (2013). Impact of School inspections on teaching and learning – describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, 25 (1), S. 3–43.
- Ehren, M. C. M., Perryman, J. & Shakleton, N. (2014). Setting expectations for good education: how Dutch school inspections drive improvement. *School Effectiveness and School Improvement*. <http://dx.doi.org/10.1080/09243453.2014.936472>.
- Fatke, R. (2010). Fallstudien in der Erziehungswissenschaft. In B. Friebertshäuser, A. Langer & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*, S. 159–172. Weinheim & München: Juventa.
- Füssel, H.-P. & Leschinsky, A. (2008). Der institutionelle Rahmen des Bildungswesens. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick*, S. 131–203. Reinbek: Rowohlt Taschenbuch Verlag.
- Gärtner, H. & Wurster, S. (2009). *Befragung zur Wirkung von Schulvisitation in Brandenburg. Ergebnisbericht*. Institut für Schulqualität der Länder Berlin und Brandenburg e.V.: Berlin.
- Gärtner, H., Hüsemann, D. & Pant, H. A. (2009). Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. Die Brandenburger Schulleiterbefragung. *Empirische Pädagogik*, 23 (1), S. 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), S. 489–508.
- Grek, S., Lawn, M., Ozga, J. & Segerholm, C. (2013). Governing by inspection? European inspectorates and the creation of a European education policy space. *Comparative Education*, 49 (4), S. 486–502.
- Huber, F. (2008). Konsequenzen aus der externen Evaluation an Bayerns Schulen. In W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB)*, S. 265–278. Münster u. a.: Waxmann.
- Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation: Überblick und Stand der Forschung. *Zeitschrift für Erziehungswissenschaft*, 14 (2), S. 259–283.
- Kotthoff, H.-G. & Böttcher, W. (2010). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 295–325. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lambrecht, M., Kotthoff, H. G. & Maag Merki, K. (2008). Taktieren oder Öffnen? Die Pilotphase Fremdevaluation in Baden-Württemberg zwischen Entwicklung und Kontrolle. In W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive*, S. 279–291. Münster: Waxmann.
- Lamnek, S. (2010). *Qualitative Sozialforschung. Lehrbuch*. 5., überarb. Aufl. Weinheim: Beltz.

- Landwehr, N. (2011). Thesen zur Wirkung und Wirksamkeit der externen Schulevaluation. In C. Quesel, V. Husfeldt, N. Landwehr & P. Steiner (Hrsg.), *Wirkungen und Wirksamkeit der externen Schulevaluation*, S. 35–69. Bern: hep Verlag.
- Louis, K. S., Leithwood, K., Wahlstrom, K. L. & Anderson, S. E. (2010). *Learning from Leadership: Investigating the Links to Improved Student Learning. Final Report of Research to the Wallace Foundation*. University of Minnesota, University of Toronto.
- Luginbuhl, R., Webbink, D. & Wolf, I. de (2009). Do Inspections Improve Primary School Performance? *Educational Evaluation and Policy Analysis* 31 (3), S. 221–237.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 11. Aufl. Weinheim: Beltz.
- Pietsch, M. (2011). *Nutzung und Nützlichkeit der Schulinspektion in Hamburg. Ergebnisse der Hamburger Schulleitungsbefragung*. Behörde für Schule und Berufsbildung, Institut für Bildungsmonitoring, Schulinspektion. Hamburg.
- Pietsch, M., Janke, N. & Mohr, I. (2013). Führt Schulinspektion wirklich nicht zu besseren Schülerleistungen? Eine Einschätzung zur Belastbarkeit vorliegender Wirksamkeitsstudien aus programmtheoretischer Perspektive. In K. Schwippert, M. Bonsen & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung – Diskussionen, Befunde und Perspektiven*, S. 167–185. Münster: Waxmann.
- 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), S. 446–470.
- Projektleitung „Selbstständige Schule“ (2008). *Selbstständige Schulen in regionalen Bildungslandschaften. Eine Bilanz*. Troisdorf: Bildungsvlag EINS.
- Rosenthal, L. (2004). Do school inspections improve school quality? OFSTED inspections and school examination results in the UK. *Economics of Education Review* 23 (2), S. 143–151.
- 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), S. 63–75.
- Slavin, R. E. (2005). Sand, Bricks, and Seeds: School Change Strategies and Readiness for Reform. In D. Hopkins (Hrsg.), *The Practice and Theory of School Improvement. International Handbook of Educational Change*, S. 265–279. Dordrecht: Springer.
- Sommer, N. (2011). Wie beurteilen schulische Gruppen die erlebte Schulinspektion? Ergebnisse einer Befragung. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulin-spektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht*, S. 137–164. Münster u. a.: Waxmann.
- Wilcox, B. & Gray, J. (1996). *Inspecting schools: holding schools to account and helping schools to improve*. Buckingham: Open University Press.
- Wissinger, J. (2011). Schulleitung und Schulleitungshandeln. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf*, S. 98–115. Münster u. a.: Waxmann.
- Wolf, I. de & Janssens, F. J. G. (2008). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education* 33 (3), S. 379–396.

- Wurster, S. & Gärtner, H. (2013). Schulen im Umgang mit Schulinspektion und deren Ergebnissen. *Zeitschrift für Pädagogik* 59 (3), S. 425–444.
- Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2013). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. In I. van Ackeren, M. Heinrich & F. Thiel (Hrsg.), *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund*. 12. Beiheft der Zeitschrift Die Deutsche Schule. Münster u. a.: Waxmann, S. 19–49.

*Oliver Böhm-Kasper, Thomas Brüsemeister,
Fabian Dietrich, Lisa Gromala, Martin Heinrich,
Maike Lambrecht, Bianca E. Preuß, Matthias Rürup,
Odette Selders, Jochen Wissinger*

Schulinspektion als Steuerungsimpuls zur Schul- entwicklung und seine Realisierungsbedingungen auf einzelschulischer Ebene

*Ergebnisse eines triangulativ orientierten
Verbundprojekts*

1. Schulinspektionen im Kontext einer evidenzbasierten Schulsystemsteuerung

Schulinspektionen können zunächst einmal anhand ihrer äußeren Erscheinungsform beschrieben werden: Sie stellen eine Form der externen Evaluation von Schule dar, d. h., sie sind ein „Verfahren der systematischen evaluativen Bestandsaufnahme der Arbeitsbedingungen, Arbeitsweisen und Arbeitsergebnisse der einzelnen Schule“ (Rürup 2008, S. 467). Vor allem im Rahmen eines Schulbesuchs soll bei Schulinspektionen also systematisch objektiviertes Wissen generiert werden (vgl. van Ackeren et al. 2011, S. 172). Vor diesem Hintergrund können Schulinspektionen weiter als Bestandteil einer evidenzbasierten Schulsystemsteuerung beschrieben werden, in der sich die Hoffnung auf eine gezielte Gestaltbarkeit des Schulsystems reproduziert und erneuert: „Nicht mehr mittels einer zentralistischen Planung oder einer hierarchischen Steuerung soll das schulische Bildungssystem reformiert bzw. ‚optimiert‘ werden. (...) Stattdessen wird auf eine dezentrale Steuerungswirkung empirischer Evidenzen gesetzt“ (Dietrich & Lambrecht 2012, S. 58). Schulinspektionen gehören damit zu den bildungspolitischen Instrumenten, die nach PISA eingeführt wurden, um belastbares Steuerungswissen sowohl für schulische als auch bildungsadministrative und bildungspolitische Akteure zu generieren und so mittel- bis langfristig die Schulqualität zu erhöhen und Schülerleistungen zu steigern. Dem Instrument wird sowohl eine Rechenschaftslegungs- als auch eine Schulentwicklungsfunktion zugeschrieben (z. B. Böttcher & Kotthoff 2007), wobei Letztere in der deutschsprachigen Diskussion dominiert (vgl. Husfeldt 2011). Daraus ergibt sich ein komplexeres Verständnis von Schulinspektionen, wonach das Instrument als Ausdruck des Versuchs einer Rationalisierung der Schulsystemsteuerung beschrieben werden kann, die sich insbesondere hierzulande in der programmatischen Steuerungsvorstellung einer „Schulentwicklung durch Einsicht“ (vgl. Böttger-Beer & Koch 200; Rürup 2008) ma-

nifestiert: Die Schulinspektion soll nicht durch Sanktionen wirken, sondern durch den „zwanglosen Zwang“ der systematisch generierten Evidenz (vgl. Dietrich & Lambrecht 2012, S. 68). Dies kommt auch prägnant in der offiziellen Beschreibung der externen Schulevaluation in Mecklenburg-Vorpommern zum Ausdruck; dort heißt es: „Evaluation setzt Fakten an die Stelle von Vermutungen und Mutmaßungen“ (LISA 2006, S. 6).

So weit die programmatische Verortung des Steuerungsinstrumentes Schulinspektion. Vor diesem Hintergrund hat das Verbundprojekt „Schulinspektion als Steuerungsimpuls zur Schulentwicklung und seine Realisierungsbedingungen auf einzelschulischer Ebene“¹, dessen Forschungsansatz und zentrale Ergebnisse hier im Folgenden vorgestellt werden, versucht, die (Aus-)Wirkungen der Schulinspektion in vier Bundesländern in Zusammenhang mit der schulinspektionsbezogenen Handlungskoordination im Mehrebenensystem Schule empirisch zu bestimmen.

2. Ansatz des Verbundprojekts

Im Folgenden wird zunächst die Forschungsperspektive Educational Governance vorgestellt, die die Forschungsaktivitäten des Gesamtverbundes angeleitet hat (Kapitel 2.1); es folgen das Forschungsdesign (Kapitel 2.2) sowie der triangulative Ansatz des Verbundprojekts (Kapitel 2.3).

2.1 Forschungsperspektive Educational Governance

Die internationale und deutschsprachige Schulinspektionsforschung hat sich lange auf zwei Aspekte konzentriert: die Akzeptanz des Instruments bei unterschiedlichen Akteuren des Schulsystems und die Wirksamkeit des Instruments auf die Schul- und Unterrichtsentwicklung sowie auf die Schülerleistungen (vgl. Böttcher & Keune 2010; Husfeldt 2011; Lambrecht & Rürup 2012). Dabei sind die Forschungsbefunde zur Wirksamkeit der Schulinspektion nicht nur in Deutschland, sondern auch in Ländern mit einer längeren Schulinspektionstradition bisher uneindeutig. Dies trifft insbesondere auf die Frage zu, ob die ungewollten Nebenwirkungen die gewünschten Wirkungen überlagern. Ausgangspunkt für das Verbundprojekt war daher die Überlegung, dass die zentralen Wirkmechanismen des Instruments Schulinspektion genauer aufgeschlüsselt werden müssen, als dies bis dahin der Fall war. Im Rahmen des Verbundprojekts wurden daher Ansätze der Schulentwicklungs- und Schuleffektivitätsforschung (z. B. Bensen et al. 2008) mit der Perspektive der Educational Governance (vgl. Altrichter et al. 2007) verbunden.

1 Das Verbundprojekt wurde von 2010 bis 2013 vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der Förderrichtlinie „Steuerung im Bildungssystem“ (SteBis) gefördert und unter der Leitung von Thomas Brüsemeister (Justus-Liebig-Universität Gießen, Hauptantragsteller), Matthias Rürup (HSU Hamburg/Bergische Universität Wuppertal, Verbundkoordination), Oliver Böhm-Kasper (Universität Bielefeld), Martin Heinrich (Leibniz Universität Hannover) und Jochen Wissinger (Justus-Liebig-Universität Gießen) durchgeführt.

Die Governance-Forschung fokussiert auf die „Handlungskoordination zwischen Akteurkonstellationen in einem Mehrebenensystem“ (ebd., S. 10) und ermöglicht so eine komplexere Beschreibung von Steuerungs-, Umstrukturierungs- und Reformprozessen im Bildungssystem als klassische steuerungstheoretische Konzepte. Steuerungsinstrumente und deren Wirkungen werden hier nicht mehr isoliert, sondern eingebettet in die komplexe Handlungskoordination im Mehrebenensystem Schule betrachtet. Ziel von Governance-Analysen ist die Rekonstruktion von sogenannten bereichsspezifischen Governance-Regimen, also von „empirischen Ausprägungen der Handlungskoordination von Akteuren“ (Kussau & Brüsemeister 2007, S. 41), über die deutlich wird, „wie die Handlungskoordination in einem spezifischen gesellschaftlichen System durch spezifische Relationierungen (Handlungen) von differenzierten Akteuren, Strukturen, Intentionen und Wirkungen aufgebaut, aufrechterhalten und transformiert wird“ (Altrichter & Heinrich 2007, S. 73).

Vor diesem Hintergrund wurde im Rahmen des Verbundprojekts untersucht, wie die Schulinspektion in vier Bundesländern konzeptionalisiert, implementiert, weiterentwickelt und im Durchgang durch die verschiedenen Ebenen des Schulsystems verändert wird und welche intendierten und nicht intendierten Wirkungen sich im Rahmen einer Akteurkonstellation von Kultusadministration, regionalen schulbezogenen Interessenträgern und Einzelschulen einstellen. Zur Bearbeitung dieses Gegenstands wurden fünf Fragestellungen formuliert, die im Forschungsverbund arbeitsteilig bearbeitet und beantwortet wurden:

1. Welche Interdependenzbeziehungen lassen sich zwischen den Schulinspektionen und den spezifischen Qualitätssicherungs- und Qualitätsentwicklungsanstrengungen auf der Ebene der Einzelschule ausmachen?
2. Sind vorfindbare administrative Variationen der Konzepte der Schulinspektion mit entsprechend unterschiedlichen Rezeptionen, Akzeptanzen und Verarbeitungen von Schulinspektionsimpulsen auf Ebene der Einzelschule verbunden?
3. Wie verändern sich Zielvorstellungen, erwartete Wirkungen und Wirkungsmechanismen akteurspezifisch bei der Rezeption des Instruments Schulinspektion im Mehrebenensystem Schule?
4. Wie unterscheiden sich die spezifischen Rezeptionen und Wahrnehmungen der Schulinspektion im regionalen schulischen Umfeld?
5. Inwieweit sind Wahrnehmungen und Verarbeitungen von Schulinspektionsimpulsen abhängig von der Eigensinnigkeit der Akteure in ihrem konkreten Umfeld?

Der Ansatz des Verbundprojekts stellt demnach die Frage nach der Wirksamkeit von Schulinspektionen („Macht Schulinspektion Schulen besser?“) in den Kontext der komplexen Handlungskoordination im Mehrebenensystem Schule und nimmt damit eine Reihe von forschungslogischen Verschiebungen vor: von der Frage der Wirksamkeit hin zur Frage der Wirkungsmechanismen, von der wissenschaftlichen Reproduktion der politisch-administrativen Perspektive und deren normativer Steuerungsvorstellung hin zur interdependenten Handlungskoordination im Mehrebenensystem Schule und von der Programmatik einer evidenzbasierten Schulsystemsteuerung zur empirischen Schärfung eines sogenannten evidenzbasierten Governance-Regimes.

2.2 Forschungsdesign des Verbundprojekts

In die Analysen des Verbundprojekts wurden Inspektionssysteme aus vier Bundesländern und die jeweilige schulinspektionsbezogene Akteurkonstellation in diesen Bundesländern einbezogen. Die Entscheidung für einen Ländervergleich beruhte auf der Annahme, dass sich die untersuchten Inspektionssysteme in wesentlichen Punkten voneinander unterscheiden und sich abhängig vom Inspektionstyp unterschiedliche Wirkungen und Formen der Handlungskoordination einstellen. Leitend bei der Auswahl der Bundesländer war die Klassifikation in wahrheits- und angemessenheitsorientierte Verfahren (vgl. Rürup 2008). Vor diesem Hintergrund wurden die folgenden Bundesländer in die Untersuchungen einbezogen²:

- die standardisiert-rechenschaftsorientierte Schulinspektion in Niedersachsen,
- die ebenfalls standardisierte, aber weniger rechenschaftsorientierte externe Evaluation in Mecklenburg-Vorpommern,
- die externe Evaluation in Thüringen, die als Teil des Entwicklungsvorhabens Eigenverantwortliche Schule (EVAS) konzipiert wurde, sowie
- die qualitativ-schulentwicklungsorientierte Fremdevaluation in Baden-Württemberg.

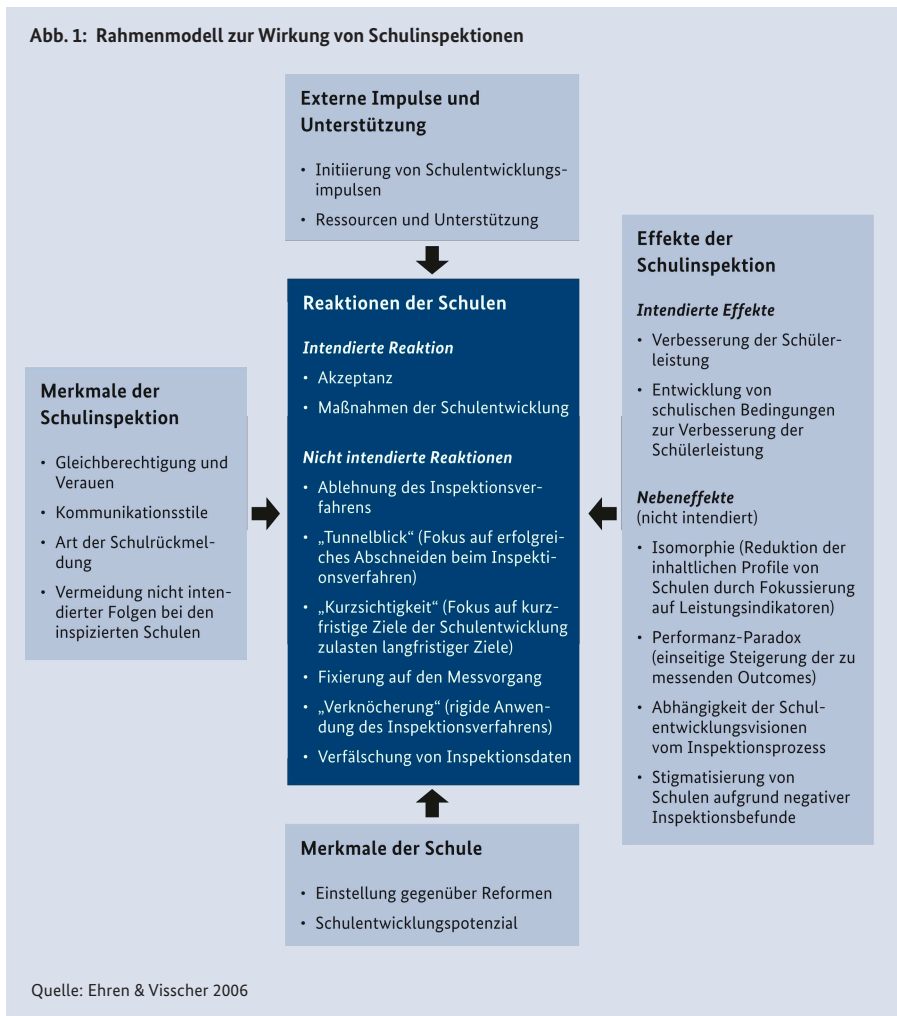
Das Untersuchungsdesign umfasste eine quantitative und eine qualitative Teilstudie. **Die quantitative Teilstudie** bestand aus einer quasi-längsschnittlichen Online-Befragung, die sich an alle Schulleitungen der allgemeinbildenden Schulen der vier am Projekt beteiligten Bundesländer richtete.³ Auf Basis des Ehren-Visscher-Modells (2006, s. Abbildung 1) wurden zu zwei Zeitpunkten Stand und Dynamik der Schulentwicklung in Abhängigkeit von Impulsen der Schulinspektion erhoben.⁴

Das Modell zielt auf eine angemessene Beschreibung von sowohl erwünschten als auch unerwünschten Auswirkungen externer Schulevaluation (vgl. Husfeldt 2011, S. 268). Hierzu werden fünf Dimensionen unterschieden: (1) Merkmale der Schulinspektion, (2) schulentwicklungsrelevante Merkmale der Schule, (3) externe

2 Die hier vorgenommene Kurzcharakterisierung der verschiedenen Schulinspektionskonzepte bezieht sich auf den Stand der Entwicklung der Konzepte zum Zeitpunkt der Erhebungen im Rahmen des Verbundprojekts, die schwerpunktmäßig 2010 und 2011 stattfanden.

3 Die bewusste Eingrenzung der Erhebung auf Schulleitungen bzw. auf in das Inspektionsverfahren involvierte Personen (z. B. Konrektoren oder Steuergruppenleiter) hat mehrere Gründe: Schulleiter (bzw. deren Vertreter) sind auf jeden Fall Ansprechpartner bei Inspektionsverfahren. Inwieweit andere Mitglieder des Lehrerkollegiums involviert sind, ist vorab nicht bekannt und wäre nur mit unverhältnismäßigem Aufwand flächendeckend zu recherchieren gewesen. Eine Befragung von Lehrerkollegien ist logistisch aufwendiger und sollte eher postalisch durchgeführt werden, um einen akzeptablen Rücklauf zu erhalten. Diese Vorgehensweise ist für einen Ländervergleich unverhältnismäßig aufwendig und mit zu hohen Kosten verbunden, da bei einem solchen Vorgehen die Mehrebenenstruktur der zu erhebenden Daten durch entsprechend hohe Stichprobenumfänge berücksichtigt werden müsste. Zudem ist die emotionale Verarbeitung und professionelle Bearbeitung durch die Schulleitung ein wichtiger Einflussfaktor auf die Rezeption einer anstehenden Schulinspektion (vgl. Brimblecombe et al. 1996).

4 An der ersten Erhebung nahmen insgesamt 1.470 Schulleiterinnen und Schulleiter (52 Prozent weiblich, 48 Prozent männlich) teil, die den Online-Fragebogen komplett ausgefüllt haben. Im Rahmen der zweiten Erhebung konnten 1.193 Schulleiterinnen und Schulleiter dazu motiviert werden, an der Befragung teilzunehmen.



Impulse und Unterstützung, (4) Reaktionen der Schule auf die Schulinspektion sowie (5) Effekte und Nebeneffekte der Schulinspektion. Die Merkmale der Schulinspektion werden als unabhängige Variablen betrachtet, die unterschiedliche Reaktionen bei den Schulen hervorrufen können. Diese sind wiederum durch externe Impulse und Unterstützungen beeinflussbar und vom generellen Innovationswillen der schulischen Akteure abhängig. Die schulischen Reaktionen auf die Schulinspektion führen letztendlich zu erwünschten und unerwünschten Wirkungen auf unterschiedlichen Handlungsebenen.

Die Bezeichnung „Rahmenmodell“ macht deutlich, dass hier zunächst Zusammenhänge zwischen Merkmalen der Inspektionsverfahren und intendierten sowie nicht intendierten Folgen aufseiten der Schulen in den Blick genommen werden. Eine konkrete theoretische Ausformulierung der schulinternen Wirkmechanismen enthält das Modell jedoch nicht. Zudem waren die theoretisch postulierten Zusammenhänge des Rahmenmodells insgesamt bislang lediglich im Rahmen von Fallstu-

dien (vgl. Ehren & Visscher 2008), aber nicht flächendeckend quantitativ überprüft worden.

Obwohl in diesem Rahmenmodell ein wesentlicher intendierter Effekt einer Schulinspektion die Steigerung der Schülerleistung ist, wurde die Erhebung von Schülerleistungen im Rahmen der quantitativen Teilstudie des Verbundprojekts bewusst ausgeklammert, und zwar aus zwei Gründen: Zum einen liegt bereits eine Reihe empirischer Studien zu schülerleistungsbezogenen Effekten der Schulinspektion vor (z. B. Rosenthal 2004; Luginbuhl et al. 2009; McCrone et al. 2007); zum anderen steht dieser Aspekt nicht genuin im Vordergrund der Schulinspektionskonzepte in Deutschland, sondern die Wirkungen der Schulinspektionen auf die Schulentwicklung. Innerschulische Vorgänge gelingender oder ggf. auch misslingender Schulentwicklung standen daher auch im Zentrum der Untersuchungen des Gesamtverbundes.

Die Grenzen der quantitativen Teilstudie sind in zwei Bereichen zu verorten: Zum einen kann eine methodisch angemessene Überprüfung der Kausalwirkungen im Ehren-Visscher-Modell (2006) nur über experimentelle oder zumindest quasi-experimentelle Designs gewährleistet werden. Dies ist aufgrund des Forschungsgegenstandes jedoch mittlerweile ausgeschlossen. Bereits 2011 konstatiert Sommer: „(...) der Nachweis kausaler Zusammenhänge ist schwierig, das Zeitfenster für einen quasi-experimentellen Vergleich der Entwicklungen in den Ländern mit und ohne Schulinspektion hat sich in Deutschland nahezu geschlossen“ (Sommer 2011, S. 138). Zum Zweiten entspricht der im Rahmen der Online-Erhebung erfolgte Rücklauf nicht den ursprünglichen Erwartungen. Es ist daher mit Selbstselektionseffekten zu rechnen, die die Generalisierbarkeit der gewonnenen Ergebnisse infrage stellen. (Das Problem der Selbstselektion wäre auch bei der Ziehung von Zufallsstichproben aufgetreten, da sich gerade zu Beginn der Etablierung von Schulinspektionsverfahren die Schulen freiwillig für die Durchführung einer Inspektion zur Verfügung gestellt haben.) Leider gibt es noch zu wenig Wissen darüber, ob und welche systematischen Motivlagen bei Schulleitungen dazu führen, sich reflexiv mit Schulinspektionsverfahren auseinanderzusetzen. Vor diesem Hintergrund erscheint es auch nicht weiterführend, durch Matchingverfahren (z. B. Propensity Score Matching) statistisch äquivalente Vergleichsgruppen zu bilden, wenn theoretisch oder empirisch fundiertes Wissen über entsprechende Kovariaten fehlt.

Die **qualitative Teilstudie** bestand aus narrativ eingeleiteten Experteninterviews (vgl. Gläser & Laudel 2009) mit zentralen, regionalen und schulischen Akteuren (N = 81). Sie thematisierten die Konzeption, Wahrnehmung und Entwicklung der Schulinspektion sowie darauf bezogene Erfahrungen unterschiedlicher schulinspektionsbezogener Akteursgruppen: (1) Vertreter und Vertreterinnen von Ministerien, (2) Vertreter und Vertreterinnen der Schulinspektionssysteme, (3) Schulaufsichten, (4) Schulträger, (5) Schulleitungen, (6) Lehrerinnen und Lehrer, (7) Eltern.

Aus dem Analyseansatz der Educational-Governance-Perspektive ist grundsätzlich die Aussage abzuleiten, dass sich die Komplexität von Steuerungsprozessen im Bildungswesen daraus ergeben, dass vielfältige Akteure in verschiedenen Handlungskontexten mit je unterschiedlichen Sichtweisen und Interessen aufeinandertreffen. Die Möglichkeiten der Schulpolitik, ihre Sichtweisen und Interessen

einseitig und vollständig durchzusetzen, sind so systematisch begrenzt. Die Anlage der qualitativen Interviewstudie des Verbundprojekts ermöglicht dementsprechend, Schulinspektionen als Ausdruck einer interdependenten Handlungskoordination zu analysieren.

Das qualitative Datenmaterial wurde im Rahmen der am Verbundprojekt beteiligten Teilprojekte mit unterschiedlichen thematischen Schwerpunkten und unterschiedlichen methodischen Zugriffen analysiert (s. Kapitel 2.3):

- Analyse von subjektiven Wirkungsvorstellungen einzelner Akteursgruppen in Abhängigkeit der Inspektionsvarianten mithilfe der qualitativen Inhaltsanalyse (vgl. Mayring 2003),
- thematische Kodierung (vgl. Flick 2007) von akteurspezifischen Positionierungen und Verhältnisbestimmungen innerhalb der schulinspektionsbezogenen Akteurkonstellation sowie Auswirkungen der Schulinspektion in der schulisch-regionalen Akteurkonstellation,
- objektiv-hermeneutische Rekonstruktionen latenter Strukturlogiken der schulinspektionsbezogenen Handlungskoordination (vgl. Oevermann 2000; Dietrich 2014).

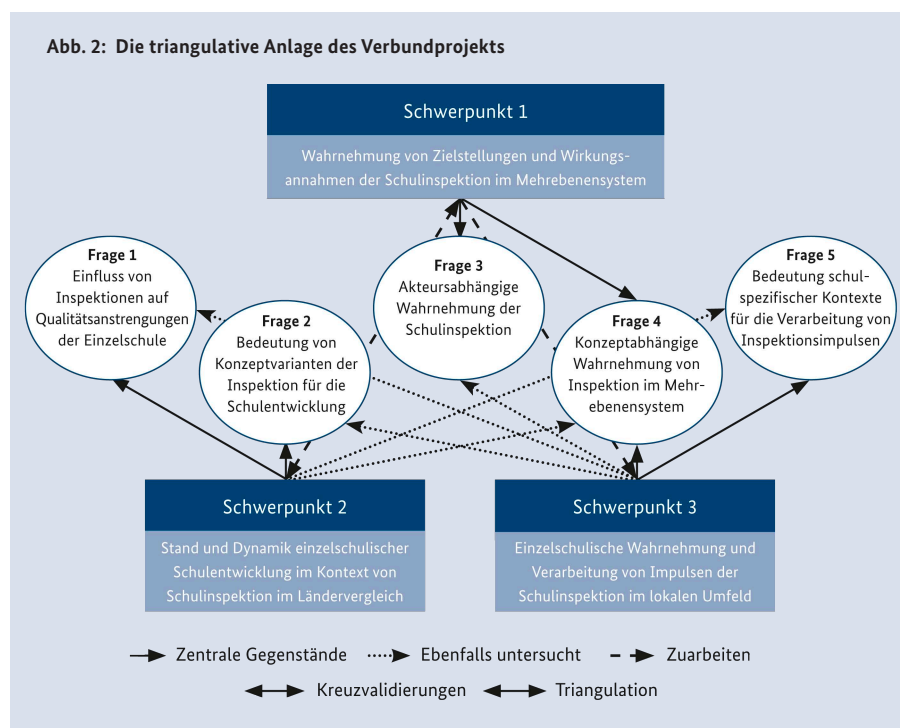
Die Untersuchungen wurden durch **Dokumentenanalysen** der offiziellen Darstellungen der Schulinspektionskonzeptionen ergänzt, die der Charakterisierung und dem Nachvollzug der Entwicklung der untersuchten Schulinspektionskonzepte dienen.

2.3 Der triangulative Ansatz des Verbundprojekts

Die Darstellung des Forschungsdesigns legt bereits nahe, dass das Verbundprojekt einen triangulativen Ansatz (vgl. Flick 2011) verfolgt. Dahinter steht die Annahme, dass Triangulation dem Gegenstand inhärent ist: Die Perspektive der Educational Governance geht davon aus, dass die Analyse komplexer Governance-Regime u. a. die Integration unterschiedlicher sozialwissenschaftlicher Theorieansätze, den Einbezug unterschiedlicher Datensorten und Manifestationsformen von Politik, die Berücksichtigung sowohl von Akteuren und ihrer Handlungen als auch von Strukturen sowie einen Abgleich zwischen normativen Konzepten und empirisch vorfindbaren Konstellationen bedarf (Altrichter & Maag Merki 2010, S. 27–34). Das Verbundprojekt reagiert darauf in erster Linie mit der Kombination unterschiedlicher methodischer Zugriffe auf das Phänomen Schulinspektion (quantitativ-standardisiert, qualitativ-kodierend, rekonstruktiv), die unterschiedliche Wirklichkeitsebenen des untersuchten Gegenstandes abbilden (z. B. manifest/latent) und unterschiedliche datenbezogenen Aggregationsniveaus ergeben.

Zur Beantwortung der in Kapitel 2.1 dargestellten Fragestellungen wurden drei inhaltliche Schwerpunkte gebildet, in denen die einzelnen Teilprojekte in unterschiedlichen Konstellationen arbeitsteilig zusammengearbeitet haben (s. Abbildung 2). Im Schwerpunkt 1 stand die akteurspezifische Wahrnehmung von Zielstellungen und Wirkungsannahmen der unterschiedlichen Schulinspektionskonzepte im Mehrebenensystem Schule im Mittelpunkt (Fragestellungen 3 und 4); hierfür spielten

insbesondere die Ergebnisse der quantitativen Schulleitungsbefragung und die inhaltsanalytischen Auswertungen der qualitativen Interviews eine Rolle (s. Kapitel 3.2). Gegenstand des Schwerpunkts 2 war die Analyse von Stand und Dynamik der einzelschulischen Entwicklung im Kontext von Schulinspektion im Ländervergleich (Fragestellungen 1 und 2); hier hinein fällt die Frage, inwiefern die Schulinspektion die innerschulische Entwicklung im Sinne der intendierten Wirkungen tatsächlich beeinflusst (s. Kapitel 3.1). Die Beantwortung der Frage stützt sich in erster Linie auf Ergebnisse der quantitativen Schulleitungsbefragung, die jedoch durch Ergebnisse der qualitativ-kodierenden Analysen unterstützt und durch die objektiv-hermeneutischen Rekonstruktionen zur latenten Strukturlogik der schulinspektionsbezogenen Handlungskoordination kontextualisiert werden konnten (s. Kapitel 3.3 und 4).



Der Schwerpunkt 3 schließlich umfasste ursprünglich die vertiefte Analyse der einzelschulischen Wahrnehmung und Verarbeitung von Impulsen der Schulinspektion in ihrem lokalen Umfeld (Fragestellungen 4 und 5). Aus inhaltlichen (z. B. Nichtbestätigung von Vorannahmen zur Relevanz der unterschiedlichen Schulinspektionskonzeptionen) und forschungspragmatischen (Anonymisierungsproblematik) Gründen konzentrierte sich dieser Schwerpunkt aber zunehmend auf die eigentliche Analyse der Handlungskoordination innerhalb der schulinspektionsbezogenen Akteurkonstellation, und zwar einerseits anhand der subjektiven Sichtweisen der Akteure auf die Akteurkonstellation und andererseits über die Rekonstruktion der latenten Sinnstrukturen dieser Handlungskoordination (s. Kapitel 3.3).

Im Zentrum der folgenden Ergebnisdarstellungen steht nun die Darstellung der Ergebnisse der einzelnen Teilprojekte in Bezug auf die genannten Schwerpunkte und deren triangulative Verknüpfung.

3. Ergebnisse

Um einen systematischen und konzentrierten Überblick über die Ergebnisse des Verbundprojekts geben zu können, fokussiert die Ergebnisdarstellung auf folgende Aspekte, die sich an den im vorigen Kapitel vorgestellten thematischen Schwerpunkten des Verbundprojekts orientieren: Schulinspektion als Steuerungsimpuls für die Schulentwicklung (Kapitel 3.1), konzeptionsspezifische und akteurspezifische Wahrnehmung und Verarbeitung der Schulinspektion (Kapitel 3.2), schulinspektionsbezogene Handlungskoordination (Kapitel 3.3).

Die Darstellung folgt dabei dem Anspruch des Verbundprojekts, Ansätze der Schulentwicklungs- und Schuleffektivitätsforschung mit der Perspektive der Educational Governance zu verbinden. Analog zum triangulativen Ansatz des Verbundprojekts erfolgt die Ergebnisdarstellung dabei teilprojektübergreifend. Das Kapitel endet mit einer Zusammenfassung (Kapitel 3.4), die die Analyseergebnisse im Hinblick auf die Frage nach der Spezifik der schulinspektionsbezogenen Handlungskoordination im Mehrebenensystem Schule bündelt. Sie kann als erster Schritt der empirisch fundierten Formulierung eines evidenzbasierten Governance-Regimes gelesen werden.⁵

3.1 Schulinspektion als Steuerungsimpuls für die Schulentwicklung

Wie bereits in Kapitel 2.2 beschrieben, wurde im Rahmen der quantitativen Teilstudie das Ehren-Visscher-Modell (2006) zur idealtypischen Wirkung von Schulinspektionen als Grundlage für die Operationalisierung einer standardisierten Befragung von Schulleitungen an allgemeinbildenden Schulen in vier deutschen Bundesländern (Baden-Württemberg, Niedersachsen, Mecklenburg-Vorpommern, Thüringen) verwendet. Zur Beantwortung der Frage nach der Wirksamkeit der Schulinspektion wurden vor allem die intendierten Effekte im Bereich der Schul- und Unterrichtsentwicklung in den Blick genommen (ausführlich in Böhm-Kasper & Selders 2013).

Alle untersuchten Schulinspektionsverfahren zielen in ihrem Selbstverständnis letztendlich auf eine Anregung der Verbesserung schulischer Arbeit. Allerdings ergibt sich aus der quantitativen Teilstudie, dass in allen vier untersuchten Bundesländern die Schulleitungen kaum Auswirkungen der Schulinspektion auf Schul- und Unter-

⁵ Die inhaltliche Arbeit des Verbundprojekts wird im Rahmen des ebenfalls vom BMBF geförderten Folgeprojekts „Funktionen von Schulinspektion“ (2013–2016) unter der Leitung von Martin Heinrich (Universität Bielefeld), Thomas Brüsemeister (Justus-Liebig-Universität Gießen) und Jochen Wissinger (Justus-Liebig-Universität Gießen) fortgesetzt.

richtsentwicklungsmaßnahmen wahrnehmen. In Tabelle 1 sind die entsprechenden Mittelwerte der eingesetzten Skalen im Ländervergleich aufgeführt.⁶

Die Befunde zeigen dementsprechend ein eher ernüchterndes Bild der mit Schulinspektionen verbundenen Effekte. Die Auswirkungen der Inspektion auf Schul- und Unterrichtsentwicklung sind aus Sicht der Schulleitungen als marginal zu beschreiben. Signifikante Länderdifferenzen sind bei der Umsetzung systematischer Schulentwicklung nach Inspektionen nicht beobachtbar. Selbst im Bundesland Niedersachsen, in dem die Schulinspektion als Konsequenz für nicht erreichte Qualitätsstandards eine Nachinspektion anordnen kann, zeigen die niedersächsischen Schulleitungen keine von diesem Befund bedeutsam abweichenden Werte.⁷

Tabelle 1: Effekte der Schulinspektion aus Schulleitungssicht im Ländervergleich

	BW (N = 70–72)	MV (N = 13–14)	NI (N = 478–482)	TH (N = 25–26)	Ges. (N = 586–594)	t-Wert**	Signi- fikanz	d
Skala (Beispielitem)	Mittelwert* Standardabweichung							
Effekte auf Schul- und Unterrichts- entwicklung I (Effekte der Schul- inspektion: Umset- zung systematischer Schulentwicklung)	1.94 .68	1.89 .47	1.99 .59	2.35 .58	2.00 .60	.568	n.s.	–
Effekte auf Schul- und Unterrichtsent- wicklung II (Effekte der Schul- inspektion: Verbesse- rung der Unterrichts- kompetenz der Lehrkräfte)	2.18 .65	2.15 .66	2.23 .60	2.45 .55	2.23 .60	.679	n.s.	–

BW = Baden-Württemberg, MV = Mecklenburg-Vorpommern, NI = Niedersachsen, TH = Thüringen

* Antwortskala: 1 „trifft überhaupt nicht zu“ bis 4 „trifft voll und ganz zu“

** Die t-Tests wurden zwischen den Ländern Niedersachsen und Baden-Württemberg berechnet.⁸

- 6 Da die Effekte nicht unmittelbar als Reaktion auf die Durchführung einer Schulinspektion erwartet werden können, werden hier die Daten des zweiten Messzeitpunktes verwendet.
- 7 Etwas differenzierter erscheint das Bild, wenn nur die niedersächsischen Schulen betrachtet werden, die einer Nachinspektion unterzogen wurden. Im Report der Niedersächsischen Schulinspektion (2008) wird von 14 vollzogenen Nachinspektionen im Berichtsjahr 2008 berichtet. Es konnte festgestellt werden, dass sich 12 der 14 nachinspizierten Schulen so weit verbessert hatten, dass sie zumindest die Norm für eine ausreichende Schulqualität überschreiten konnten.
- 8 Die Beschränkung des Mittelwertvergleiches auf diese beiden Länder erfolgt aufgrund methodischer und inhaltlicher Überlegungen: Zum einen ist die Stichprobengröße der Bundesländer Thüringen und Mecklenburg-Vorpommern zu gering, um belastbare Signifikanztestungen und Berechnungen von Effektstärken (vgl. Bortz & Döring 1995, S. 574 f.) vornehmen zu können. Zum anderen stellen die Verfahren in Niedersachsen und Baden-Württemberg einen maximalen Kontrast zwischen den in unserer Ländergruppe vorfindbaren Konzeptionen der Schulinspektion dar. Das wissenschaftsorientierte und mit Sanktionen (Nachinspektion) verbundene Verfahren ist in Niedersachsen etabliert, während das auf einem Peer-Review basierende Verfahren in Baden-Württemberg beheimatet ist.

Dieses grundlegende Ergebnis der quantitativen Teilstudie wird auch von den Ergebnissen der qualitativen Teilstudien tendenziell gestützt. Die thematisch-kodierenden Analysen der Experteninterviews deuten zwar einzelfallabhängig darauf hin, dass einige Schulen Empfehlungen der Schulinspektion aufgreifen, zum Beispiel indem Eltern und Lehrkräfte sowie externe Fachberaterinnen und Fachberater und weitere Expertinnen und Experten in Schulentwicklungsstrategien und -konzepte integriert werden. Andere wiederum schildern eher gewohnheits- und routinemäßige Kooperations- und Koordinationsmuster, die von der Schulinspektion eher unbeeindruckt scheinen.

Damit reproduziert sich sowohl in der quantitativen als auch in der qualitativen Teilstudie des Verbundprojekts die auch international bereits festgestellte Uneindeutigkeit der Wirksamkeit von Schulinspektionen in Bezug auf die Schul- und Unterrichtsentwicklung. Die überwiegende Zahl der Studien zeigt dabei lediglich geringe bzw. eben uneindeutige Auswirkungen der Schulinspektion auf die schulische Qualität. Vor allem auf die Unterrichtsentwicklung scheint die Schulinspektion wenig Einfluss zu haben (z. B. Chapman 2001; Brimblecombe et al. 1996). Böttcher und Keune resümieren hierzu: „Inspektion ist ein aufwendiges Verfahren, das den ‚Kernbereich‘ der Schule weitgehend ignoriert“ (2010, S. 158).

Ein Grund für den mangelnden Effekt der Schulinspektion auf Schul- und Unterrichtsprozesse besteht möglicherweise in der mangelnden Unterstützung der Schulen und insbesondere der Schulleitungen durch die Schulbehörden im Nachgang von Schulinspektionen. Diese werden von den im Rahmen der quantitativen Teilstudie befragten Schulleitungen kaum wahrgenommen. Besonders wenig Zustimmung erhält die Frage nach der beratenden Tätigkeit der regionalen Schulaufsicht bzw. des Schulamtes nach Bekanntgabe der Ergebnisse aus zentralen Leistungsuntersuchungen. Aus der vorliegenden empirischen Literatur ist jedoch bekannt, dass nur ein kleiner Teil der Schulen in der Lage ist, aus sich heraus und ohne externe Impulse und Unterstützung Innovationen zu initiieren. Gern werden Schulentwicklungsberater nachgefragt und auf entsprechende Materialien und Kurse zurückgegriffen. Die lokale Gemeinschaft, externe Schulentwicklungsberater oder auch Wissenschaftler spielen eine wichtige Rolle im Anstoß von Veränderungen. Sie können Schulen dazu anstoßen, über Schulentwicklung nachzudenken, oder bereits angefangene Entwicklungsprozesse durch ihre Expertise unterstützen. Aber auch Entwicklungen in der schulischen Umwelt selbst wie veränderte Schülerzahlen oder das Gebot der Inklusion von Schülern mit Förderbedarf kann schulische Entwicklungen in positiver, aber auch in negativer Weise forcieren (vgl. Hubermann & Miles 1984; Reezigt 2001; Visscher & Coe 2002).

Solche Erklärungen bleiben jedoch systemimmanent, d. h., man geht davon aus, dass alle beteiligten Akteure daran orientiert sind, Schulinspektionen zur Verbesserung von innerschulischen Entwicklungsprozessen zu nutzen, und sie lediglich in die Lage versetzt werden müssen, dementsprechend zu handeln. Einen anderen Erklärungsansatz liefern die auf der Basis der Interviewstudie durchgeführten objektiv-hermeneutischen Rekonstruktionen der Bezugnahmen schulischer Akteure auf den Steuerungsimpuls Schulinspektion. Diese verweisen auf ein Verhältnis von Schulinspektion und Schule, das als lose Kopplung (vgl. Weick 1976) beschrieben werden kann (s. Kapitel 3.3).

3.2 Akteurspezifische und konzeptionspezifische Wahrnehmung und Verarbeitung der Schulinspektion

Die in diesem Kapitel gebündelten Ergebnisse greifen zwei Vorannahmen des Projekts auf: Zum einen wird mit der Forschungsperspektive der Educational Governance davon ausgegangen, dass es für die Analyse der Wirkmechanismen der Schulinspektion von zentraler Bedeutung ist, die unterschiedlichen Rekontextualisierungen (vgl. Fend 2006) des Verfahrens durch die verschiedenen Akteure und etwaige sich daraus ergebende Inkonsistenzen in der Verarbeitung und Nutzung des Instruments zu berücksichtigen. Zum anderen wurde davon ausgegangen, dass sich abhängig vom Inspektionstyp unterschiedliche Wirkungen und Formen der Handlungskoordination einstellen. Insbesondere interessiert die Frage danach, ob unterschiedliche Inspektionsvariationen die Akzeptanz des Steuerungsinstruments v. a. bei schulischen Akteuren vermindern oder erhöhen, da die Akzeptanz des Verfahrens als wesentliche Voraussetzung einer „Schulentwicklung durch Einsicht“ gilt (s. Kapitel 1).

Akteurspezifische Rekontextualisierungen der Schulinspektion im Mehrebenensystem Schule

Auf Basis der inhaltsanalytischen Auswertung von Interviews mit Vertretern und Vertreterinnen der administrativen Leitungsebene der Schulinspektionen, Schulleitungen, Schulinspektionsteams, Schulaufsichten und Schulträgern finden sich bundeslandübergreifend durchaus akteurspezifische Rekontextualisierungen des Steuerungsimpulses Schulinspektion: Schulaufsichten sehen so zum Beispiel die Schulinspektion vermittelt über ihre Aufgabe des Abschlusses von Zielvereinbarungen mit den inspizierten Schulen betont als Kontrollinstrument. Schulleitungen nehmen Schulinspektion dagegen eher als externe Impulssetzungen zur eigenverantwortlichen Schulentwicklung wahr. Auf der administrativen Leitungsebene zeigen sich jeweils die komplexesten Vorstellungen zu Zielstellungen und Funktionsweisen. Wenig sichtbar wird in den Interviews dagegen eine eigenständige strategische Nutzung der Schulinspektionsergebnisse durch regionale Akteure der Schulgestaltung. Während sich bezogen auf die Schulträger summieren lässt, dass diese das Instrument Schulinspektion nur in geringem Maße für ihre eigenen Aufgaben und Interessen nutzen können, so ist auch bei Schulleitungen eine betont reaktiv-rezeptive Haltung zu beobachten.

Diese Ergebnisse deuten darauf hin, dass es schwierig ist, von dem einen Steuerungsimpuls Schulinspektion zu sprechen. Die akteurspezifischen Rekontextualisierungen der Schulinspektion können zu widersprüchlichen Deutungen und Nutzungen des Instruments führen, die wiederum dessen Wirkungen beeinflussen können: So könnte ein vonseiten der Schulinspektionsakteure als Instrument der Schulentwicklung kommuniziertes Konzept, das aber im Nachgang durch die Schulaufsicht als Kontrollinstrument genutzt wird, dazu führen, dass schulische Akteure ambivalent oder auch ablehnend auf die Schulinspektion Bezug nehmen. Auffällig ist jedoch ein weiteres Ergebnis des Verbundprojekts, das im Folgenden dargestellt wird: Die Akzeptanz der Schulinspektion ist in allen untersuchten Bundesländern erstaunlich hoch.

Relevanz von Konzeptvariationen für die Wahrnehmung und Verarbeitung der Schulinspektion

Neben der Frage nach dem Zusammenhang von Schulinspektion und Schulentwicklung stellt die Frage, ob die länderspezifischen Ausprägungen von Schulinspektionsverfahren eine messbare Auswirkung auf die Wahrnehmung und Akzeptanz dieses Steuerungsimpulses aus der Sicht von Schulleitungen haben, eine weitere zentrale Analyseperspektive der quantitativen Schulleitungsbefragung dar. Auch diese Frage kann allerdings insgesamt eher verneint werden.

Im Ehren-Visscher-Modell wird die Initiierung von Schulentwicklungsmaßnahmen als abhängige Variable betrachtet, deren Ausprägung u. a. von den Merkmalen der Schulinspektion – vermittelt über schulische Merkmale und Reaktionen – beeinflusst wird. Vor diesem Hintergrund zeigen sich zwar in vielen der eingesetzten Skalen signifikante Differenzen zwischen den Bundesländern; diese Unterschiede sind aber mehrheitlich mit geringen praktischen Bedeutsamkeiten verbunden. Lediglich ein Muster zieht sich durch alle verwendeten Indikatoren: Die Schulleitungen in Niedersachsen äußern eine niedrigere Akzeptanz der Schulinspektion insgesamt, thematisieren häufiger die mit einer Inspektion verbundenen Belastungen und nehmen weniger externe Unterstützungen als ihre Kolleginnen und Kollegen in den anderen drei Bundesländern wahr. Niedersachsen ist das einzige der einbezogenen Bundesländer, in dem das Instrument der Nachinspektion verwendet wird. Die mit den niedersächsischen Schulinspektionsverfahren für die Schulen einhergehende Option, als „failing school“ identifiziert zu werden und daraufhin eine Nachinspektion zu erfahren, wirkt sich offenbar auf die gesamte Akzeptanz des Verfahrens und damit einhergehender Bewertungen vonseiten der Schulleitungen aus.

Abgesehen davon scheint die spezifische Schulinspektionsvariante für die Wahrnehmung und Kommunikation des Instruments Schulinspektion aber eine eher untergeordnete Rolle zu spielen, was auch von den qualitativen Interviewanalysen bestätigt wird. Die inhaltsanalytischen Auswertungen der Experteninterviews zu aktorspezifischen Ziel- und Wirkungsvorstellungen der vier untersuchten Schulinspektionskonzeptionen ergaben hinsichtlich der Bedeutung der Konzeptvariationen ein ähnliches Bild wie die quantitative Teilstudie: Anhand der Nennungshäufigkeiten der identifizierten Vorstellungen zu Zielstellungen und Funktionsweisen der Schulinspektion lässt sich eine größere Konsistenz der Schulinspektionsvorstellungen in den verschiedenen Akteursgruppen im Vergleich zu den verschiedenen Bundesländern ermitteln. Das bedeutet, dass sich eher die Sichtweisen der Vertreter und Vertreterinnen der Leitungsebenen der Schulinspektion, der Schulinspektionsteams, der Schulaufsichten oder der Schulleitungen ähneln als die Vorstellungen der befragten Akteure je Bundesland.

Eine mögliche Erklärung für die geringe Bedeutung der Konzeptvariationen liefern Ergebnisse einer exemplarischen längsschnittlichen Dokumentenanalyse: Der eigentliche Steuerungsimpuls ist eher unspezifisch und durch eine schrittweise Implementationslogik gekennzeichnet, die daran orientiert ist, Akzeptanz herzustellen. Das Instrument ist insofern nicht durch eine konsistente Vorabbestimmung der Steuerungsimpulsion gekennzeichnet, sondern der Steuerungsimpuls entfaltet sich nach

und nach. Diese im Rahmen der Dokumentenanalyse exemplarisch durchgeführte Rekonstruktion der Implementationslogik von Schulinspektionen wird auch durch die Analysen anderer Teilprojekte und für weitere Bundesländer gestützt. So weisen die objektiv-hermeneutischen Rekonstruktionen ebenfalls auf eine „nachträgliche Erfindung“ des Steuerungsimpulses durch administrative Akteure hin (s. Kapitel 3.3). Darüber hinaus gleichen sich die zu Beginn der Implementation ggf. gegensätzlich erscheinenden Konzeptionen nach und nach an. So integriert z. B. das ursprünglich rein qualitative Verfahren in Baden-Württemberg nun auch standardisierte Vorabbefragungen analog zu den Konzeptionen in Thüringen und Mecklenburg-Vorpommern.

In diesem Zusammenhang ist bemerkenswert, dass – über alle Schulleitungen betrachtet, d. h. unabhängig vom Bundesland – eine durchschnittlich bis leicht überdurchschnittlich ausgeprägte allgemeine Akzeptanz von Schulinspektionsverfahren zu konstatieren ist (s. Tabelle 2).

Tabelle 2: Reaktionen der Schule aus Schulleitungssicht im Ländervergleich (Erster Messzeitpunkt)								
	BW (N = 76–77)	MV (N = 15)	NI (N = 486–495)	TH (N = 24–26)	Ges. (N = 603–611)	t- Wert**	Signi- fikanz	d
Skala (Beispielitem)	Mittelwert* Standardabweichung							
Allgemeine Akzeptanz SI (Schulinspektionen sind für die Arbeit der Schulen sehr wichtig.)	2.65 .65	2.82 .58	2.50 .68	2.99 .64	2.55 .68	-1.80	n.s.	-
Negative Folgen der SI (Die Schulinspektion hatte zusätzliche Belastungen zur Folge.)	2.40 .69	2.49 .67	2.66 .60	2.10 .59	2.60 .62	3.47	.001	.43
Ablehnende Sichtweise auf SI (Ich beobachte in meinem lokalen Umfeld, dass die Schulinspektion zu einer Stigmatisierung von Schulen führt.)	1.79 .58	1.77 .47	2.14 .53	1.90 .37	2.08 .54	5.45	.000	.66

BW = Baden-Württemberg, MV = Mecklenburg-Vorpommern, NI = Niedersachsen, TH = Thüringen * Antwortskala: 1 „trifft überhaupt nicht zu“ bis 4 „trifft voll und ganz zu“
 ** Die t-Tests wurden zwischen den Ländern Niedersachsen und Baden-Württemberg berechnet.

Dieser Befund wird dadurch untermauert, dass eine generell ablehnende Haltung gegenüber der Schulinspektion in den vier untersuchten Bundesländern nicht beobachtet werden kann. Auch in Bezug auf spezifische Merkmale der Schulinspektion von den Schulleitungen ist eine eher positive Bewertung durch die Schulleitungen zu beobachten. Besonders trifft dies für die Wahrnehmung des Inspektorenteams und

die Einschätzung des Inspektionsberichtes zu. Auch die Qualität der Schulinspektion erscheint in allen vier untersuchten Bundesländern aus Sicht der Schulleitungen in einem eher positiven Licht. Damit ergibt sich für die quantitative Schulleitungsbefragung der zentrale Befund einer konzeptunspezifischen hohen Akzeptanz der Schulinspektionsverfahren bei gleichzeitig marginalen Effekten des Instruments auf die Schulentwicklung. Akzeptanz und Wirkung fallen also – anders als in der Steuerungsvorstellung einer Schulentwicklung durch Einsicht konzipiert angenommen – auseinander. Auch dieser Befund ergab sich bereits in anderen Untersuchungen zu deutschen Schulinspektionssystemen (vgl. z. B. Gärtner et al. 2009).

3.3 Schulinspektionsbezogene Handlungskoordination

Die schulinspektionsbezogene Handlungskoordination wurde im Verbundprojekt über zwei methodische Zugänge erfasst: Zum einen wurden mithilfe des thematischen Kodierens akteurspezifische Positionierungen und Verhältnisbestimmungen innerhalb der schulinspektionsbezogenen (regionalen) Akteurkonstellation analysiert. Diese Analyseergebnisse stellen also die subjektiven Sichtweisen der befragten Akteure auf die schulinspektionsbezogene Akteurkonstellation dar. Zum anderen wurde mithilfe der Objektiven Hermeneutik die latente Strukturlogik der schulinspektionsbezogenen Handlungskoordination rekonstruiert, also die Ebene der Handlungskoordination, die den Akteuren selbst nicht notwendigerweise bewusst ist.

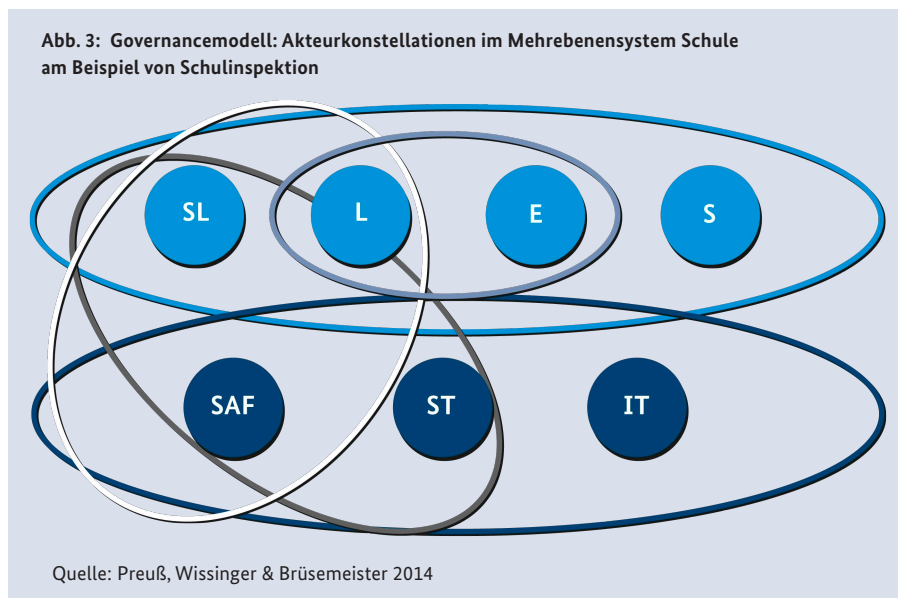
Akteurspezifische subjektive Positionierungen und Verhältnisbestimmungen innerhalb der schulinspektionsbezogenen Akteurkonstellation

Auf Basis der metatheoretischen Perspektive der Educational-Governance-Forschung (vgl. Altrichter et al. 2007; Kussau & Brüsemeister 2007; Altrichter & Maag Merki 2010) wird davon ausgegangen, dass sich die Durchsetzung der Schulinspektion nicht unabhängig von Akteurkonstellationen sowie der Rezeption und Perzeption durch die anderen Akteure einstellt, sondern in Aushandlungsprozesse, Interdependenzen und in Bearbeitungen durch die anderen Akteure eingebettet ist – im Sinne von „joint action“ (vgl. Blumer 1969).

Einfluss und Wirkung der Schulinspektion werden nicht allein aus einem hierarchisch organisierten Steuerungsverständnis heraus verstanden, sondern aus einem System über mehrere Ebenen organisierter Akteurkonstellationen (vgl. Preuß et al. 2012). Empirisch werden Koordinationsverhältnisse (vgl. Altrichter 2014) in institutionalisierten Handlungszusammenhängen eines schulischen Mehrebenensystems untersucht. Jeder Akteur handelt in wechselseitigen Abhängigkeiten zu anderen Akteuren. Die Governance-Perspektive zieht in anderen Worten auch eine Umorientierung der Untersuchung von Handlungen hin zur Untersuchung von Handlungskonstellationen nach sich (s. Abbildung 3). In Bezug auf das Verbundprojekt bedeutet das: In den thematisch kodierten Auswertungen der qualitativen Experteninterviews

zeigt sich, wie die „belief systems“ (Sabatier 1993) einzelner Akteure durch den Einfluss von Schulinspektion überhaupt erst klar zutage treten und dass diese für die Handlungskoordination im Kontext der Schulinspektion eine Bedeutung haben.

Abb. 3: Governancemodell: Akteurkonstellationen im Mehrebenensystem Schule am Beispiel von Schulinspektion



Schulinspektion intensiviert das Interdependenzbewusstsein der einzelnen Akteure. Diesen wird unter dem Einfluss von Schulinspektion sowohl ihre eigene Rolle in der Konstellation bewusst(er) als auch die Rolle der anderen. Vor diesem Hintergrund werden im Folgenden die akteurspezifischen subjektiven Positionierungen und Verhältnisbestimmungen innerhalb der schulinspektionsbezogenen Akteurkonstellation entlang der verschiedenen Akteursgruppen skizziert.

Inspektionsteams. Das Inspektionsteam nimmt sich als neutraler Partner wahr und versucht, über Lob- und Vertrauenskultur schulische Akzeptanz zu gewinnen. Das Instrument Schulinspektion wird nicht als Kontrolle gesehen, sondern als hilfreich für Schulentwicklung. Als problematisch sieht das Inspektionsteam, Wissen über Schule zurückzuhalten; insbesondere wenn die Inspektorinnen und Inspektoren von der Profession her Lehrerinnen und Lehrer sind. Sie haben viel Erfahrungswissen über Schule, sollen dies aber nicht preisgeben und sollen Schule nicht beraten.

Schulleitungen. Die Schulleitungen stehen aufgrund der ihnen rechtlich zugewiesenen Gesamtverantwortung für die Schule, deren Qualitätssicherung und -entwicklung vor der Herausforderung, Lehrkräften, Eltern und Schülerinnen und Schülern die Idee der Schulinspektion vermitteln zu müssen. Die Rolle der Schulleitung ist in diesem Zusammenhang verantwortungsvoll und komplex. Im „transformative leadership concept“ nach Geijsel (2001) sind entsprechende Aspekte beschrieben: Dazu gehört u. a. die aktive Unterstützung der Lernprozesse des Lehrerkollegiums und der gesamten Schulorganisation. Schulleitungen sind dementsprechend auch

zentrale Ansprechpartner und Ansprechpartnerinnen der Schulinspektion, und von ihnen wird erwartet, dass sie in Fragen der Schulentwicklung initiativ sind (vgl. Wissinger 2014). Daraus ergibt sich, dass die Handlungsmöglichkeiten aller weiteren Akteure in der Schule von der Qualität der Führung abhängen. Neu ist, dass sich zwischen Schulleitungen und Inspektion ein strukturelles Bündnis bildet, das Lehrkräfte den Schulleitungen in gewisser Weise „entrücken“ kann. Dadurch kann Schulentwicklung sogar erschwert werden; es kann also das Gegenteil des mit der Inspektion Beabsichtigten bewirkt werden. Unter den Bedingungen der durch die Schulinspektion veränderten regionalen Handlungskonstellationen wünschen sich die Schulleitungen mit Blick auf die Qualitätsentwicklung ihrer Schule sowohl im Prozess der Schulinspektion als auch zugunsten einer nachhaltigen Entwicklung im Anschluss danach Unterstützung von außen, insbesondere durch die Schulaufsicht oder auch durch Fachberaterinnen und Fachberater.

Lehrkräfte. Einerseits nehmen die Lehrkräfte Schulinspektion als Aufsicht und Kontrolle wahr, andererseits als Chance, von außen schulische Praxis in ihren Vor- und Nachteilen gespiegelt zu bekommen. Sie kooperieren in erster Linie mit der Schulleitung, zeigen aber auch Rückzugsverhalten, insbesondere durch unangenehme Erfahrungen im Zusammenhang mit der Rückmeldesituation. Diese beziehen sich auf Situationen der Kränkung, die dadurch entstehen, dass die Fremdwahrnehmung der Schulinspektion und die Eigenwahrnehmung der Lehrkräfte nicht übereinstimmen. Von der Schulinspektion wünschen sie sich ein persönliches Feedback über die Qualität ihres Unterrichts. In Einzelfällen nutzen die Lehrkräfte Schulinspektion, um bestehende Konflikte mit der Schulleitung zu klären. Teilweise nutzen Lehrkräfte die Inspektion als „Bühne“ für bestehende Spannungen mit der Schulleitung, was darauf hindeutet, dass das Verhältnis von Schulleitung und Lehrkräften ein wechselseitiges Abhängigkeitsverhältnis ist (vgl. Preuß et al. 2012) und Fragen von Erfolg oder Misserfolg durch Schulinspektion angestoßener Schul- und Unterrichtsentwicklungsprozesse im Kontext der Thematisierung der Schule als Organisation (vgl. Wissinger 2013) und der Kooperation im Kollegium einer Schule gesehen werden müssen (vgl. Keller-Schneider et al. 2013).

Schulaufsicht. Die Schulaufsicht hat ein positives Verständnis von Schulinspektion. Sie wird als ergänzendes und hilfreiches Instrument der Aufsicht verstanden. Die Schulaufsicht ist durch die Einführung der Schulinspektion und im Rahmen der Durchführung von Zielvereinbarungen mit der Schule mit der neuen Rolle und Aufgabe konfrontiert, nicht mehr nur zwischen Bildungsverwaltung und Schule vermittelnd zu kommunizieren, sondern auch den Akteur Schulinspektion zu berücksichtigen. Durch den neuen Akteur Schulinspektion werden für die Schulaufsicht Fragen aufgeworfen, wie sie das Verhältnis zwischen Aufsicht, Beratung und Inspektion balancieren soll. Ein ungewollter Effekt in der Beziehung zur Inspektion besteht darin, dass diese der Aufsicht verschärft den Kontrollaspekt zuschiebt. Außerdem beinhaltet die Zusammenarbeit zwischen Schulleitung und Inspektion eine strukturelle Konkurrenz zur Schulaufsicht.

Schulträger. Die Schulträger sehen sich in einer passiven Rolle, da sie aufgrund ihrer gesetzlichen Funktion das pädagogische Wissen des Inspektionsberichts kaum nutzen können. Einige Schulträger gehen jedoch mit der Schule das ideale

Bündnis ein, gemeinsam Bildung zu optimieren. Bildung hat aus ihrer Sicht einen so hohen Stellenwert, dass es sich für den Schulträger lohnt, zu investieren und Fragen jenseits von Zuständigkeiten zu beantworten. Dies geschieht jedoch unbeeinflusst von der Schulinspektion. Durch die Nichteinbindung der Schulträger in das Inspektionsgeschehen geht, so die Hypothese, Potenzial verloren bzw. wird gar nicht erst genutzt.

Eltern. Auch die Eltern nehmen Schulinspektion ambivalent wahr: Einerseits interessieren sie sich für Schulinspektion dann, wenn es um einen Vorteil für ihr eigenes Kind geht. Andererseits sehen sie sich – wie die Schulträger – über Schulinspektion wenig informiert. Auch hier bleibt offensichtlich Potenzial ungenutzt.

In diesen kurzen Darstellungen der subjektiven Positionierungen und Verhältnisbestimmungen der an Schulinspektion beteiligten Akteure deutet sich an, dass sich aus den qualitativ-kodierenden Analysen zwar keine eindeutigen Zusammenhänge zwischen Schulinspektion und Schulentwicklungsprozessen ablesen lassen, es aber allein durch das Auftauchen des neuen Akteurs Schulinspektion sehr wohl zu neuen Verhältnisbestimmungen innerhalb der Akteurkonstellation kommt, die von den Akteuren auch als solche manifest wahrgenommen werden. Insbesondere die Beziehungen zu den Schulleitungen sind intensiv, die im Verfahren auch formal besonders herausgehoben werden. Bearbeitungsintensiv scheint außerdem – trotz positiver Bezugnahmen aufeinander – das Verhältnis von Schulinspektion und Schulaufsicht. Allein daraus lässt sich jedoch auf Ebene der selbst berichteten Wirkungen noch kein eindeutig negativer oder positiver Effekt auf Schulentwicklung feststellen.

Die latente Strukturlogik der schulinspektionsbezogenen Handlungskoordination

Die im Rahmen des Verbundprojekts durchgeführten objektiv-hermeneutischen Rekonstruktionen des qualitativen Interviewmaterials ermöglichen im Sinne einer rekonstruktiven Reformulierung des Educational-Governance-Ansatzes (vgl. Dietrich 2014) eine Relationierung der Bezugnahmen auf Schulinspektion im Mehrebenensystem und damit explizit die Rekonstruktion der Strukturlogik der schulinspektionsbezogenen Handlungskoordination. Dabei geht es jedoch nicht – wie im Abschnitt zuvor – um die Abbildung manifest-subjektiver Bewertungen und Sichtweisen, sondern jeweils um die latente Strukturlogik der deutenden Umgangsweisen mit dem Steuerungsimpuls. Dies ermöglicht darüber hinaus, zwischen der in Kapitel 1 skizzierten Programmatik einer evidenzbasierten Schulsystemsteuerung und dem empirisch vorfindbaren schulinspektionsbezogenen Governance-Regime zu differenzieren. Aus den objektiv-hermeneutischen, auf der Interviewstudie basierenden Rekonstruktionen ergeben sich dabei die folgenden Schlaglichter auf die Etablierung der Schulinspektion.

Analog zu den Ergebnissen der längsschnittlich angelegten Dokumentenanalyse zur Implementationslogik der Schulinspektion (s. Kapitel 3.2) laufen die Rekonstruktionen der steuerungstheoretischen Vorstellung entgegen, die Einführung der

Schulinspektionen begründe sich in einer vorgängig ausformulierten Programmatik einer „Schulentwicklung durch Einsicht“, der zufolge die Schulinspektion in erster Linie mittels der Bereitstellung verobjektivierter Informationen über die Schulen innerschulische Entwicklung initiieren und fundieren solle. Vielmehr scheinen die Akteure auf der **administrativen Ebene** mit der Möglichkeit oder Herausforderung konfrontiert, den politischen Beschluss zur Einführung der Inspektion erst im Nachhinein inhaltlich zu füllen und zu legitimieren. In diesem Zusammenhang spielt die in Kapitel 1 umrissene Programmatik evidenzbasierter Steuerung eine Rolle, indem sie zentraler Referenzpunkt der manifesten Darstellung der landesspezifischen Ausgestaltungen der Schulinspektionen ist. Sie dient damit in erster Linie der Legitimation und einer Ausfüllung einer programmatischen Leerstelle. Jedoch laufen die mit der Schulinspektion verknüpften latenten Steuerungsvorstellungen und -hoffnungen auf der administrativen Ebene jener Vorstellung der Entfaltung eines steuernden zwanglosen Zwangs der im Zuge der Inspektionen generierten, besseren – weil methodisch kontrolliert und systematisiert erzeugten – Evidenzen (vgl. Dietrich & Lambrecht 2012) mitunter diametral entgegen. Jenseits der fallspezifischen Ausprägungen eint die rekonstruierten Fälle eine Fokussierung auf die performative Kraft der Schulinspektion: Demnach sind es der Schulinspektionsbesuch, dessen Ankündigung und dessen Vorbereitung, die dieser Konzeption zufolge steuernde Wirkung entfalten – unabhängig von den Evaluationsergebnissen. In diesem Sinne arbeitet die Schulinspektion in Richtung einer Etablierung bzw. weiteren Tradierung einer Schulentwicklung als systematisierte zyklische Qualitätsentwicklung. Indem die schulischen Akteure mit der Schulinspektion konfrontiert werden, müssen sie sich mit der in dieses Instrument eingeschriebenen Logik der „Qualitätsentwicklung“ auseinandersetzen und sich dieses damit zumindest bis zu einem gewissen Grad aneignen (vgl. Dietrich 2012). Dieser latente Steuerungsmechanismus erlaubt einen weitgehenden Verzicht auf unmittelbare mit den Inspektionsergebnissen verbundene Kontrollmomente. Damit laufen die manifeste Programmatik einer „Entwicklung durch Einsicht“ und die rekonstruierte latente Steuerungskonzeption zwar auseinander, ihr Spannungsverhältnis verläuft jedoch quer zur Frage, inwieweit Schulinspektion ein Unterstützungs- oder Kontrollinstrument darstellt.

Vor dem Hintergrund der dominanten Orientierung am programmatischen Ziel, Schulentwicklungsprozesse zu initiieren, sind die rekonstruierten Bezugnahmen der Schulinspektorinnen und Schulinspektoren durch eine noch pointiertere Distanzierung von einem Verständnis von Schulinspektion als Kontrollinstrument geprägt. Geleitet werden sie in ihrer Praxis von dem paradox anmutenden Ziel, die schulischen Akteure zu einer „Schulentwicklung durch Einsicht“ zu bewegen: Ihr Auftreten zielt nicht allein auf eine empirisch zutreffende Abbildung und Bewertung der Situation vor Ort. Gleichzeitig repräsentieren sie die Schulinspektion und versuchen in ihrem Auftritt, den schulischen Akteuren die Schulinspektion nahezubringen und sie so zu einer produktiven Nutzung der generierten Daten zu animieren, was auf eine pädagogische Adressierung der schulischen Akteure verweist. Das sich darin andeutende zentrale Strukturproblem ergibt sich daraus, dass die Programmatik einer „Schulentwicklung durch Einsicht“ letztlich keine autonome

Umgangsweise mit der Schulinspektion seitens der schulischen Akteure vorsieht: Eine solche müsste auch in Rechnung stellen und akzeptieren, dass die schulischen Akteure die Relevanz der generierten Evidenzen selbst und autonom auf der Basis einer eigenen und letztlich eigensinnigen Perspektive bewerten. Genau dies würde jedoch den mit dem Steuerungsinstrument verbundenen Steuerungsanspruch, der erst den mit der Implementierung verbundenen Aufwand rechtfertigt, unterminieren. Das Konzept der „Entwicklung durch Einsicht“ setzt vielmehr voraus, dass die schulischen Akteure quasi zwangsläufig auf der Basis der Evaluationsergebnisse „zur Einsicht kommen“. Mit der Brüchigkeit dieser Konzeption sehen sich die Schulinspektoren in ihrer Praxis vor Ort und in der Interaktion mit den schulischen Akteuren konfrontiert.

Spätestens der Schulinspektionsbesuch stellt die schulischen Akteure vor die Aufgabe, die abstrakten programmatischen Formeln, die die Schulinspektion begleiten, in ein Verhältnis zur eigenen Praxis zu setzen. Mit Blick auf die Schulleitungen deutet sich an, dass der Nutzen der Schulinspektion für die Schule zumindest nicht unmittelbar evident erscheint. Gleichwohl folgt daraus keine offensive Kritik oder Infragestellung. Vielmehr wird ungeachtet der geringen Anschlussfähigkeit versucht, die Relevanz der Schulinspektion zu plausibilisieren, um so im Kollegium für das notwendige Maß an Akzeptanz zu sorgen. Jedoch verweist die affirmative Umgangsweise nicht auf eine substanzielle Auseinandersetzung mit der Schulinspektion. Letztlich scheint der „Steuerungsimpuls“ – zumindest vorerst – eher hinsichtlich einer Etablierung innerschulischer formalisierter Evaluationsschleifen wirkmächtig: Zwar ist damit nicht gegeben, dass die implementierte Form der Qualitätsentwicklung die unmittelbare schulische Praxis tangiert. Unabhängig davon wird jedoch eine für die Schuladministration unmittelbar nachvollziehbare Interaktionslogik in der Schule etabliert: So werden innerschulisch – und im Austausch mit der Schulaufsicht in Gestalt von Zielvereinbarungen – Ziele formuliert, die operationalisiert und damit formal überprüfbar erreicht oder verfehlt werden können. Gleichzeitig wird damit die Schule als Handlungseinheit sichtbar – unabhängig davon, inwieweit diese Gestalt nach innen Kohärenz stiftend auf die pädagogische Praxis wirkt. Die innerschulische Realität wird damit in eine Form übersetzt, die anschlussfähig an die Schuladministration ist, womit die Schulinspektion das oben beschriebene administrative Steuerungsproblem tatsächlich bearbeitet.

Die Bezugnahme der Lehrkräfte ist durch die Frage nach der Relevanz der Schulinspektion für ihre unmittelbare unterrichtliche Praxis geprägt. Darin begründet sich die in den Interviews erkennbar werdende Bedeutsamkeitszuschreibung der halbstündigen Unterrichtshospitationen. Entgegen den Konzeptionen werden diese z. T. als bedrohliche, auf die individuelle Unterrichtspraxis bezogene Überprüfungssituationen gedeutet. Diese Deutung sowie die sich jenseits der positiven Kommentierungen des „Blicks von außen“ bemerkbar machende ablehnende, z. T. in offene Empörung umschlagende, jedoch recht diffus bleibende Haltung gegenüber der Inspektion können auf ein mangelndes Passungsverhältnis zwischen der jeweiligen Schulkultur (vgl. Helsper 2008) bzw. dem professionellen Habitus der Lehrenden und der Strukturlogik der Adressierung der schulischen Akteure durch

den Steuerungsimpuls gedeutet werden. Demzufolge sind es also nicht zuvorderst einzelne Ergebnisse der Inspektionen, die irritieren, auch nicht der vermeintliche Kontrollcharakter des Steuerungsinstrumentes, der ja zumindest nicht auf die Kontrolle der individuellen Praxis der Lehrerinnen und Lehrer zielt, sondern die Art des Umgangs mit schulischer Praxis – etwa die konsequente Adressierung der Schule als eine Handlungseinheit, deren Praxis als in formalisierter und mehr oder weniger standardisierter Form evaluierbar und damit letztlich extern steuerbar definiert wird. Kehrseite dieser Adressierung ist die Nichtadressierung und damit die Missachtung der Lehrerinnen und Lehrer als professionelle und als traditionell primäre Ansprechpartner bezüglich der Ausgestaltung der unterrichtlichen Praxis.

Diese Lesart konvergiert schließlich mit der oben beschriebenen Bezugnahme der Schulinspektorinnen und Schulinspektoren auf das Konzept der Schulinspektion. Deren pädagogische Adressierung und Wunsch, über die Bekanntgabe der Inspektionsergebnisse hinaus beratend tätig werden zu dürfen, können als handlungspraktische Reaktion auf das in der Gesamtschau erkennbar werdende und als lose Kopplung zu beschreibende Verhältnis von Schulinspektion und schulischer Praxis gedeutet werden. Die lose Kopplung (vgl. Weick 1976) ermöglicht auf schulischer Seite mindestens partiell, die schulkulturellen und berufshabituellen Eigensinnigkeiten von der Strukturierungsmacht der Schulinspektion abzuschirmen. Dies begründet plausibel das Ausbleiben einer manifesten Kritik an der Schulinspektion und darüber hinausgehend die rhetorisch positive Bezugnahme auf das Programm einer Evidenzbasierung, die sich in den quantitativen Analysen zeigen. Das bedeutet: Die positiven Äußerungen zur Schulinspektion und die verhältnismäßig hohen allgemeinen Akzeptanzwerte der quantitativen Schulleitungsbefragung verweisen nicht zwingend auf eine materiale Auseinandersetzung schulischer Akteure mit den Inspektionsergebnissen im Sinne einer Evidenzbasierung von Schulentwicklung. Andersherum stattdessen das Programm der Evidenzbasierung, so sie sich in entsprechenden evaluativen Befragungen abbildet (z. B. Pietsch 2011), die Schulinspektion mindestens insoweit mit Legitimität aus, wie diese nicht am Grad der Erfüllung der programmatisch behaupteten Wirksamkeit bemessen wird. Gleichzeitig werden mit den Schulinspektionen, im Zuge der Vorbereitung der Inspektion seitens der schulischen Akteure, des Inspektionsbesuchs, der Ergebnispräsentation und der sich in aller Regel anschließenden Zielvereinbarungen auf der Basis der Inspektionsberichte mit der Binnenlogik der Schulverwaltung kompatible Kommunikationsformen etabliert. In diesem Sinne bearbeitet die Schulinspektion das in den Bezugnahmen auf dieser Ebene aufscheinende administrative Steuerungsproblem; inwieweit damit die schulische Praxis tangiert wird, inwieweit damit also im emphatischen Sinne Schulentwicklung initiiert wird, bleibt jedoch zumindest offen.

3.4 Zusammenfassung: Zur empirischen Rekonstruktion des schulinspektionsbezogenen Governance-Regimes

Die vorangegangene Darstellung der Ergebnisse des Verbundprojekts „Schulinspektion als Steuerungsimpuls zur Schulentwicklung und seine Realisierungsbedingungen auf einzelschulischer Ebene“ lassen sich im Hinblick auf die empirische Rekonstruktion des schulinspektionsbezogenen Governance-Regimes folgendermaßen zusammenfassen und systematisieren:

- **Vager Steuerungsimpuls:** Es lassen sich zwar akteurspezifische Rekontextualisierungen der Schulinspektion ausmachen, die bundeslandspezifischen Konzeptvariationen spielen dagegen für die Wahrnehmung des Instruments eine eher untergeordnete Rolle, was u. a. auf die Vagheit des Steuerungsimpulses zurückzuführen ist, der in erster Linie darauf ausgerichtet ist, Akzeptanz herzustellen. Darüber hinaus nähern sich die Inspektionskonzeptionen sukzessive aneinander an.
- **Herstellung von Anschlussfähigkeit:** Zu beobachten sind Verschiebungen der Handlungskoordination innerhalb der untersuchten Akteurkonstellation. Dies drückt sich allgemein z. B. in einem verstärkten Interdependenzbewusstsein der einzelnen Akteure aus. Darüber hinaus findet vermittelt über das Instrument der Schulinspektion ein formaler Umbau der Schulsystemsteuerung statt. Auffällig ist hier insbesondere eine Veränderung der schulischen Legitimationsweisen; diese funktionieren nun maßgeblich über die Etablierung formaler Qualitätsmanagementzirkel. Unter anderem dadurch gelingt es, die schulische Realität in eine Form zu übersetzen, die anschlussfähig an die Schuladministration ist, und sie so steuerbar und überprüfbar zu machen. Insofern wird das administrative Steuerungsproblem mithilfe der Schulinspektion tatsächlich bearbeitet.
- **Performative Kraft der Schulinspektion und Vermittlungsbedürftigkeit von Evidenzen:** Die zentrale Wirkungsvorstellung der Schulinspektion lässt sich als Rationalisierung der Schulentwicklung durch Einsicht in Evidenzen charakterisieren. Insbesondere die objektiv-hermeneutischen Rekonstruktionen deuten jedoch darauf hin, dass die Schulinspektion weniger über die Evaluationsergebnisse als vielmehr durch die bloße Tatsache ihrer Existenz – d. h. ihre performative Kraft – Wirkungen entfaltet. Insgesamt ist fraglich, wie selbstwirksam die produzierten Evidenzen tatsächlich sind. Vielmehr scheint es einer Vermittlung der generierten Evidenzen und damit einer pädagogischen Adressierung der schulischen Akteure zu bedürfen. Ein eigenständiger Umgang schulischer Akteure mit den generierten Evidenzen wäre im Rahmen des Konzepts einer Schulentwicklung durch Einsicht – ernst genommen – dagegen eher kontraproduktiv.
- **Lose Kopplung von Schulinspektion und Schulentwicklung:** Bei relativ hoher Akzeptanz des Instruments auf Schulleitungsebene ist bundesländerübergreifend eine eher geringe bzw. uneindeutige Wirksamkeit des Steuerungsimpulses Schulinspektion auf die Schulentwicklung auszumachen. Als Erklärungsansatz kann hierfür die These herangezogen werden, dass sich die lose Kopplung im Mehrebenensystem Schule im Verhältnis von Schule und Schulinspektion reproduziert. Dies bedeutet, dass durch Schulinspektionen zwar die kommunikative

Anschlussfähigkeit zwischen Administration und schulischer Praxis erhöht werden kann; inwieweit sich daraus jedoch faktisch die intendierten Effekte in Bezug auf die Schulentwicklung und die Schulqualität ergeben, bleibt offen.

4. Schlussreflexion: Das triangulative Potenzial des Verbundprojekts

Die triangulative Anlage des Verbundprojekts hat in Bezug auf die folgenden Aspekte zu einem signifikanten Erkenntnismehrwert geführt:

- Durch die Verknüpfung der Analysen der Schulinspektionskonzeptionen mit den quantitativen Wirkungsanalysen kann die Bedeutung der Schulinspektionskonzeptionen für die Wahrnehmung und Wirkung des Instruments, wie sie beispielsweise im Ehren-Visscher-Modell postuliert wird, relativiert werden. Die längsschnittliche Analyse von Dokumenten der baden-württembergischen Fremdevaluation und die daraus abgeleitete Hypothese von Schulinspektionen als vagen Steuerungsimpulsen bieten hierfür außerdem einen Erklärungsansatz.
- Durch die Verknüpfung der Analyse der Programmatik einer evidenzbasierten Schulinspektion mit den am Interviewmaterial mithilfe der objektiven Hermeneutik rekonstruierten schulinspektionsbezogenen Handlungs- und Strukturlogiken konnte die Vorstellung einer Rationalisierung der Schulinspektion durch die Generierung empirischer Evidenzen relativiert werden.
- Für das immer wieder beobachtete Phänomen der hohen Akzeptanz deutschsprachiger Schulinspektionsverfahren bei schulischen Akteuren bei gleichzeitig geringen bzw. indifferenten Wirkungen konnte durch die Kombination von Ergebnissen der quantitativen Teilstudie und den objektiv-hermeneutischen Rekonstruktionen des Interviewmaterials ein Erklärungsansatz entwickelt werden: Die lose Kopplung im Mehrebenensystem Schule scheint sich im Verhältnis von Schule und Schulinspektion zu reproduzieren.

Im Rahmen des Verbundprojekts hat sich insbesondere die Kombination von qualitativen bzw. rekonstruktiven und quantitativen Verfahren bewährt. Hier ergab sich eine bemerkenswerte Arbeitsteilung. Im Allgemeinen wird qualitativen Verfahren in triangulativen Forschungsansätzen die Rolle der Exploration, quantitativen Verfahren dagegen die Rolle der Überprüfung der qualitativ-explorativ gewonnenen Thesen zugesprochen. In diesem Fall erwies sich jedoch ein umgekehrtes Vorgehen als ausgesprochen fruchtbar: Mithilfe der qualitativen und rekonstruktiven Analysen konnten z. T. ungeklärte quantitative Phänomene differenziert und aufgeklärt werden. Gleichzeitig hat sich gezeigt, dass sich Wirkungen der Schulinspektionen auf Schulentwicklungsprozesse nur bedingt in Form hochaggrierter quantitativer Daten abbilden lassen. Auf Basis qualitativer Einzelfallanalysen können dagegen spezifische schulische Akteurkonstellationen und Formen der Handlungskoordination herausgearbeitet werden, die die Anschlussfähigkeit der Schulinspektionsergebnisse an Schulentwicklungsprozesse fördern bzw. hemmen können. Potenziell ist damit eine zumindest teilweise Aufklärung der uneindeutigen Wirkungsweise der Schulinspektion möglich.

Die Ergebnisse des Verbundprojekts machen außerdem auf die Bedeutung einer spezifischen Unterscheidung aufmerksam, die mit Husfeld (2011) als Unterscheidung zwischen Wirksamkeit einerseits und Wirkung andererseits beschrieben werden kann. Diese Unterscheidung liegt quer zu der beispielsweise im Ehren-Vischer-Modell vorgenommenen Unterscheidung zwischen intendierten und nicht intendierten Effekten:

„Wirksam ist eine Maßnahme (...) dann, wenn sie einen Prozess so beeinflusst, dass gewünschte Ziele erreicht werden können. An der Zielerreichung ist der Grad der Wirksamkeit ablesbar. Die Wirkung einer Maßnahme ist im Gegensatz zur Wirksamkeit auch unabhängig von expliziten Zielen zu betrachten“ (Husfeldt 2011, S. 260).

Auf die Ergebnisse des Verbundprojekts gewendet bedeutet das: Ob Schulinspektionen die Schulentwicklung befördern und so durch sie Schulen vor dem Hintergrund eines spezifischen Modells guter Schule „besser“ werden, ist eine offene Frage. In jedem Fall stellen sie jedoch eine relativ erfolgreiche Bearbeitung klassischer bildungsadministrativer Steuerungsprobleme dar, indem sie vor dem Hintergrund einer losen Kopplung des Schulsystems die kommunikative Anschlussfähigkeit zwischen Administration und schulischer Praxis erhöhen. Schulinspektionen sind so betrachtet nicht wirkungslos.

5. Literatur

- Ackeren, I. van, Zlatkin-Troitschanskaia, O., Binnewies, C., Clausen, M., Dormann, C., Preisendörfer, P., Rosenbusch, C. & Schmidt, U. (2011). Evidenzbasierte Schulentwicklung: Ein Forschungsüberblick aus interdisziplinärer Perspektive. *Die Deutsche Schule*, 103 (2), S. 170–184.
- Altrichter, H. (2014). Governance: Steuerung und Handlungskoordination bei der Transformation von Bildungssystemen. In H. J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungssystem: Analysen zur Mehrebenenperspektive, Steuerung und Koordination*. Wiesbaden: VS.
- Altrichter, H. & Heinrich, M. (2007): Kategorien der Governance-Analyse und Transformationen der Systemsteuerung in Österreich. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance: Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: VS, S. 55–103.
- Altrichter, H. & Maag Merki, K. (Hrsg.) (2010). *Handbuch Neue Steuerung im Bildungssystem*. Wiesbaden: VS.
- Altrichter, H., Brüsemeister, T. & Wissinger, J. (Hrsg.) (2007). *Educational Governance: Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: VS.
- Blumer, H. (1969). *Symbolic Interactionism: Perspective and Method*. New Jersey: Prentice-Hall.
- Böhm-Kasper, O. & Selders, O. (2013). „Schulinspektionen sollten regelmäßig durchgeführt werden“? Ländervergleichende Analyse der Wahrnehmung und Akzeptanz von Schulinspektionsverfahren. In I. van Ackeren, M. Heinrich & F. Thiel,

- (Hrsg.): *Evidenzbasierte Steuerung im Bildungssystem? Die Deutsche Schule*. Beiheft 12, Münster u.a.: Waxmann, S. 121–153.
- Böttcher, W. & Keune, M. (2010). Funktionen und Effekte der Schulinspektion. In W. Böttcher, J. N. Dicke & N. Högbe (Hrsg.), *Evaluation, Bildung und Gesellschaft*. Münster, u. a.: Waxmann, S. 151–164.
- Böttcher, W. & Kotthoff, H. G. (Hrsg.) (2007). *Schulinspektion, Evaluation, Rechenschaftslegung und Qualitätsentwicklung*. Münster u. a.: Waxmann.
- Böttger-Beer, M. & Koch, E. (2008). Externe Schulevaluation in Sachsen: Ein Dialog zwischen Wissenschaft und Praxis. In W. Böttcher, W. Bos, H. Döbert & H.-G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive*. Münster u. a.: Waxmann, S. 253–264.
- Bonsen, M., Bos, W. & Rolff, H.-G. (2008). Zur Fusion von Schuleffektivitäts- und Schulentwicklungsforschung. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung: Daten, Beispiele, Perspektiven (15)*. Weinheim, München: Juventa, S. 11–39.
- Brimblecombe, N., Shaw, M. & Ormston, M. (1996). Teachers' Intention to Change Practice as a Result of OFSTED School Inspections. *Educational Management & Administration*, 24 (4), S. 339–354.
- Chapman, C. (2001). Changing Classrooms through Inspection. *School Leadership & Management*, 21 (1), S. 59–73.
- Dietrich, F. (2012). Schulinspektion im Kontext von „Deregulierung“: Eine rekonstruktive Annäherung an die Adaption des Steuerungsimpulses auf der Ebene der Schulleitung. In S. Hornberg & M. Parreira do Amaral (Hrsg.), *Deregulierung im Bildungswesen*. Münster u. a.: Waxmann, S. 187–204.
- Dietrich, F. (2014). Objektiv-hermeneutische Governanceanalysen. In K. Maag Merki, R. Langer & H. Altrichter (Hrsg.), *Educational Governance als Forschungsperspektive: Strategien, Methoden und Forschungsansätze in interdisziplinärer Perspektive*. Wiesbaden: Springer VS, S. 199–227.
- 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, Heft 1, S. 58–71.
- Ehren, M. C. M. & Visscher, A. J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54 (1), S. 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), S. 205–227.
- Fend, H. (2006). *Neue Theorie der Schule: Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: VS.
- Flick, U. (2007). *Qualitative Sozialforschung: Eine Einführung*. Reinbek: Rowohlt.
- Flick, U. (2011). *Triangulation: Eine Einführung*. Wiesbaden: VS.
- Gärtner, H., Hüsemann, D. & Pant, H. A. (2009). Wirkungen von Schulinspektion aus der Sicht betroffener Schulleitungen: Die Brandenburger Schulleiterbefragung. *Empirische Pädagogik* 23 (1), S. 1–18.
- Geijsel, F. (2001). *Schools and innovations: Conditions fostering the implementation of educational innovations*. University of Nijmegen.

- Gläser, J. & Laudel, G. (2009). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruktiver Untersuchungen*. Wiesbaden: VS.
- Helsper, W. (2008). Schulkulturen: Die Schule als symbolische Sinnordnung. *Zeitschrift für Pädagogik*, 53 (1), S. 63–80.
- Huberman, A. M., Miles, M. B. (1984). *Innovation up close: How school improvement works. Environment, development, and public policy. Public policy and social services*. New York: Plenum Press.
- Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation: Überblick zum Stand der Forschung. *Zeitschrift für Erziehungswissenschaft*, 14 (2), S. 259–282.
- Keller-Schneider, M., Albisser, S. & Wissinger, J. (Hrsg.) (2013). *Professionalität und Kooperation in Schulen: Beiträge zur Diskussion über Schulqualität*. Bad Heilbrunn: Klinkhardt.
- Kussau, J. & Brüsemeister, T. (2007). Educational Governance: Zur Analyse der Handlungskoordination im Mehrebenensystem der Schule. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance: Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: VS, S. 15–54.
- Lambrecht, M. & Rürup, M. (2012). Bildungsforschung im Rahmen einer *evidenz-based policy*: Das Beispiel der Schulinspektion. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung*. Wiesbaden: VS, S. 57–77.
- LISA Landesinstitut für Schule und Ausbildung Mecklenburg-Vorpommern (Hrsg.) (2006). *Gute Schule: Externe Evaluation von Schulen in Mecklenburg-Vorpommern*. Schwerin. URL: http://www.bildung-mv.de/export/sites/lisa/de/schule/entwicklung/selbst_schule/dok_s_schule/GS_Ext_Evaluation.pdf [04.10.2011].
- Luginbuhl, R., Webbink, D. & Wolf, I. de (2009). Do Inspections Improve Primary School Performance? *Educational Evaluation and Policy Analysis*, 31 (3), 221–237.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz UTB.
- McCrone, T., Coghlan, M., Wade, P. & Rudd, P. (2007). *Evaluation of the impact of Section 5 inspection: Research Report*. National Foundation for Educational Research.
- Niedersächsische Schulinspektion (Oktober 2008). *Periodischer Bericht*. Bad Iburg. <http://www.nibis.de/nibis3/uploads/2nlqa2/files/materialien/jahresbericht2008.pdf> [11.11.2014].
- Oevermann, U. (2000). Die Methode der Fallrekonstruktion in der Grundlagenforschung sowie der klinischen und pädagogischen Praxis. In K. Kraimer (Hrsg.), *Die Fallrekonstruktion: Sinnverstehen in der sozialwissenschaftlichen Forschung*. Frankfurt a. M.: Suhrkamp, S. 58–157.
- Pietsch, M. (2011). *Nutzung und Nützlichkeit der Schulinspektion Hamburg. Ergebnisse der Hamburger Schulleitungsbefragung*. Hamburg: Institut für Bildungsmonitoring. Online: <http://www.schulinspektion.hamburg.de/index.php/file/download/1744> [29.01.2013].
- Preuß, B., Brüsemeister, T. & Wissinger, J. (2012). Einführung der Schulinspektion und die Rolle der Schulleitung aus governanceanalytischer Perspektive. *Empirische Pädagogik*, 26 (1), S. 101–120.

- Preuß, B., Wissinger, J. & Brüsemeister, T. (2014). Einführung der Schulinspektion: Struktur und Wandel regionaler Governance im Schulsystem. In H. J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungssystem: Analysen zur Mehrebenenperspektive, Steuerung und Koordination*. Wiesbaden: Springer VS.
- Reezigt, G. (2001). *A framework for effective school improvement: Final report of the ESI project: Work Package 11 of the ESI project SOE2-CT97-2027, capacity for change and adaptation of schools: the case of effective school improvement*. Groningen: GION, Institute for Educational Research, University of Groningen.
- Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review*, 23, S. 143–151.
- Rürup, M. (2008). Typen der Schulinspektion in den deutschen Bundesländern. *Die Deutsche Schule*, 100 (4), S. 467–477.
- Sabatier, P. A. (1993). Advocacy-Koalitionen, Policy-Wandel und Policy-Lernen: Eine Alternative zur Phasenheuristik. In A. Héritier (Hrsg.), *Policy-Analyse. Kritik und Neuorientierung. Politische Vierteljahresschrift 34, Sonderheft 24*, Opladen: S. 116–148.
- Visscher, A. J. & Coe, R. (2002). *School Improvement through performance Feedback*. Swets & Zeitlinger BV: Lisse.
- Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly*, 21, S. 1–19.
- Wissinger, J. (2013). Schulleitungshandeln und Förderung der Professionalität unter Lehrpersonen: Eine Analyse institutioneller und struktureller Entwicklungsbedingungen. In M. Keller-Schneider, S. Albisser & J. Wissinger (Hrsg.), *Professionalität und Kooperation in Schulen*. Bad Heilbrunn: Klinkhardt, S. 185–208.
- Wissinger, J. (2014). Schulleitung und Schulleitungshandeln. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf*. Münster u. a.: Waxmann.

3. Qualitätsentwicklung von Schulen in sozialräumlich deprivierter Lage

*Isabell van Ackeren, Kathrin Racherbäumer,
Christina Funke, Marten Clausen*

Herausforderungen für und Bewältigungs- strategien von Lehrkräften an Schulen in sozialräumlich deprivierter Lage

Zusammenfassung

Schulen in sozial deprivierten Kontexten sind im deutschen Sprachraum bislang deutlich weniger beforscht, als dies international der Fall ist. Dabei kann davon ausgegangen werden, dass die Arbeit in diesen Schulen im Hinblick auf die Forderung und Förderung benachteiligter Schüler und Schülerinnen besonders herausfordernd und gleichzeitig bedeutsam ist, um Übergänge in Ausbildung und Beruf zu befördern. Die spezifischen Herausforderungen werden im Beitrag vor dem Hintergrund des Forschungsstandes skizziert. Zu den besonders relevanten Aspekten der datengestützten Schulentwicklung, des darauf bezogenen Schulleitungshandelns und der Unterrichtsgestaltung werden Befunde aus einem BMBF-geförderten Forschungsprojekt im Fallstudiendesign auf der Grundlage quantitativer und qualitativer Daten präsentiert. Dabei werden Gemeinsamkeiten und Unterschiede zwischen sogenannten erwartungswidrig guten und erwartungskonform schwachen Schulen beschrieben.

International und zunehmend auch national gibt es ein verstärktes Interesse an Schulen in schwierigen sozialen Kontexten, die mit verschiedenen Restriktionen und fehlenden Ressourcen ihrer Schülerinnen und Schüler bzw. des weiteren sozialen Umfelds für schulisches Lernen umgehen müssen. Auf der Ebene der Einzelschule wirken sich diese Kontextbedingungen mit Blick auf die Schülerleistungen insofern ungünstig aus, als dass diese Schulen in landesweiten Vergleichsarbeiten zumeist unterdurchschnittliche Resultate erzielen (auf der Grundlage von VERA-Daten in NRW: Bonsen et al. 2010, S. 27 ff).

Gleichwohl zeigen Ergebnisse der Schuleffektivitätsforschung, dass die Einzelschule Einfluss auf die (Leistungs-)Entwicklung von Schülerinnen und Schülern nehmen kann (Rutter 1980; Mortimore 1988; Scheerens 2004; Gröhlich 2011). *Erwartungswidrig gute Schulen* zeichnen sich – im Vergleich zu Schulen mit ähnlich schwierigen Rahmenbedingungen – durch überdurchschnittliche mittlere Prüfungsergebnisse am Ende der Pflichtschulzeit aus. Hier stellt sich die Frage, wie es diesen Schulen gelingt, trotz ungünstiger Rahmenbedingungen solche guten Resultate zu erzielen.

International gibt es bereits verschiedene Studien, die für erwartungswidrig gute Schulen (bzw. international u. a. „improving schools in challenging circumstances“,

vgl. Harris et al. 2006; zusammenfassend Muijs et al. 2004; Sammons et al. 2011) unterschiedliche Qualitätsfaktoren auf der schulischen Prozessebene identifizieren konnten. Hierbei handelt es sich u. a. um Merkmale des Schulleitungshandelns, der Unterrichtsqualität sowie der Gestaltung interner und externer Kooperationsbeziehungen zur Erweiterung des Sozialkapitals der Schule. Vor dem Hintergrund divergierender bildungspolitischer und schulstruktureller Rahmenbedingungen (z. B. in England etwa mit Blick auf Optionen der Entlassung bzw. den Austausch von Schulleitungen, die Möglichkeit der Schulschließung, das Zusammenwirken von Schulrankings und elterlicher Schulwahlfreiheit) sind diese Ergebnisse jedoch nicht ohne Weiteres auf den deutschen Kontext übertragbar.

Der Fokus dieses Beitrages liegt auf der Präsentation von Ergebnissen des Projekts „Strategien der Qualitätsentwicklung von Schulen in schwieriger Lage“ zu den durch die Akteure beschriebenen Herausforderungen an Schulen in sozialräumlich deprivierter Lage einerseits und zu adaptiven Bewältigungsstrategien der Lehrkräfte und Schulleitung andererseits. Den theoretischen Rahmen zur Analyse der Rahmenbedingungen und der Reaktionen der Akteure bildet die Kontingenztheorie, die im ersten Teil des Beitrages erläutert wird (Kapitel 1). Darauf folgt eine knappe Darstellung des (inter-)nationalen Forschungsstands (Kapitel 2), mit besonderem Fokus auf die Dimensionen des datengestützten Schulleitungshandelns und der Unterrichtsqualität. Im nächsten Schritt werden zentrale Fragestellungen des Projekts hergeleitet sowie das methodische Vorgehen erläutert (Kapitel 3). Auf die Darstellung der Ergebnisse (Kapitel 4) folgt eine vertiefende Analyse, Diskussion und theoretische Einordnung der Ergebnisse (Kapitel 5).

1. Schulqualität und sozialräumlicher Kontext

In der aktuellen Diskussion um neue Ansätze der Steuerung und der Erweiterung von Autonomie von Schulen im Kontext der inneren Schulreform, ist der Einbezug organisationstheoretischer Modelle von zentraler Bedeutung. Schulen sollen zunehmend die Aufgabe übernehmen, im reflexiven Umgang mit dem eigenen Umfeld die Qualität der schulischen und unterrichtlichen Prozesse kontinuierlich weiterzuentwickeln (vgl. Rolff 2013).

Als „erfolgreich“ gilt in diesem Zusammenhang eine Organisation, die ihre Ziele koordiniert verfolgt und Stärken und Schwächen der Akteure gezielt einsetzt bzw. angeht (vgl. Böttcher 2002). Eine zentrale Aufgabe der schulischen Organisation ist es, Strategien zur adaptiven Bewältigung herausfordernder Kontextbedingungen zu entwickeln.

Vor dem Hintergrund dieser allgemeinen organisationstheoretischen Annahmen bietet insbesondere die Kontingenztheorie (Kieser & Walgenbach 2010) Ansatzpunkte zur Analyse von Bildungseinrichtungen in schwieriger Lage, da sie die Frage fokussiert, inwiefern externe Kontextbedingungen die Strukturgestaltung intern beeinflussen (vgl. Creemers et al. 2000; Preisendörfer 2011; Reynolds & Teddlie 2000).

Die ursprünglich in der Organisationssoziologie entwickelte, auf den Bildungsbereich übertragene *Kontingenztheorie* geht – nach dem Verständnis von Creemers et al. (2000) – davon aus, dass die Schuleffektivität von spezifischen situationalen Faktoren innerhalb und außerhalb der Organisation abhängig ist. Insbesondere die Unterscheidung zwischen Dimensionen der internen versus der externen Situation erscheint geeignet, die spezifische Situation von Schulen in sozial deprivierter Lage zu beschreiben. Während auf der Ebene der internen Situation z. B. die Größe und das Alter der Organisation subsumiert werden, werden unter den Dimensionen der externen Situation gemeinhin die Umweltbedingungen verstanden (vgl. Preisendörfer 2011). Dies führt zu der These, dass Schulen in deprivierten Regionen durch ähnliche Organisationsmerkmale charakterisiert sind und sich in diesen von anderen Schulen unterscheiden und entsprechend differenziert agieren müssen (vgl. Reynolds 1996 und 1998). Hinsichtlich weniger effektiver Schulen in benachteiligten Regionen kann angenommen werden, dass die Aktivitäten der am Bildungsprozess beteiligten Akteure unzureichend zwischen den situativen Faktoren einerseits sowie den schulischen und unterrichtlichen Arbeitsprozessen andererseits vermitteln.

Dem *Kompensationsmodell* (in Anlehnung an Teddlie et al. 2000) liegt die Annahme zugrunde, dass Schulen in deprivierter Lage Kompensationsleistungen für fehlende außerschulische Unterstützungsleistungen ihrer Schülerschaft erbringen müssen, um verbesserte Lernerträge zu erzielen. Dazu müssen sie sich zunächst auf grundlegende Bedarfe konzentrieren, wie eine geordnete Lernumgebung oder eine hohe Erwartungshaltung, mit denen Schüler und Schülerinnen außerschulisch kaum konfrontiert werden, bevor in einem zweiten Schritt strukturelle, längerfristige Entwicklungsprozesse eingeleitet werden. Dies impliziert, dass Lehrkräfte an diesen Schulen zeitlich mehr arbeiten müssen, um entsprechende Resultate zu erzielen, da sie zunächst „kompensieren“ müssen (Whitty 2001; McDougall 2006). Demnach lässt sich das Kompensationsmodell in die Kontingenztheorie integrieren, da es auf der Prozessebene eine Handlungsstrategie beschreibt, mit den situationalen Faktoren umzugehen. Dabei ist anzumerken, dass der Grundgedanke des Kompensationsmodells, der einen mittelschichtorientierten Normalfall als Bezugsgröße impliziert, den es über die Kompensation herkunftsbedingter fehlender Ressourcen zu erreichen gilt, kritisch zu hinterfragen ist (vgl. hierzu auch Bremm 2015).

Die *Additivitätshypothese* geht davon aus, dass Schulen in benachteiligten Regionen ein höheres Risiko niedriger Lernerträge haben, und zwar auch nach Kontrolle des sozioökonomischen Hintergrunds (Reynolds & Teddlie 2000). Dahinter steht die Annahme einer abnehmenden Motivation der Lehrkräfte bzw. auf der Seite der Schüler und Schülerinnen die Annahme eines fehlenden Anregungsmilieus durch die Abwesenheit leistungsstarker Schüler und Schülerinnen, sodass negativ kumulative Effekte der Kontextmerkmale benachteiligter Schulen und der Reaktion auf die Lehr- und Lernvoraussetzungen durch die schulischen Akteure erwartet werden können (vgl. auch Baumert et al. 2006). Dementsprechend könnte man bei der Additivitätshypothese auch von einer nicht stattfindenden oder erfolglosen Adaptation situationaler Rahmenbedingungen auf der Mikroebene sprechen, da die bestehenden Kontextbedingungen nicht konstruktiv aufgegriffen werden.

Die Einordnung sowohl des Kompensationsmodells als auch der Additivitätshypothese geht über den ursprünglichen Anspruch der Kontingenztheorie als Strukturtheorie, die das Handeln der Akteure auf der Mikroebene nicht fokussiert, hinaus (vgl. Preisendörfer 2011, S. 94) und greift damit gleichermaßen einen wesentlichen Kritikpunkt, der gerade die fehlende Berücksichtigung der Mikroebene moniert, auf.

Neben den benannten Theorien und Modellen, die im Kontext erwartungswidrig guter Schulen bislang zur theoretischen Rahmung rezipiert wurden, fokussieren Theorien des organisationalen Lernens ebenfalls die schulische Mikroebene. Ansätze wie Organisationsentwicklung (Lewin 1963), Change Management (Bea und Göbel 2010) und Lernende Organisationen (Dalin 1999; Senge & Klostermann 2008) bieten wesentliche Ansatzpunkte einer vertiefenden Betrachtung von Schulentwicklungsprozessen, indem zum Beispiel Prozesse der Remodellierung fokussiert werden (vgl. zusammenfassend auch Dederling 2012; Feldhoff 2011; Gourmelon et al. 2014). Im Rahmen der vorliegenden Studie liegt der Forschungsfokus jedoch nicht primär auf *Schulentwicklungsprozessen*, vielmehr geht es darum, zunächst den Ist-Stand schulischer Praxis an erwartungswidrig guten bzw. erwartungskonform abschneidenden Schulen zu ermitteln. Dabei wird angenommen, dass eine Passung zwischen *dem Handeln* der Lehrkräfte bzw. der Schulleitung und den externen Kontextbedingungen, womit an dieser Stelle sowohl die *sozialräumliche* Einbettung der Schule als auch die *sozialstrukturelle* Zusammensetzung der Schülerschaft gemeint sind, zum Lernerfolg beitragen kann.

Vor diesem Hintergrund bieten handlungstheoretische Ansätze des organisationalen Lernens, wie etwa von March und Olsen (1975) oder von Argyris und Schön (2008), Ansatzpunkte, die die Interaktion zwischen dem Handeln der Organisationsmitglieder (hier Lehrkräfte und Schulleitungen) und der Reaktion der Umwelt in den Blick nehmen. So sind kollektive Deutungen der Organisationsrealität Produkt und Ausgangspunkt von Organisationalem Lernen (vgl. Feldhoff 2011, S. 58). Organisationales Lernen findet dann statt, wenn sich die kollektive Wissensbasis („theory of action“), die nach Argyris und Schön (2008) jeweils spezifisch für jede Organisation ist, ändert. Änderungsprozesse werden nicht zuletzt durch Umweltbeobachtungen und einen damit einhergehenden Informationsabgleich von Umwelt und derzeit geltenden Standards initiiert (vgl. Vahs 2012, S. 463).

Demnach können auch an Schulen in ähnlichen sozialräumlichen Kontexten divergierende kollektive Deutungen der situationalen Kontexte entstehen, die wiederum unterschiedliche Handlungsstrategien evozieren können. Von einer gelungenen Passung von Organisation und Umwelt kann jedoch nur dann gesprochen werden, wenn die entwickelte Organisation(skultur) auch zu entsprechend messbaren Erfolgen führt. Im schulischen Kontext können hierzu Leistungsdaten von Schülern und Schülerinnen, Übergangsquoten, aber auch Daten zum schulischen Wohlbefinden von Schülern und Schülerinnen sowie Lehrern und Lehrerinnen herangezogen werden.

2. Erfolgreiche Schulen in sozialräumlich deprivierter Lage

In der Literatur sind Charakteristika von Schulen dokumentiert, die trotz schwieriger Kontextfaktoren erwartungswidrig gute bzw. im Zeitverlauf verbesserte Lernergebnisse erzielen konnten (zusammenfassend z. B. Muijs et al. 2004; van Ackeren 2008). Die in vielen Beiträgen beschriebenen Kennzeichen von *effective schools* bzw. *improving schools* konzentrieren sich im Kern insbesondere auf das spezifische Handeln der Schulleitung, auf die Qualität der internen und externen Beziehungen und Netzwerke (einschließlich einer lernförderlichen Schulkultur und eines guten Schulklimas) sowie auf das unterrichtliche Lehren und Lernen (vgl. dazu Funke & Clausen 2015) einschließlich der Professionsentwicklung in diesem Bereich. Dabei werden der Leitung und dem Management von Schule zumeist besondere Bedeutung beigemessen, da diese auf die benannten Qualitätsdimensionen grundlegend Einfluss nehmen, Veränderungen im Sinne der Qualitätsentwicklung einleiten und Vernetzung herstellen können. Gleichwohl versteht sich die Auflistung solcher Merkmale nicht als Prioritätenliste, zumal in vielen Fällen Angaben zu Effektstärken fehlen. Zudem erfordert die Entwicklung leistungsschwacher Schulen komplexe Handlungsstrategien der Schulleitungen, die den je spezifischen, multikausalen Problemkonstellationen gerecht werden (Schwier 2005). Die als bedeutsam identifizierten Felder – datengestütztes Leitungshandeln und Unterrichtsqualität – werden nachfolgend dargestellt.

2.1 Datengestütztes Schulleitungshandeln

Vor allem vorliegende Befunde aus internationalen Studien zu erfolgreichen Schulen in schwieriger sozialer Lage deuten darauf hin, dass der Schulleitung erhöhte Aufmerksamkeit in der Schulentwicklung beizumessen ist (Reynolds et al. 2001; Harris et al. 2006; Sammons et al. 2011; Klein 2015). Indirekte Wirkungen auf die Effektivität von Schule konnten zumindest für den anglophonen Raum in verschiedenen Metastudien belegt werden (vgl. z. B. Hendriks & Steen 2012). Als Hauptaufgaben von Schulleitungen benennt Wissinger (2011, S. 104) „die Verbesserung der Schülerleistungen als Zielpunkt schulischen Handelns und als Antwort auf spezifische soziale und ethische Kontexte des Lehrens und Lernens“, womit ein adaptives, kontextsensibles Handeln der Schulleitungen markiert wird. Dass das konkrete Schulleitungshandeln sowohl situations- als auch kontextspezifisch geprägt ist, bestätigen Leithwood und Riehl (2003).

Keys et al. (2003) kommen auf der Grundlage einer systematischen Sichtung des Forschungsstands zu dem Ergebnis, dass sich ein kooperativer/partizipativer Führungsstil, der durch eine hohe Erwartungshaltung an die Lehrkräfte sowie die Schüler und Schülerinnen gekennzeichnet ist, als charakteristisch für *improving schools in challenging circumstances* erweist. Die Arbeitsweise der Schulleitungen sollte sich ferner durch ein hohes Maß an Transparenz auszeichnen, das durch demokratische Entscheidungsprozesse und Partizipation aller schulischen Akteursgruppen flankiert wird (McDougall et al. 2006; Flintham 2006). In diesem Kontext wird mit Be-

zug auf den sogenannten distributiven Führungsstil das Delegieren von Aufgaben an spezielle Arbeitsgruppen positiv hervorgehoben, um eine breite Teilhabe an Schulentwicklungsprozessen zu befördern und Schulentwicklungskapazität in der Breite aufzubauen. Verschiedene Autoren (Reynolds 2001; Hopkins 2001; Thomas et al. 1998; Sammons et al. 2011) beschreiben einen Zusammenhang zwischen dem Datenreichtum und dem Erfolg von Schulen, die unter schwierigen Bedingungen arbeiten. Dabei steht das datengestützte Handeln für eine systematische, rationale Strategie unter Berücksichtigung der spezifischen organisationalen Ausgangsbedingungen. Zur schulischen Qualitätssicherung und Qualitätsentwicklung greifen erfolgreiche Schulleiter und Schulleiterinnen systematisch auf schulinterne und -externe Daten zurück (vgl. Hopkins 2001; Reynolds 2001; Leithwood 2010; Racherbäumer et al. 2013b); dies ist Ausdruck eines systematischen Führungshandelns, das vom Interesse an der prozess- und ergebnisbezogenen Situation und Arbeitsleistung der Schule getragen ist. Diese Vorgehensweise zeigt deutliche Schnittmengen zum Modell „Leadership for Learning“. Hallinger et al. (2011) betonen in ihrem Modell „Leadership for Learning“ die Wechselwirkung zwischen der Schulleitung einerseits und verschiedenen Aspekten der Schulqualität, die schließlich zum schulischen Erfolg der Schüler/innen beitragen, andererseits. Zudem berücksichtigen die Autoren die kontextuelle Einbettung einer Schule mit Blick auf das Schulleitungshandeln (Hallinger et al. 2011). Bezug nehmend auf die dargestellten Befunde kommt der Schulleitung die Rolle des sogenannten *Change Agent* zu, dem es durch eine visionäre Grundhaltung und einen professionellen Führungsstil nicht nur gelingt, die Qualität von Schul- und Unterrichtsprozessen sowie ihren Ergebnissen zu verbessern, sondern auch an der Außenwirkung der Schule zu arbeiten und so deren Image nachhaltig aufzuwerten (vgl. Leithwood et al. 2010; Hallinger & Heck 1998; Harris & Chapman 2001; zusammenfassend Klein 2015).

2.2 Unterricht

Während die Schulleitung die organisationalen Rahmenbedingungen wesentlich gestaltet, um die Grundlage für eine gute Schulqualität insgesamt und damit auch eine entsprechende Unterrichtsqualität zu legen, müssen die einzelnen Lehrkräfte auf der Mikroebene des Unterrichts der Heterogenität ihrer Schüler und Schülerinnen produktiv begegnen. Funke und Clausen (2015) haben dargestellt, dass verschiedene Studien zu erfolgreichen *schools in challenging circumstances* die unterrichtlichen Qualitätskriterien bestätigen, die für guten Unterricht allgemein formuliert wurden (Helmke 2012; insgesamt Muijs et al. 2004; zur Klassenführung Calkins et al. 2007). Nachfolgend werden die Qualitätskriterien guten Unterrichts nach Helmke (2012) hinsichtlich ihrer Bedeutung für Schulen in sozial deprivierter Lage beleuchtet.

Wie auch Helmke (2012) als allgemeines Qualitätskriterium guten Unterrichts auf eine gute *Klassenführung* und damit verbunden auf eine hohe Nutzung der Lernzeit verweist, so zeigen ebenfalls Befunde internationaler Untersuchungen zum Unterricht an erfolgreichen Schulen in segregierter Lage, dass dieses Merkmal besonders wichtig ist, werden doch durch ein geordnetes Lernumfeld und disziplinierte Verfah-

rensrountinen erst die Bedingungen geschaffen, unter denen Lernen erfolgen kann (Muijs et al. 2004; Calkins et al. 2007; Keys et al. 2003). Insgesamt zeigt sich jedoch, dass mit Blick auf ein geordnetes Lernumfeld neben dem Fokus auf die Einzelklasse auch ein ganzheitlicher Ansatz auf Schulebene notwendig ist (Muijs 2010). Ein Faktor, der sich in Bezug auf Schülerleistungen vor allem für Schüler und Schülerinnen mit ungünstiger sozialer Herkunft als wirksam gezeigt hat, ist die lehrerübergreifende Kohärenz der gewählten Ansätze (Kyriakides 2005) zur Verhaltenssteuerung (behaviour management) (Kerr & West 2010). Insgesamt sollen durch ein geordnetes Lernumfeld Zeitverluste minimiert werden. Auch eine Studie über sieben erfolgreiche Schulen in segregierter Lage in den USA hat die Zeit zur Aufgabebearbeitung als Schlüssel zum Erreichen von Fortschritt identifiziert (Carter 1999; nach Keys et al. 2003).

Mit Blick auf das Qualitätsmerkmal der Angebotsvielfalt scheint insgesamt eine (lehrerübergreifende) Beständigkeit in den Unterrichtsmethoden für segregierte Schülerschaften wichtig zu sein und mit verbesserten Ergebnissen im Zusammenhang zu stehen (Kyriakides 2005; Kerr & West 2010; Harris & Chapman 2001).

Hinsichtlich des Aspekts der *Klarheit und Strukturiertheit* scheint Einvernehmen darin zu bestehen, dass ein klar strukturierter Unterricht eine wichtige Rolle spielt, um Lernprozesse an Schulen in sozial deprivierter Lage zu befördern (Harris & Chapman 2001; Campbell et al. 2004; Ledoux & Overmaat 2001).

Lehrer und Lehrerinnen haben Erwartungen an die Fähigkeiten und Fertigkeiten ihrer Schüler und Schülerinnen, die deren Leistungen im Sinne des Pygmalion-Effekts beeinflussen. Dusek und Joseph (1983; nach Hattie 2009) haben festgestellt, dass das soziale Milieu, aus dem Schüler und Schülerinnen stammen, signifikant und positiv mit den Lehrererwartungen in Verbindung steht. Je niedriger das soziale Herkunftsmilieu, umso niedriger sind die Erwartungen der Lehrkräfte an die Leistungen der Schüler und Schülerinnen. Demgemäß berichten auch Harris und Chapman (2001), dass in Schulen in segregierten Quartieren oft niedrige Erwartungen daran bestehen, was die Schüler und Schülerinnen erreichen können. Besonders schwerwiegend erweist sich dieser Befund vor dem Hintergrund, dass insbesondere sozioökonomisch benachteiligte Schüler und Schülerinnen am besten lernen, wenn die Lehrkräfte hohe Erwartungen an sie stellen (Gamoran 1993; Gamoran et al. 1997; Jussim & Harber 2005; nach OECD 2011). Demgemäß konnten bei Lehrkräften in effektiven Schulen in segregierter Lage höhere Erwartungen an die Schülerleistung identifiziert werden als bei Mitarbeitern und Mitarbeiterinnen in weniger effektiven Schulen mit ähnlicher Schülerzusammensetzung (Vaughn 2013, S. 24).

Hinsichtlich der Merkmale *Kompetenzorientierung, Aktivierung* sowie die *Konsolidierung und Sicherung* legen internationale Studien dar, dass Schülerschaften an Schulen in sozioökonomisch benachteiligten Lagen von einem Curriculum profitieren, das für ihr tägliches Leben relevant und fächerübergreifend ist (Hopkins & Reynolds 2001; Rossi et al. 1997).

Eine *positive Schülerorientierung* benennen Helmke (2012) und Hattie (2009) als zentrales Merkmal guten Unterrichts. Dies bestätigen mit Blick auf Schulen in benachteiligter Lage auch Calkins et al. (2007). Sie berichten, dass die Fähigkeit von Lehrkräften, mit sozioökonomisch benachteiligten Kindern eine Beziehung aufzubauen, von einigen Forschern als Schlüsselfaktor erfolgreicher Schulen identifiziert

wurde. Auch in PISA 2009 wird mit Bezug auf sozioökonomisch schlechter gestellte Schülerinnen und Schüler angemerkt, dass sie weniger Probleme mit Disziplin haben und mehr lernen, wenn die Beziehung zu ihren Lehrkräften gut ist (Crosnoe et al. 2004) und sie der Auffassung sind, dass den Lehrkräften viel an ihrem schulischen Erfolg liegt

Hinsichtlich des Merkmals des *Umgangs mit Heterogenität* gelingt es Schulen, die trotz ihrer Lage in einem sogenannten sozialen Brennpunkt erfolgreich arbeiten, den Blick auf die individuellen Potenziale der Schüler und Schülerinnen aufrechtzuerhalten und sich an deren Lern- und Förderbedarfen zu orientieren (Herrmann 2010).

Entsprechend dem Kompensationsmodell hat sich für Schulen in benachteiligter Lage auch die Kompensation von Benachteiligung als erfolgreich erwiesen (McDougal et al. 2006).

3. Fragestellung und Forschungsdesign des Projekts „Strategien der Qualitätsentwicklung von Schulen in schwieriger Lage“

Das Projekt „Strategien der Qualitätsentwicklung von Schulen in schwieriger Lage“ verfolgte das Ziel, Qualitätsentwicklungsstrategien durch ein kontrastives Fallstudien-Design von effektiven versus weniger effektiven Schulen in sozialräumlich deprivierter Lage herauszuarbeiten. Dazu stand zunächst die Analyse der internen und externen Situation der Schule im Vordergrund, um daran anknüpfend zu eruieren, welche Strategien auf Organisations- und Unterrichtsebene praktiziert werden, um den umschriebenen Herausforderungen zu begegnen. Schließlich sollte der Vergleich der Schulen unter Berücksichtigung ihrer erreichten Schüler- und Schülerinnenleistung die Möglichkeit bieten, genutzte Strategien mit Blick auf ihre Wirkung einzuschätzen.

Zusammenfassend ergeben sich daraus die folgenden Forschungsfragen, die vertiefend auf Gruppenunterschiede zwischen erwartungsgemäß bzw. erwartungswidrig gut abschneidenden Schulen hin untersucht wurden:

- Wie werden Schulen in schwieriger Lage durch die Lehrkräfte der entsprechenden Schule charakterisiert?
- Welche Strategien verfolgen Schulleitungen und Lehrkräfte in der Bewältigung der beschriebenen Herausforderungen?
- Inwiefern gibt es Qualitätsentwicklungsstrategien, die insbesondere von erwartungswidrig guten Schulen verfolgt werden?

Dem Projekt liegt ein Fallstudienansatz zugrunde, durch den im Rahmen einer „umfassenden, systematischen Darstellung der einzelschulischen Wirklichkeit“ (Horstkemper & Tillmann 2008, S. 298) die einzelschulische Situation differenzierter im Sinne eines holistischen Verständnisses beschrieben werden kann. Dabei wurde ausgehend von einem theoretischen Sampling, mit dem neben erwartungskonform schwachen auch erwartungswidrig gute Schulen ausgewählt wurden, ein fallkontrastiver Zugang gewählt. Auf der Grundlage vorliegender Daten der Vergleichsarbeiten der Jahrgangsstufe 8 (VERA 8, in NRW auch: Lernstand 8) wurden in der Metro-

polregion Rhein-Ruhr acht Schulen identifiziert und für die Teilnahme gewonnen, die dem ungünstigen Standorttyp der Stufe 5 angehören (vgl. Isaac 2011).

Auf der Grundlage der Lernstand-8-Daten wurden den Schulen in Abhängigkeit von ihrem Standort sowie ihrer Schulform Erwartungswerte hinsichtlich der Leistungen der Schülerinnen und Schüler in den Lernstanderhebungen zugeschrieben. Hierdurch ist es möglich, die Ergebnisse der einzelnen Schule im Vergleich der Schulen in NRW insgesamt sowie mit Schulen bzw. Schulformen ähnlicher Standortvoraussetzungen zu vergleichen. Davon ausgehend wurden im Sample auf der Grundlage schulformspezifischer Regressionsanalysen mit dem kontinuierlichen Sozialindex und dem Migrantenanteil der Schulen als Prädiktoren (Jahrgänge 2009 bis 2011) zwei Gruppen von Schulen in sozioökonomisch benachteiligter Lage gebildet:

- Vier Schulen, die vom Erwartungswert nicht abweichen, bilden die Gruppe der erwartungsgemäßen (leistungsschwachen) Schulen (eine Hauptschule, zwei Gesamtschulen, ein Gymnasium).
- Vier Schulen, die von diesem Wert signifikant positiv abweichen, bilden die Gruppe der erwartungswidrig guten Schulen (zwei Hauptschulen, eine Gesamtschule, ein Gymnasium).

Ein solches Sample ermöglicht einen kontrastierenden Vergleich von Merkmalen und Prozessen an Schulen, die allesamt mit ungünstigen Bedingungen umgehen müssen, sich jedoch hinsichtlich ihrer Ergebnisse – dem Erfolg der Schüler und Schülerinnen bei den Lernstanderhebungen – im Durchschnitt unterscheiden. Die Datenerhebung erfolgte multimethodisch unter Einbezug mehrerer am Bildungsprozess beteiligter Akteure, um möglichst viele Perspektiven auf die einzelschulische Wirklichkeit zu erhalten.

Neben Leitfadeninterviews mit verschiedenen Akteuren (Lehrpersonen, Schulleitungen sowie Schüler und Schülerinnen), Dokumentenanalysen (z. B. von Schulprogrammen) und Unterrichtsbeobachtungen wurde eine schriftliche Schülerbefragung durch eine Vollerhebung der Stufen sechs, acht und zehn durchgeführt (N = 2232). Zudem gab es eine Fragebogenerhebung der Lehrkräfte. Die Lehrkräfte wurden u. a. zu Strategien schulischer und unterrichtlicher Prozesse sowie zum Schulleitungshandeln befragt. An der Befragung beteiligten sich insgesamt 198 Lehrkräfte, wobei 78 Lehrkräfte an erwartungswidrig guten Schulen unterrichteten. Die ungleiche Verteilung der Gesamtanzahl ist den unterschiedlich großen Kollegien der Schulen in den zwei Vergleichsgruppen geschuldet. Der Rücklauf an erwartungswidrig guten Schulen betrug im Durchschnitt insgesamt 43 %, an erwartungskonform abschneidenden Schulen insgesamt 33,6 % (zur detaillierten Stichprobenbeschreibung vgl. Funke & Clausen 2015).

Die Datenauswertung der Leitfadeninterviews erfolgte inhaltsanalytisch deduktiv durch ein theorie- und empiriegestütztes Codesystem sowie induktiv, indem Codes für „neue“ Strategien generiert wurden (vgl. Racherbäumer et al. 2013). Zur Datenauswertung der quantitativen Daten der Lehrer- und Lehrerinnen-Fragebogenerhebung wurden durch eine theoriegeleitete Faktorenanalyse Skalen gebildet, die sich insgesamt als reliabel erwiesen, und deskriptive Auswertungen vorgenommen (vgl. Funke & Clausen 2015). Mit Blick auf die hier betrachteten Gruppen der erwartungs-

konform abschneidenden bzw. der erwartungswidrig gut abschneidenden Schulen wurden die quantitativen und qualitativen Daten trianguliert und kontrastiert.

4. Ergebnisse

Im Folgenden werden quantitative Ergebnisse zu allen acht betrachteten Schulen vorgestellt und erläutert. Die vertiefenden qualitativen Daten beschränken sich auf die fallkontrastive Darstellung der Hauptschulen einerseits und den Gesamtschulen andererseits. Diese Auswahl ist dadurch begründet, dass sich die Hauptschulen mit Blick auf ihre situationalen Kontextbedingungen sowie ihrer Schulgröße, Anzahl der Lehrkräfte etc. stark ähneln, nicht jedoch mit Blick auf ihre Leistungen in VERA 8. Ferner ist zu betonen, dass alle drei Hauptschulen von einer möglichen Schulschließung „bedroht“ sind.

Die drei Gesamtschulen ähneln sich ebenfalls mit Blick auf ihre organisationalen Rahmenbedingungen (Schüler- und Schülerinnenanzahl, Lehrkräfte etc.), nicht jedoch mit Blick auf die Leistungen in VERA 8.

Die vertiefende Betrachtung der organisationalen Rahmenbedingungen der Gymnasien zeigte hingegen deutliche Unterschiede mit Blick auf die Größe der beiden Schulen, die Wettbewerbssituation zu anderen Schulen (die erwartungswidrig gute Schule befindet sich in einer hohen Wettbewerbssituation und sieht sich mit stetig sinkenden Schüler- und Schülerinnenanzahlen konfrontiert) sowie einen zum Forschungsbeginn erfolgten Wechsel der Schulleitung an der erwartungskonform abschneidenden Schule, was insbesondere die Betrachtung und Interpretation des aktuellen Schulleitungshandeln auf den Erfolg der Schule erschwert.

4.1 Analyse der internen und externen situativen Kontextdimensionen

Alle Fallschulen gehören dem ungünstigen Standorttyp der Stufe 5 an, womit die nachfolgend beschriebenen externen situativen Kontextbedingungen impliziert sind:

- über 40 % der Schüler und Schülerinnen haben (unabhängig von ihrer Staatsangehörigkeit) einen Migrationshintergrund;
- über 25 % der Schüler und Schülerinnen erhalten Sozialgeld nach SGB II oder kommen aus Familien, die den geregelten Eigenanteil im Rahmen der Lernmittelfreiheit nicht aufbringen können und damit zur Unterstützung Sozialhilfe nach SGB XII erhalten,
- die elterliche Wohnung liegt in einem Wohngebiet, dessen Einwohner/innen ein sehr geringes Einkommen aufweisen;
- die Schüler und Schülerinnen kommen aus einem Wohngebiet, in dem der Anteil der Empfänger/innen von SGB-II-Leistungen sehr hoch ist und das einen sehr hohen Ausländeranteil und Arbeitslosenanteil aufweist (vgl. Qualitäts- und UnterstützungsAgentur – Landesinstitut für Schule NRW 2014).

Die Dimensionen der internen Situation der Schulen variierten mit Blick auf die betrachtete Schulform, die Anzahl der Schüler und Schülerinnen und Lehrkräfte und die Schulgeschichte insbesondere bei den betrachteten Gymnasien sehr stark.

Darüber hinaus wurden zur Erfassung der internen Situation die Akteursperspektiven auf schulische Prozesse der Lehrer und Lehrerinnen eingeholt. Hierzu haben die Lehrkräfte im Rahmen eines offenen Antwortformats schriftlich angegeben und zum Teil erläutert, mit welchen größten Herausforderungen ihre Schule aus ihrer Sicht konfrontiert ist. Die Aspekte der wahrgenommenen mangelnden Unterstützung durch die Eltern, Sprachprobleme und Verhaltensauffälligkeiten der Schüler und Schülerinnen wurden insgesamt am häufigsten benannt.

Tabelle 1: Von Lehrkräften an Schulen in segregierter Lage beschriebene Herausforderungen bezogen auf den eigenen Schulstandort; N = 153; Angabe der Nennungen in Prozent der Lehrkräfte, die eine Angabe gemacht haben; Mehrfachnennungen möglich		
	Erwartungskonform abschneidende Schule	Erwartungswidrig gut abschneidende Schule
Mangelnde Unterstützung der Schüler und Schülerinnen durch die Eltern	44 %	50 %
Sprachprobleme von Kindern mit und ohne Migrationshintergrund	22 %	34 %
Problematisches Sozialverhalten der Schüler und Schülerinnen	65 %	30 %

Die Auswertung der offenen Antworten weist auf Unterschiede zwischen den Lehrkräften an erwartungskonform und erwartungswidrig gut abschneidenden Schulen hin.

Die wahrgenommene *mangelnde Unterstützung durch die Eltern* zeigt sich nach Angabe aller Lehrkräfte zum Beispiel in der defizitären Unterstützung der Kinder bei der grundlegenden Organisation ihres Alltages (z. B. morgendliches Wecken, Zubereiten von Mahlzeiten und gemeinsames Essen) über die Beschaffung von Schulmaterialien bis hin zur mangelnden emotionalen Zuwendung der Eltern zu ihren Kindern („Die Eltern interessieren sich nicht für ihre Kinder“). Sofern die Lehrkräfte auf die Ursachen dieser mangelnden Unterstützung näher eingehen, benennen sie u. a. Arbeitslosigkeit der Eltern, die zum Teil mit finanzieller Not bis hin zu Armut einhergeht. Diese belastenden familialen Lebensumstände lassen die schulischen Angelegenheiten und Probleme der Kinder aus Sicht der Lehrkräfte für die Eltern oftmals in den Hintergrund treten. Obgleich alle Schulen dem Standorttyp 5 angehören, sodass von ähnlichen soziokulturellen und sozioökonomischen Hintergründen der Schüler und Schülerinnen ausgegangen werden kann, wird die mangelnde Unterstützung durch die Eltern an erwartungswidrig gut abschneidenden Schulen häufiger als Herausforderung von den Lehrkräften formuliert bzw. erkannt.

Defizite in der Bildungssprache von Kindern mit und ohne Migrationshintergrund werden von den Lehrern und Lehrerinnen erwartungswidrig gut abschneidender Schulen häufiger als Herausforderung benannt als von Lehrern und Lehrerinnen erwartungskonform abschneidender Schulen. Zum Teil führen die Lehrkräfte aus, dass sowohl der Wortschatz als auch das grammatische Verständnis der Schüler und Schülerinnen in der Unterrichtssprache förderbedürftig sei und dies nicht nur auf die Gruppe der Schüler und Schülerinnen mit Migrationshintergrund beschränkt sei. In diesem Zusammenhang werden zum Teil auch sprachliche Verständigungsschwierigkeiten der Lehrer und Lehrerinnen mit den Eltern angeführt.

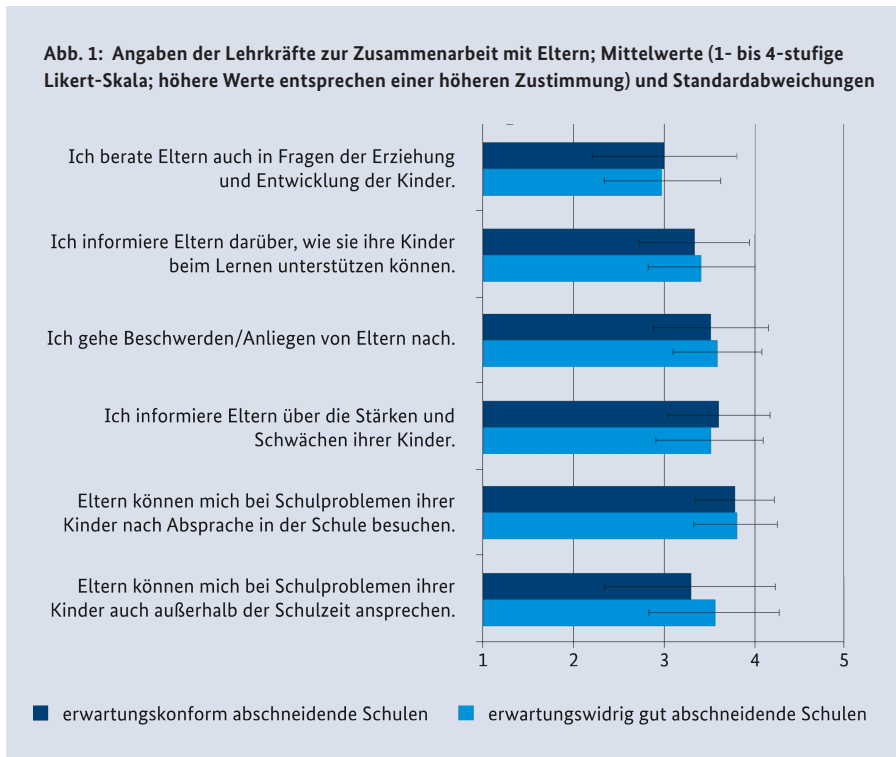
Unter dem Aspekt *Problematisches Sozialverhalten der Schüler und Schülerinnen* wurden unterschiedlichste Dimensionen von Verhaltensproblematiken von den Lehrkräften subsumiert. So beziehen sich einige Lehrkräfte nur auf den Unterricht und beschreiben Motivationsprobleme, Interessenlosigkeit oder Disziplinprobleme. Andere benennen Gewalt auf dem Schulhof, Schulabstinz oder mangelndes Selbstwertgefühl infolge von Schulversagen als zentrale Herausforderung. Hinsichtlich dieser Subkategorien zeigen sich tendenziell Übereinstimmungen zwischen den Schulformen, so trifft insbesondere der Aspekt des Schulversagens auf die beteiligten Gymnasien nicht zu, während sich die befragten Hauptschullehrkräfte mit dieser Problematik fast durchgängig konfrontiert sehen. Insgesamt benennen Lehrer und Lehrerinnen erwartungskonform schwach abschneidender Schulen das Verhalten der Schüler und Schülerinnen als häufigste Herausforderung – die Lehrkräfte erwartungswidrig gut abschneidender Schulen weniger häufig. Hier ergeben sich mit Blick auf die Ergebnisse der Schüler- und Schülerinnenbefragung zur Disziplin- und Störungsfreit im Unterricht deutliche Übereinstimmungen: Auch aus Sicht der Schüler und Schülerinnen kommen Unterrichtsstörungen und Disziplinprobleme im Unterricht an erwartungskonform abschneidenden Schulen häufiger vor (vgl. Racherbäumer et al. 2013).

4.2 Bewältigungsstrategien

Anknüpfend an die Kontingenztheorie ist zu erwarten, dass Lehrer und Lehrerinnen sowie Schulleitungen erwartungswidrig guter Schulen effektivere Strategien entwickelt haben, um die beschriebenen Herausforderungen konstruktiv zu bewältigen.

4.2.1 Strategien zum Umgang mit fehlender häuslicher Unterstützung

Im Rahmen der standardisierten Befragung der Lehrkräfte wurde zunächst abgefragt, wie sich die Zusammenarbeit mit Eltern gestaltet.



Die Angaben der Lehrkräfte machen deutlich, dass ein insgesamt hohes Bestreben der Zusammenarbeit aus Sicht der befragten Lehrkräfte vorhanden ist. Zwischen den Angaben der Lehrkräfte erwartungswidrig gut und erwartungskonform abschneidender Schulen bestehen hinsichtlich der hier abgefragten Items keine statistisch signifikanten Unterschiede. Die Analyse der Interviews der Lehrkräfte zeigt ebenfalls, dass Lehrkräfte der Zusammenarbeit mit Eltern grundsätzlich eine hohe Bedeutung zusprechen und hier zum Teil viel Zeit investieren. Die nachfolgenden Zitate stammen von Lehrkräften, die an Hauptschulen beschäftigt sind.

„(...) aber ich glaub', man muss sich ein bisschen von dem Gedanken verabschieden, dass man die Eltern jetzt komplett in den Lernprozess mit einbeziehen kann oder dass man komplett auf die Unterstützung des Elternhauses bauen kann. Und am Anfang habe ich mit meiner Kollegin (...) super viele Elterngespräche geführt – immer wieder und immer wieder, und da kommt aber letztendlich recht wenig raus dann. Von den Vereinbarungen wird dann halt wenig eingehalten, und deshalb hab' ich jetzt so für mich gedacht, dass ich's jetzt – also mich gar nicht mehr so auf die Eltern konzentriere, also es sei denn, da ist wirklich 'ne gute Zusammenarbeit möglich, was auch mit einigen super funktioniert. (...) ich geb' z. B. dann auch einigen, die nehmen ihre Klassenarbeiten gar nicht mehr nach Hause, weil bis ich die dann wieder sehe oder bis ich meine Unterschrift bekomme, dann ... vergeht einfach Zeit, das ist dann Zeit, die ich dann wieder da hinterher sein muss, die dann die eigentlich dem Unterricht zugutekommen könnte“ (Lehrerin einer erwartungskonform abschneidenden Hauptschule).

Auf der Grundlage des Interviews ist grundsätzlich zu sehen, dass die Lehrkräfte dieser Hauptschule bestrebt sind, die Eltern einzubeziehen, um zum Beispiel die häusliche Lernsituation der Kinder zu verbessern. Dies gelingt jedoch nur zum Teil, was einerseits zu Resignation und andererseits zu einer Abwendung von der häuslichen Lebenswelt des Schülers/der Schülerin führen kann, wie die nachfolgende Sequenz der Schulleiterin derselben Schule verdeutlicht:

„Also ein Großteil der Eltern sind Hartz-IV-Empfänger und haben einfach die Ressourcen nicht, die unsere Schüler bräuchten für ihre Schule. Das fängt mit materiellen Dingen an, dass, wie Sie eben sagten, Stifte fehlen, Geodreiecke fehlen, Hefte nicht da sind, teilweise keine Turnsachen in der richtigen Größe da sind, also diese ganzen materiellen Dinge, die fehlen. Dann die Haltung der Eltern, viele Eltern halten sich aus Schule raus, kommen dann noch nicht mal zu den Elternabenden und zu den Elternsprechtagen und sind dann auch nicht diejenigen, die uns unterstützen“ (Schulleiterin einer erwartungskonform abschneidenden Schule).

Bedeutsam an dieser Sequenz ist insbesondere der letzte Teilsatz, in dem die Schulleiterin indirekt die Erwartung formuliert, dass die Eltern sie (die Lehrer und Lehrerinnen) unterstützen sollen, um – so wird im weiteren Verlauf des Interviews deutlich – die Lernerfolge der Schüler und Schülerinnen zu befördern. Dies steht im Widerspruch zu der vorab geschilderten Erkenntnis, dass die Ressourcen sowohl auf materieller als auch auf personeller Ebene gar nicht vorhanden seien.

In der Zusammenschau der qualitativen Daten von erwartungskonform versus erwartungswidrig gut abschneidenden Haupt- und Gesamtschulen entsteht der Eindruck, dass Lehrkräfte erwartungswidrig gut abschneidender Schulen die mangelnde Unterstützung der Eltern zwar als deutliche Herausforderung akzentuieren, diesen Aspekt aber auf der schulischen Prozessebene nicht weiterverfolgen. So taucht diese Thematik in den Interviews bei Schulleitungen und Lehrkräften erwartungswidrig gut abschneidender Schulen zwar immer wieder auf, indem u. a. von erfolglosen Strategien berichtet wird, die mit Blick auf zeitlich und inhaltliche Ressourcen als belastend erlebt werden.

Gleichwohl kompensieren Lehrkräfte z. B. das Fehlen schulischer Materialien bzw. mangelnde Ernährung, indem sie selbst für Ersatzmaterialien sorgen bzw. externe Netzwerke initiieren, um zum Beispiel grundsätzlich für alle Schüler und Schülerinnen ein kostenfreies Essen zu besorgen, wie die Schulleitung der erwartungswidrig gut abschneidenden Hauptschule berichtet.

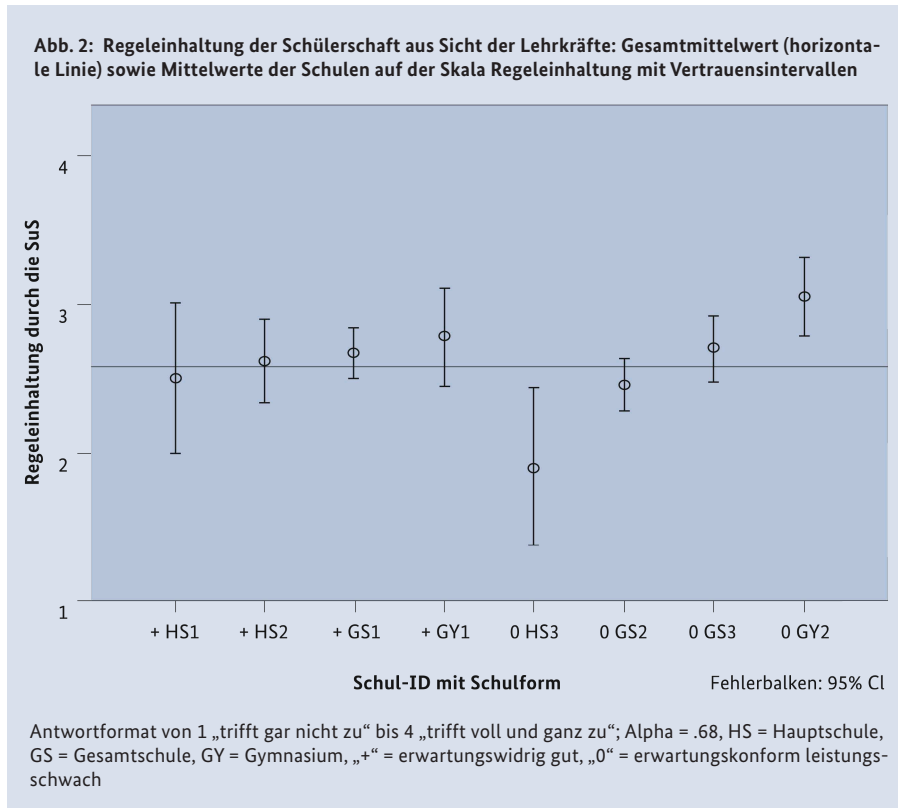
4.2.2 Sprachprobleme von Kindern mit und ohne Migrationshintergrund

Die Ergebnisse der Daten- und Dokumentenanalyse sowie der Interviews zeigen, dass dezidierte Sprachstandserfassungen an erwartungswidrig guten Schulen zu Beginn der 5. Klasse sowie Förderkonzepte zum Standard gehören. Auffallend ist ferner, dass die Leistungsentwicklung auch mit Blick auf die sprachlichen Fertigkeiten zum Teil systematisch längsschnittlich weiterverfolgt wird und somit einen wesentlichen Bestandteil einer datengestützten Unterrichtsentwicklung darstellt. An drei der er-

wartungskonform abschneidenden Schulen wird der Sprachstand zwar ebenfalls dokumentiert, jedoch werden Sprachförderkonzepte, die darüber hinaus evaluiert werden, nicht gleichermaßen benannt. Aus Sicht der Schüler und Schülerinnen wird auf Sprachprobleme im Unterricht beim Erklären durch die Lehrkraft selbst bzw. beim Lesen von Texten an erwartungswidrig guten Schulen tendenziell (jedoch nicht statistisch signifikant) mehr eingegangen (erwartungskonform abschneidende Schule: $M = 3,12$; $SD = 0,953$; erwartungswidrig gut abschneidende Schule: $M = 3,21$; $SD = 0,921$).

4.2.3 Strategien zum Umgang mit Verhaltensauffälligkeiten der Schülerinnen und Schüler

Die von den Lehrkräften beschriebenen Verhaltensauffälligkeiten beziehen sich neben Schulabstinenz und Gewaltbereitschaft insbesondere auf Disziplinprobleme im Unterricht. Hier zeigt die Auswertung der Lehrer- und Lehrerinnenbefragung zwar tendenziell, dass es erwartungskonform abschneidenden Schulen schlechter gelingt, Regeln im Unterricht umzusetzen, was zu einem insgesamt hohen Anteil an Störungen führt (vgl. Abbildung 2 sowie Funke & Clausen 2015). Darüber hinaus dokumentiert die Grafik aber auch große Unterschiede innerhalb der hier gebildeten



Gruppen der erwartungswidrig gut und erwartungsgemäß abschneidenden Schulen. So wird hier auch deutlich, dass es an den beiden Gymnasien am besten gelingt, die Einhaltung der Regeln umzusetzen. Deutliche Unterschiede zeigen sich zwischen den erwartungswidrig guten Hauptschulen sowie der erwartungskonform abschneidenden Hauptschule, die insgesamt auch die größten Disziplinprobleme aus Sicht der Schüler und Schülerinnen hat.

Betrachtet man nunmehr die Strategien der Lehrkräfte, um dem herausfordernden Sozialverhalten sowie den unterrichtlichen Disziplinproblemen zu begegnen, wird durch die befragten Lehrkräfte die Lehrer-Schüler-Beziehung fokussiert. Mit Blick auf die Prozessebene im schulischen Alltag geht es aus Sicht der befragten Lehrkräfte schließlich darum, über eine positive Beziehungsebene die *Lernbereitschaft* der Schüler und Schülerinnen zu unterstützen. Im Gegensatz zu Schülern und Schülerinnen in anderen Kontexten sind diese Schüler und Schülerinnen aus Sicht einiger Lehrkräfte durch ihre soziale Situation (z. B. wenig familiäre Unterstützung, Armut, Wohnsituation, häusliche Gewalt) zum Teil so stark belastet, dass es ihnen schwerfällt, sich auf das schulische Lernen einzulassen. Diese Bereitschaft zum Lernen bei Schülern und Schülerinnen an Schulen in deprivierter Lage herzustellen und aufrechtzuerhalten erscheint sowohl den befragten Lehrkräften erwartungswidrig gut als auch erwartungskonform abschneidender Schulen zentral, wie z. B. das nachfolgende Zitat einer Lehrkraft einer erwartungskonform guten Gesamtschule verdeutlicht:

„Die Arbeit an dieser Schule ist besonders herausfordernd insofern, dass man erst die Kinder für sich gewinnen muss, bevor man überhaupt unterrichten kann, das heißt, ein großer Teil unserer Arbeit ist auch der, dass wir eine gute Autorität aufbauen. Eine Autorität, die darauf fußt, dass wir die Kinder kennen, dass die Kinder uns akzeptieren. Und das muss einhergehen mit dem Lehren. Lehren alleine ohne Kontakt ist in dieser Lage mit diesen Kindern schwierig“ (Lehrkraft einer erwartungswidrig guten Gesamtschule).

Die Schüler und Schülerinnen sehen die Lehrkräfte nicht nur als Fachvermittler, sondern die Lehrer und Lehrerinnen fungieren in ihrer Rolle als Klassenlehrer und -lehrerin zum Teil als feste Bezugspersonen im Sinne eines „Klassenvaters“ bzw. einer „Klassenmutter“, was wiederum aus Sicht einer Schulleiterin einer erwartungswidrig guten Hauptschule wesentlich zur *Leistungsmotivation* der Schüler und Schülerinnen beiträgt:

„Unsere Kinder sind eigentlich nicht zum Lernen so motiviert, die möchten im Grunde gar nicht, und vieles, was sie hinterher tun, vor allen Dingen wenn sie älter werden, ich wage das mal zu sagen, machen die hinterher für uns. Die kommen irgendwann, so ich sage mal so in der Neun irgendwann, kommen sie dahinter, dass wir auf ihrer Seite sind (...) Und wenn wir dann wirklich mal bestimmte Dinge dann auch einfordern, dann würden die das von alleine vielleicht nicht machen, aber weil diese Anbindung, diese persönliche Anbindung da ist, machen sie es dann doch“ (Lehrkraft einer erwartungswidrig guten Hauptschule).

Gerade mit Blick auf die Bedeutung der Lehrer-Schüler-Beziehung wird die Relevanz unterschiedlicher organisationaler Rahmenbedingungen sichtbar, da hier divergierende Strategien erwartungswidrig guter bzw. erwartungskonform ab-

schneidender Haupt- und Gesamtschulen zum Tragen kommen. So berichten die Schulleitungen an drei erwartungswidrig guten Schulen, dass sie die Rahmenbedingungen an ihren Schulen modifiziert haben, u. a. um die Lehrer-Schüler-Beziehung zwischen Klassenlehrern und -lehrerinnen und Schülern und Schülerinnen intensivieren zu können.

An einer erwartungskonform abschneidenden Gesamtschule werden derzeit organisationale Rahmenbedingungen verändert, um die Lehrer-Schüler-Beziehung zu intensivieren. Auch hier wird die Rolle der Klassenlehrer und -lehrerinnen gestärkt, indem Lernzeiten und Arbeitsstunden im Rahmen des Ganztages zeitlich ausgeweitet werden. Die befragten Lehrkräfte dieser Schule begreifen ihre Schule aufgrund der schwierigen Kontextbedingungen auch als Familie, sodass Strukturen zur Intensivierung der Klassenlehrer-Schüler-Beziehung gestärkt werden. Bei den anderen erwartungskonform abschneidenden Schulen sind solche systematischen Veränderungen jedoch nicht bekannt. Dazu wurde an den hier betrachteten erwartungswidrig guten Hauptschulen das Klassenlehrerprinzip stark ausgebaut, sodass die Klassenlehrer und -lehrerinnen mit erhöhten Stundenanteilen auch fachfremd in ihren Klassen unterrichten. An der erwartungswidrig guten Gesamtschule wurde die Fachleistungsdifferenzierung weitestgehend aufgelöst:

„Vorteile hat es (die Auflösung der Fachleistungsdifferenzierung; Anm. Autorinnen) gebracht, dass meistens der Lehrer drinbleibt, der die ganze Zeit schon Mathematik unterrichtet hat, das ist oft der Klassenlehrer. Und das heißt, es ist eine Lernatmosphäre in der Klasse da, für E- und G-Kinder. Wenn es aufgeteilt wurde, haben oft die Lehrer gewechselt, einer musste ja, zumindest einer, G oder E musste ja von jemand Neuem übernommen werden (...) Das kann ein sehr guter Mathematiklehrer sein, und trotzdem kracht das Ding. Das heißt, da ist es dann schwierig, in der Sieben oder in der Acht, wenn man die Kinder nicht vorher kennt, eine Lernatmosphäre aufzubauen, weil der Kontakt nicht gewachsen ist“ (Lehrkraft einer erwartungswidrig guten Gesamtschule)

Tendenziell entsteht der Eindruck, dass Lehrkräfte und Schulleitungen erwartungswidrig guter Schulen durch die Intensivierung der Lehrer-Schüler-Beziehung den speziellen Bedürfnissen ihrer Schüler und Schülerinnen konstruktiver begegnen und damit Verhaltensproblemen möglicherweise vorbeugen.

5. Zusammenfassung der Ergebnisse und Diskussion

Ziel des Projekts „Strategien der Qualitätsentwicklung von Schulen in schwieriger Lage“ ist es, durch einen Vergleich erwartungswidrig guter und erwartungsgemäß abschneidender Schulen in sozial deprivierten Kontexten solche Strategien zu identifizieren, die sich positiv auf die Schüler- und Schülerinnenleistungen auswirken. Dabei wurde anknüpfend an die Kontingenztheorie einerseits und handlungstheoretische Annahmen im Kontext organisationalen Lernens andererseits vermutet, dass Schulen, die eine höchstmögliche Passung zwischen ihrem sozialräumlichen Kontext und ihrem schulischen Handeln erreichen, Strategien entwickelt haben, die sich positiv auf die Schüler- und Schülerinnenleistungen auswirken. Dazu wurden

acht Schulen, die dem Standorttyp 5 angehörten und sich damit in ähnlich herausfordernder Lage befinden sollten, ausgewählt. Vier der Schulen konnten unter Berücksichtigung ihrer Ergebnisse in Lernstand 8 als erwartungswidrig gut eingestuft werden, während die übrigen vier erwartungskonforme Leistungen erbrachten.

Die Analyse der Situation aus Sicht der Akteure zeigte deutliche Unterschiede mit Blick auf die wahrgenommenen Herausforderungen der Akteure zwischen den gebildeten Gruppen. Dieser Befund kann unterschiedlich interpretiert werden. Eine Möglichkeit wäre, dass es die Schulen trotz formal gleicher Zuordnung zum Standorttyp 5 in der tatsächlichen Schulrealität mit einer anderen Schüler- und Schülerinnenklientel zu tun haben. Im Rahmen der Schüler- und Schülerinnenbefragung wurden jedoch die soziokulturellen Hintergrundmerkmale der Schüler und Schülerinnen kontrolliert, sodass ein solcher substanzieller Unterschied auszuschließen ist. Demnach kann als zweite Möglichkeit angenommen werden, dass der Blick der Lehrkräfte auf ihre schulischen Kontextbedingungen möglicherweise ein anderer ist oder dass auf der organisationalen Ebene Strategien entwickelt wurden, die zum Beispiel Verhaltensprobleme von Schülern und Schülerinnen weniger befördern. Letztere Sichtweise ist durch die Handlungstheorie von Argyris & Schön (2008) erklärbar, da anzunehmen ist, dass an den einzelnen Schulen divergierende kollektive Deutungsmuster entwickelt wurden, die wiederum unterschiedliche Strukturen und Handlungen ermöglichen bzw. initiieren oder auch verhindern. Hierfür ist die Wahrnehmung der geringen elterlichen Unterstützung, die sowohl von Lehrern und Lehrerinnen erwartungskonform als auch erwartungswidrig gut abschneidender Schulen als wesentliche Herausforderung formuliert wird, ein Beispiel. So beschreibt die Lehrkraft einer erwartungswidrig guten Schule einen Lernprozess, der zu veränderten Handlungsrountinen in ihrem schulischen Alltag führt. Offen bleibt, ob ihr zunächst individueller Lernprozess (Veränderung der *theory of action*) auch zu veränderten Handlungsrountinen auf der schulischen Organisationsebene führt, sodass man mit Argyris und Schön von *double loop learning* sprechen kann. Kontrastierend dazu ist der Bericht der Schulleitung einer erwartungskonform abschneidenden Schule zum Umgang mit der fehlenden elterlichen Unterstützung ein Beispiel einer fehlenden *Adaption*, indem die Beobachtungen und Handlungen eben keiner kritischen Reflexion unterzogen werden und entsprechend effektives Handeln ausbleibt. Anzumerken ist jedoch, dass ein Nachweis divergierender kollektiver Deutungsmuster und deren Wirkung auf organisationales Handeln durch rekonstruktive Verfahren der schulischen Prozesse nicht vorgenommen wurden.

Zur Identifikation von Strategien auf unterrichtlicher Prozessebene sowie auf strukturell-organisationaler Ebene wurden Lehrer und Lehrerinnen, Schüler und Schülerinnen sowie Schulleitungen sowohl durch standardisierte Fragebögen, die im Wesentlichen auf der Basis gängiger Skalen der Schul- und Unterrichtsqualität entwickelt wurden, als auch durch leitfadengestützte Interviews befragt. Mit Blick auf die Ergebnisse der standardisiert quantitativen Erhebungen konnten kaum substanzielle Unterschiede zwischen den betrachteten Gruppen festgestellt werden (vgl. hierzu ausführlich auch Funke & Clausen 2015; Racherbäumer et al. 2013a; Racherbäumer et al. 2013b; Racherbäumer & van Ackeren 2015). Rückblickend ist kritisch zu hinterfragen, ob die vorhandenen quantitativen Instrumente zur Erfassung der

allgemeinen Schul- und Unterrichtsqualität geeignet sind, die spezifische Situation von Schulen in sozialräumlich deprivierter Lage so genau zu beschreiben, dass Unterschiede innerhalb dieser Schulen abbildbar sind. Gleichwohl konnten auf der Basis der qualitativ gewonnenen Daten Unterschiede festgestellt werden, die insbesondere zeigen, dass es auf ein gelungenes Passungsverhältnis der schulischen Ausgangsbedingungen einerseits und der schulischen Prozesse, Strukturen und Einstellungen andererseits ankommt, das durch organisationales Lernen hergestellt wird. Hier wurden zum Beispiel Strategien der Intensivierung der Schüler-Lehrer-Beziehung als wesentliches Merkmal von den Akteuren benannt, die Realisierung derselben auf der einzelschulischen Prozessebene gestaltet sich jedoch je nach internen situationalen Kontexten (z. B. mit Blick auf die Schulgröße) sehr heterogen. Insgesamt muss die beschriebene Intensivierung der Lehrer-Schüler-Beziehung jedoch auch kritisch hinterfragt werden, insbesondere vor dem Hintergrund der von den Interviewten genutzten Metaphorik der „Klassenmutter“ bzw. des „Klassenvaters“, die letztlich eine Rollenkonfusion impliziert. Diese kann in einem System, in dem Lehrkräfte Leistungen bewerten und Entscheidungen für die weitere Bildungslaufbahn treffen müssen, sowohl für Schüler und Schülerinnen als auch für die Lehrer und Lehrerinnen dauerhaft belastend werden.

Insgesamt ist es kaum möglich, Implikationen für konkretes strategisches Handeln für Schulen in sozialräumlich deprivierter Lage auszusprechen, ohne die einzelschulische Wirklichkeit zu kennen. Gleichwohl zeigen die vorliegenden Befunde insgesamt die Bedeutung und das Potenzial der systematischen datengestützten Analyse auf der Ebene der Einzelschule auf, die eine wichtige Basis für langfristige Veränderungen bieten kann. Ferner wurde an verschiedenen Stellen deutlich, dass der Schulleitung eine entscheidende Rolle zugesprochen werden muss, wenn es um die Initiierung und nachhaltige Implementation von Strategien des konstruktiven Umgangs mit erschwerten Kontextbedingungen geht.

Abschließend kann die Bildung des Samples auf der Basis der VERA-8-Daten kritisch hinterfragt werden, da der Erfolg der Schulen auch das Ergebnis gezielter Testcoachings sein kann. Hier wäre zukünftig der Einbezug weiterer Output-Ergebnisse der Einzelschulen sicherlich sinnvoll, wobei der alleinige Blick auf die Leistungen der Schüler und Schülerinnen als Referenzwert zur Erfassung schulischer Qualität auch sehr verkürzt ist.

Literatur

- Ackeren, I. van (2008). Schulentwicklung in benachteiligten Regionen – Eine exemplarische Bestandsaufnahme von Forschungsbefunden und Steuerungsstrategien. In Lohfeld, W. (Hrsg.), *Gute Schulen in schlechter Gesellschaft*. Wiesbaden: VS, S. 47–58.
- Argyris, C. & Schön, D. A. (2008). *Die lernende Organisation. Grundlagen, Methode, Praxis*. Stuttgart: Schäffer-Poeschel.

- Baumert, J., Stanat, P. & Watermann, R. (2006). *Herkunftsbedingte Disparitäten im Bildungswesen. Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS Verlag.
- Böttcher, W. & Keune, M. (2012). Externe Evaluation und die Steuerung der Einzelschule: Kontrolle oder Entwicklung? In Ratermann, S. & Stöbe-Blossey, S. (Hrsg.), *Governance von Schul- und Elementarbildung*, S. 63–80. Wiesbaden: Springer VS Verlag.
- Bonsen, M., Bos, W., Gröhlich, C., Harney, B., Imhäuser, K., Makles, A., Schräpler, J.-P., Terpoorten, T., Weishaupt, H. & Wendt, H. (2010). *Zur Konstruktion von Sozialindizes. Ein Beitrag zur Analyse sozialräumlicher Benachteiligung von Schulen als Voraussetzung für qualitative Schulentwicklung*. Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Bremm, N. (2015). Schulentwicklungsforschung. In Gogolin, I., Georgi, V., Krüger-Potratz, M., Lengyel, D. & Sandfuchs, U. (Hrsg.), *Handbuch Interkulturelle Pädagogik*. Julius Klinkhardt.
- Calkins, A., Guenther, W., Belfiore, G. & Lash, D. (2007). *The Turnaround Challenge. Why America's Best Opportunity to Dramatically Improve Student Achievement Lies in Our Worst-Performing Schools*. Main Report. Mass Insight Education & Research Institute. http://www.massinsight.org/publications/turnaround/51/file/1/pubs/2010/04/15/TheTurnaroundChallenge_MainReport.pdf [Zugriff am 22.10.2014].
- Creemers, B., Scheerens, J. & Reynolds, D. (2000). Theory Development in School Effectiveness Research. In Teddlie, C. & Reynolds, D. (Hrsg.), *The International Handbook of School Effectiveness Research*. London: Falmer Press, S. 283–298.
- Crosnoe, R., Kirkpatrick, J., Elder & Glen H. (2004). Intergenerational Bonding in School: The Behavioral and Contextual Correlates of Student-Teacher Relationships. In *Sociology of Education* 77, 1, S. 60–81.
- Ditton, H. (2013). Kontexteffekte und Bildungsungleichheit. Mechanismen und Erklärungsmuster. In Becker, R. & Schulze, A. (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*, S. 173–206. Wiesbaden: Springer VS.
- Faber, G. (2013). Klassenzusammensetzung als Kontextfaktor für die Leistungs- und Selbstkonzeptentwicklung. In Becker, R. & Schulze, A. (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*, S. 325–351. Wiesbaden: Springer VS.
- Feldhoff, T. (2011). *Schule organisieren*. Wiesbaden: VS Verlag.
- Flintham, A. (2006). *What's Good About Leading Schools in Challenging Circumstances?* Nottingham: National College for Leadership of Schools and Children's Services.
- Funke, C. & Clausen, M. (2015). Unterrichtliche Praxis mit segregierten Schülerschaften. Erscheint in Fölker, L., Hertel, T. & Pfaff, N. (Hrsg.), *Brennpunkt(-)Schule: Analysen, Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren*. Verlag Barbara Budrich.

- Gamoran, A., Porter, A. C., Smithson, J. & White, P. A. (1997). Upgrading High School Mathematics Instruction. Improving Learning Opportunities for Low-Achieving, Low-Income Youth. In *Educational Evaluation and Policy Analysis* 19, 4, S. 325–338.
- Gourmelon, A., Mroß, M. & Seidel, S. (2014). *Management im öffentlichen Sektor. Organisationen steuern – Strukturen schaffen – Prozesse gestalten*. Leck: Verlag Rehm.
- Hallinger, P. & Heck, R. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980–1995. In *School School Effectiveness and School Improvement*, 9 (2), S. 157–191.
- Hallinger, P. (2011): Leadership for learning: lessons from 40 years of empirical research. In *Journal of Educational Administration*, Vol. 49 iss. 2, S. 125–142.
- Hattie, J. A. C. (2009): *Visible Learning. A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London & New York: Routledge.
- Harris, A., Chapman, C., Muijs, D., Russ, J. & Stoll, L. (2006). Improving Schools in Challenging Contexts: Exploring the Possible. *School Effectiveness and School Improvement*, 17 (4), S. 409–424.
- Harris, A. & Chapman, C. (2001). *Leadership in Schools Facing Challenging Circumstances*. London: Department for Education and Skills.
- Helmke, A. (2012): *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. 4. überarbeitete Auflage. Seelze: Klett-Kallmeyer.
- Hendriks, M. & Steen, R. (2012). Results from School Leadership Effectiveness Studies (2005–2010). In Scheerens, J. (Hrsg.), *School Leadership Effects Revisited. Review and meta-analysis of empirical studies*, S. 65–129. Dordrecht: Springer.
- Herrmann, J. (2010). *Abschlussbericht zum Projekt „Schulentwicklung im System – Unterstützung von Schulen in schwieriger Lage“*. Hamburg: Landesinstitut für Lehrerbildung und Schulentwicklung. <http://li.hamburg.de/contentblob/3025990/data/pdf-abschlussbericht-zum-projekt-schulentwicklung-im-system-pdf-datei.pdf> [Zugriff am 21.09.2011].
- Hopkins, D. (2001): *Meeting the Challenge. An improvement guide for schools facing challenging circumstances*. London: Department for Education and Skills.
- Hopkins, D. & Reynolds, D. (2001). *The Past, Present and Future of School Improvement*. London: Department for Education and Skills.
- Horstkemper, M. & Tillmann, K. J. (2008). Schulformvergleiche und Studien zu Einzelschulen. In Helsper, W. & Böhme, J. (Hrsg.), *Handbuch der Schulforschung* (2., durchgesehene und erweiterte Auflage, S. 285–320. Wiesbaden: VS.
- Isaac, K. (2011). Neues Standorttypenkonzept. Faire Vergleiche bei Lernstandserhebungen. In *Schule NRW* 06/11, S. 300–301.
- Keys, W., Sharp, C., Greene, K. & Grayson, H. (2003). *Successful Leadership in Urban Contexts*. Nottingham: National College for School Leadership.
- Kerr, K. & West, M. (2010): *Social Inequality. Can Schools Narrow the Gap?* British Educational Research Association. www.bera.ac.uk/system/files/Insight2-web_1.pdf [Zugriff am 05.09.2014].
- Kieser, A. & Walgenbach, P.: *Organisation*. Stuttgart: Schäffer-Poeschel.
- Klein, E. D. (eingereicht). *Zwillinge oder entfernte Verwandte? Strategien der Schulleitung in Schulen in deprivierter Lage in Deutschland und den USA*.

- Kyriakides, Leonidas (2005): Extending the Comprehensive Model of Educational Effectiveness by an Empirical Investigation. In *School Effectiveness and School Improvement* 16, 2, S. 103–152.
- Leithwood, K., Harries, A. & Strauss, T. (2010). *Leading School Turnaround*. San Francisco: Jossey-Bass.
- Leithwood, K. A. & Riehl, C. (2003). *What We Know About Successful School Leadership*. Philadelphia: Laboratory for Student Success, Temple University.
- Ledoux, G. & Overmaat, M. (2001). *Op zoek naar succes: Een onderzoek naar basisscholen die meer en minder succesvol zijn voor autochtone en allochtone leerlingen uit achterstandsgroepen*. Amsterdam: SCO-Kohnstamm Instituut.
- March, J., G. & Olsen, J. P. (1976). *Ambiguity and Choice in Organizations*. Bergen: Universitetsforlaget.
- McDougall, D., Gaskell, J., Flessa, J., Kugler, J., Jang, E., Herbert, M., Pollon, D., Russell, P. & Fantilli, R. (2006): *Improving Student Achievement in Schools Facing Challenging Circumstances: Draft Report*. Ontario Institute for Studies in Education: Toronto.
- Mortimore, P. (1991). *Bucking the Trends: Promoting Successful Urban Education*. Paper presented at the Times Educational Supplement Greenwich Annual Lecture, London.
- Muijs, D. (2010): Changing Classroom Learning. In Hargreaves, A., Lieberman, A. Fullan, M. & Hopkins David (Hrsg.), *Second International Handbook of Educational Change*. Springer International Handbooks of Education, S. 857–867. Netherlands: Springer.
- Muijs, D., Harris, A., Chapman, C., Stoll, L. & Russ, J. (2004). Improving Schools in Socioeconomically Disadvantaged Areas. A Review of the Research Evidence. In *School Effectiveness and School Improvement*, 15 (2), S. 149–175.
- OECD (2011): *PISA 2009 Ergebnisse: Potenziale nutzen und Chancengerechtigkeit sichern – Sozialer Hintergrund und Schülerleistungen* (Band II). <http://dx.doi.org/10.1787/9789264095359-de> [Zugriff am 08.10.2014].
- Preisendörfer, P. (2011): *Organisationssoziologie. Grundlagen, Theorien und Problemstellungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Racherbäumer, K., Funke, C., Ackeren, I. van & Clausen, M. (2013a). Schuleffektivitätsforschung und die Frage nach guten Schulen in schwierigen Kontexten. In Becker, R. & Schulze, A. (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*, S. 239–267. Wiesbaden: Springer VS.
- Racherbäumer, K., Funke, C., Ackeren, I. van & Clausen, M. (2013b). Qualitätsentwicklung an Schulen in weniger begünstigter Lage. Empirische Befunde zur Datennutzung und zum Schulleitungshandeln. In Ackeren, I. van, Heinrich, M. & Thiel, F. (Hrsg.), *Steuerung im Bildungssystem, Beiheft DDS – Die Deutsche Schule*, S. 226–254. Münster: Waxmann.
- Racherbäumer, K. & Ackeren, I. van (2015). Was ist eine (gute) Schule in schwieriger Lage? Befunde einer Studie im kontrastiven Fallstudiendesign. Erscheint in Fölker, L., Hertel, T. & Pfaff, N. (Hrsg.), *Brennpunkt(-)Schule: Analysen, Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren*. Verlag Barbara Budrich.

- Reynolds, D. (1998). The Study of Remediation of Ineffective Schools: Some Further Reflections. In Stoll, L. & Myres, K. (Hrsg.), *No Quick Fixes. Perspectives on Schools in Difficulty*, S. 163–175. London: Falmer Press.
- Reynolds, D. (1996). Turning around Ineffective Schools: Some evidence and some Speculations. In Gray, J., Reynolds, D., Fitz-Gibbon, C. & Jesson, D. (Hrsg.), *Merging Traditions. The Future of Research and School Improvement*, S. 150–165. New York: Redwood Books.
- Reynolds, D. & Teddlie, C. (2000): The Processes of School Effectiveness. In Teddlie, C. & Reynolds, D. (Hrsg.), *The International Handbook of School Effectiveness Research*, S. 134–159. London: Falmer Press.
- Reynolds, D., Hopkins, D., Potter, D. & Chapman, C. (2001). *School improvement for Schools Facing Challenging Circumstances: A Review of Research and Practice*. London: Department for Education and Skills.
- Rossi, R. J., Stringfield, S. C., Daugherty, M. S., Goya, K. A., Herman, R., Montgomery, A. F., Nesselrodt, P. S., Royal, M. A., Vergun, P. B. & Young, E. M. (1997). *Education Reform and Students at Risk. Studies of Education Reform*. Office of Educational Research and Improvement. US Department of Education.
- Rolff, H.-G. (2013). *Schulentwicklung kompakt. Modelle, Instrumente, Perspektiven*. Weinheim: Beltz.
- Sammons, P., Gu, Q., Day, C. & Ko, J. (2011). Exploring the Impact of School Leadership on Pupil Outcomes. Results From a Study of Academically Improved and Effective Schools in England. *International Journal of Educational Management*, 25 (1), S. 83–101.
- Schwier, B. (2005). Ein blinder Fleck. Schulentwicklung als Schulverbesserung von „schwachen“ Schulen in England und Wales. Probleme und Entwicklungsansätze. In *Zeitschrift für Pädagogik*, 51 (2005) 3, S. 380–396
- Teddlie, C., Stringfield, S. & Reynolds, D. (2000). Context Issues within School Effectiveness Research. In Teddlie, C. & Reynolds, D. (Hrsg.), *The International Handbook of School Effectiveness Research*, S. 160–185. London: Falmer Press.
- Vahs, D. (2012). *Organisation. Ein Lehr- und Managementbuch*. Stuttgart: Schaeffer Poeschel.
- Vaughn, K. A. (2013). *Dispositions of Highly Effective Teachers and Student Achievement: Are We Closing the Gap?* Paper, präsentiert auf dem Jahreskongress der American Educational Research Association (AERA) in San Francisco, USA, 29.04.2013.
- Whitty, G. (2001). Education, Social Class and Social Exclusion. In *Education and Social Justice 1* (2001) 1, S. 2–8.
- Wissinger, J. (2011). Schulleitung und Schulleitungshandeln. In Terhart, E., Bennewitz, H. & Rothland, M. (Hrsg.), *Handbuch der Forschung zum Lehrerberuf*, S. 98–115. Münster u. a.: Waxmann.

Laura Fölker, Thorsten Hertel, Nicolle Pfaff

Schulische Praxis und Schulentwicklung unter Bedingungen sozialräumlicher Segregation – Befunde einer fallvergleichenden Studie zu zwei Schulen in benachteiligten Stadtteilen

Zusammenfassung

Der Beitrag stellt Ergebnisse einer fallrekonstruktiv angelegten Studie vor, die Formen der schulischen Integration von Lernenden an Schulen in segregierten Quartieren untersucht. Vor dem Hintergrund eines Erfahrungsraums, der von dauerhafter Prekarität schulischer Ordnung sowie von der Gefährdung des Schulimages geprägt ist, entwickeln die schulischen Akteure der beiden untersuchten Schulen je spezifische Schülerbilder sowie pädagogische und organisationale Praktiken. Diese bilden den Ausgangspunkt sowohl für Formen der Integration von Lernenden wie auch für die Umsetzung schulstruktureller Reformimpulse.

Bildungsbezogene Ungleichheit stellt ein beständiges Problem des deutschen Schulsystems dar. Sie betrifft vor allem Kinder aus sozial benachteiligten Lagen und in besonderem Maße solche mit Migrationshintergrund. Sie sind in Bildungsbeteiligung, schulischem Kompetenzerwerb und hinsichtlich der Chancen am Übergang von der Schule in die Berufsausbildung gegenüber ihren herkunftsdeutschen Mitschülerinnen und Mitschülern deutlich im Nachteil (für einen Überblick vgl. Stanat & Edele 2011; Diefenbach 2010). Soziale und bildungsbezogene Ungleichheiten manifestieren sich dabei auch stadträumlich. Während unter dem Begriff der urbanen Segregation zunächst die „ungleiche Verteilung von verschiedenen Gruppen der Bevölkerung auf die Quartiere einer Stadt“ verstanden wird (Häussermann 2008, S. 335), zeigt sie sich in bestimmten Quartieren deutscher Großstädte als ein Prozess der ökonomischen Verdrängung sozial benachteiligter Gruppen, der in besonderem Maße Migrantinnen und Migranten trifft. So entstehen Viertel, in denen zum großen Teil Menschen aus marginalisierten Schichten leben und von denen wiederum überproportional viele einen Migrationshintergrund aufweisen (vgl. z. B. Friedrichs & Triemer 2009, S. 109 ff.). In diesen Stadtteilen überschneiden sich also soziale und ethnische Segregation.

Dies hat dann häufig Folgen für die dort ansässigen Schulen. Die beschriebenen Dynamiken befördern Entmischungsprozesse an einzelnen Schulen, in deren Folge sich an ihnen Lernende aus den benachteiligten Milieus im Stadtteil konzentrieren – und das oftmals in höheren Relationen, als es die soziale Zusammensetzung des Quartiers nahelegen würde (vgl. Baur & Häussermann 2009). Als Ursachen hierfür

gelten, neben ökonomischen und ethnischen Verdrängungsprozessen auf der Ebene des Wohnraums, unter anderem das ausweichende Schulwahlverhalten ökonomisch besser gestellter Familien (z. B. ebd.; Maroy & van Zanten 2011) und bestimmte Verwaltungsentscheidungen, etwa zum Zuschnitt von Schuleinzugsgebieten (vgl. z. B. Radtke & Stosic 2009).

Segregierte Schulen sind im vergangenen Jahrzehnt in verschiedenen Forschungstraditionen zum Gegenstand gemacht worden (vgl. z. B. Ditton & Krüsken 2007; Radtke 2007; Lohfeld 2008; Baur & Häussermann 2009; Terpoorten 2014). Im angloamerikanischen Raum wird insbesondere im Kontext der Schuleffektivitätsforschung bereits seit Längerem zur Situation segregierter Schulen geforscht (vgl. Clarke 2005; Muijs et al. 2005). Hingegen liegen im deutschsprachigen Wissenschaftsraum bislang kaum empirische Studien zu diesem Thema vor (vgl. van Ackeren 2008). Diese Leerstelle wurde vom BMBF mit einem inhaltlichen Feld im Rahmen der Förderlinie „Steuerung im Bildungssystem“ aufgegriffen. Das Projekt „*Bildungsbezogene Integration unter vielfachen Bildungsrisiken – Fallstudien zu zwei Schulen mit multikulturellen Schülerschaften in benachteiligten Quartieren*“ zielte als eines von zwei Projekten innerhalb dieses Feldes auf die Rekonstruktion der Formen und Praktiken des pädagogischen Arbeitens an Schulen in benachteiligten Stadtteilen. Als ein steuerungsbezogen ebenso wie für die schulpädagogische Grundlagenforschung relevanter Fokus stellte sich dabei in der rekonstruktiven Arbeit die Frage heraus, welche strukturellen Herausforderungen sich für Schulen durch eine hohe soziale und ethnische Segregation ergeben und wie sie diese bearbeiten: Welche Erfahrungen machen pädagogische Akteurinnen und Akteure an segregierten Schulen und wie entwickeln sie auf dieser Basis Praktiken der schulischen Integration von Lernenden?

Diesen grundlegenden und bisher für den sozialstrukturellen und bildungspolitischen Raum der BRD nicht beantworteten Fragen ging das hier dargestellte Forschungsprojekt nach (zu Problemen der Übertragbarkeit der Befunde aus anderen Ländern vgl. van Ackeren 2008, S. 56).

1. „Bildungsbezogene Integration unter vielfachen Bildungsrisiken“ – Ziele und Untersuchungsdesign

Das Projekt geht von einer methodologischen Perspektive aus, in welcher die Weltansichten und Handlungen von Akteuren als ein Ausdruck von konkreten Erfahrungszusammenhängen verstanden werden (vgl. Mannheim 1964). Menschen in ähnlichen sozialstrukturellen und soziokulturellen Erfahrungszusammenhängen werden dabei als Angehörige von Milieus gefasst, die sich in ihrem Wissen und Handeln von anderen unterscheiden. Vor dem Hintergrund dieser Annahme, begründet über das fokussierte Interesse an den pädagogischen Praktiken segregierter Schulen, wurde für die Studie ein kontrastives Fallstudiendesign gewählt (vgl. Horstkemper & Tillmann 2003). Es wurden zwei Schulen nach sozialstrukturellen Gesichtspunkten ausgewählt. Beide Schulen liegen in benachteiligten Quartieren deutscher Großstädte, die sich durch ein Zusammentreffen von sozialer, demografischer und ethnischer Segregation auszeichnen. Sie sind nach lokalen Sozial- und Bevölkerungsstatistiken

durch hohe Arbeitslosigkeit und damit verbunden durch eine überdurchschnittliche Abhängigkeit der Bewohnerinnen und Bewohner von sozialstaatlichen Leistungen (insbesondere bei Kindern und Jugendlichen) gekennzeichnet. Die Stadtteile weisen im Vergleich zu den Gesamtstädten hohe Jugendquotienten und Ausländeranteile auf. Bei den untersuchten Schulen handelte es sich um weiterführende Schulen mit mehreren Bildungsgängen, auf denen vor allem Jugendliche aus den genannten Quartieren lernen. Ihre Schülerschaften weisen vor diesem Hintergrund ähnliche sozialstrukturelle Merkmale auf wie die Quartiere, in denen diese liegen.

Zur Erfassung der Perspektiven verschiedener schulischer Akteure und zur Generierung eines umfassenden Gesamtbildes der Untersuchungsschulen wurden ausgehend von der Schulleitung unterschiedliche schulische (Lehrkräfte, Akteure der Schulsozialarbeit, Aushilfs- und Ergänzungskräfte, Lernende, Eltern) und außerschulische Akteure (lokale Initiativen, lokale Jugendhilfeeinrichtungen, Schulverwaltung) befragt. Eingesetzt wurden dabei offene Erhebungsmethoden wie Interviews und Gruppendiskussionen, die einen Zugang zu den konkreten Erfahrungen der Akteure und zu ihren sozialen Praktiken eröffnen. Bis zum Projektende wurden so an beiden Schulen insgesamt über 100 Interviews, Gruppendiskussionen und Audiomitschnitte pädagogischer Interaktionen erhoben.

Zur Analyse zentraler Orientierungen der befragten Akteure an und um segregierte Schulen auf der Basis des erhobenen qualitativen Datenmaterials wurde die Textinterpretation der dokumentarischen Methode eingesetzt (vgl. Bohnsack 2003). Das Auswertungsverfahren zielt auf die Herausarbeitung des handlungsleitenden Wissens der Befragten. Handlungsleitende Orientierungen werden als habituell verankertes und in konkreten Erfahrungen begründetes Wissen verstanden (vgl. ebd.). Sie sind damit einerseits das Ergebnis einer bestimmten Lebenspraxis in ihrer sozialstrukturellen und soziokulturellen Verankerung, andererseits strukturieren sie diese, indem sie die Grundlage für konkretes Handeln bilden. Unterschiedliche qualitative Daten stellen in diesem Sinne Dokumentationen von Erfahrungsräumen und Milieus dar.

Zur Rekonstruktion handlungsleitender Orientierungen wird in der mehrschrittigen Analyse der dokumentarischen Textinterpretation auf die wissenssoziologische Unterscheidung von kommunikativem und konjunktivem Wissen zurückgegriffen (ebd.). Als „kommunikativ“ wird dabei ein explizites Wissen bezeichnet, das zur (alltags-)theoretischen Darstellung von Sachverhalten verwendet und gesellschaftlich geteilt wird. Dieses Wissen wird in der dokumentarischen Analyse im Schritt der formulierenden Interpretation herausgearbeitet, womit eine Systematisierung der im Datenmaterial enthaltenen thematischen Gehalte erfolgt. Als „konjunktives Wissen“ wird demgegenüber im Anschluss an die wissenssoziologischen Überlegungen Karl Mannheims ein implizites und von den Akteurinnen und Akteuren selbst nur eingeschränkt explizierbares Wissen verstanden. Dieses wird als eingelagert in geteilte Erfahrungsräume und als Ergebnis von strukturidentischen Sozialisationserfahrungen gefasst. Dieser Unterschied kann prägnant am Beispiel des Begriffes der Schule verdeutlicht werden. Mit ihm werden gesamtgesellschaftlich bestimmte Vorstellungen z. B. über Strukturen und Rollenverteilungen verbunden, während sich auf der Ebene des „konjunktiven Wissens“ die Erfahrungen von Lernenden und Lehrkräften, etwa

an auf Leistungsexzellenz ausgerichteten bürgerlich-humanistischen Gymnasien gegenüber an sozialer Teilhabe und lernbezogenem Leistungsausgleich ausgerichteten Gesamtschulen, grundlegend unterscheiden¹. Die Rekonstruktion des konjunkativen Wissens und damit der handlungsleitenden Orientierungen erfolgt in der dokumentarischen Analyse in der sog. reflektierenden Interpretation. Durch differenzierte und methodisch strukturierte Auseinandersetzung mit der Form der Darstellungen werden die für einen Erfahrungsraum spezifischen Wissensbestände rekonstruiert.

Im Vergleich der Erfahrungen von Akteuren an zwei segregierten Schulen erhalten wir damit einerseits Einblicke in den konjunkativen Erfahrungsraum der sozial und ethnisch segregierten Schule und andererseits, bezogen auf Differenzen zwischen den Schulen, Hinweise auf Spezifika der Einzelschulen im Umgang mit Phänomenen der Segregation. Innerhalb unserer Rekonstruktionen kristallisierten sich dabei *vier Analyseschwerpunkte* heraus, die aufgrund ihrer Relevanz in der Folge besonders intensiv bearbeitet wurden. Dabei handelt es sich *erstens* um die gesellschaftliche und schulische Bezugnahme auf Segregation und die mit ihr einhergehenden Herausforderungen für die Akteurinnen und Akteure in den untersuchten Schulen (vgl. Richter & Pfaff 2014; Fölker et al. 2015.; Pfaff et al. 2015). *Zweitens* wurde der Aspekt der schulischen Ordnung und ihre Gefährdung (vgl. Fölker et al. 2013; Hertel 2014; Hertel 2015) rekonstruiert. *Drittens* wurden die professionellen Orientierungen der pädagogischen Kräfte gegenüber den Lernenden und ihren Familien herausgearbeitet und auf ihre Implikationen bezüglich pädagogischer Praktiken befragt (vgl. Fölker et al. 2015; Fölker & Hertel 2015) sowie *viertens* der Umgang mit Steuerungsimpulsen aus Bildungspolitik und Schulverwaltung untersucht (vgl. Fölker 2013a; 2013b; 2015). Die Auswahl dieser Schwerpunkte ist dabei bereits ein Ergebnis der Analysen: Es sind jene inhaltlichen Bezugspunkte, die sich als besonders relevant für die pädagogische Arbeit an den untersuchten Schulen herausstellten. In den Begriffen der oben dargestellten dokumentarischen Methode gesprochen sind sie Teil der konjunkativen Erfahrungsräume der von uns interviewten Akteurinnen und Akteure und stehen dabei in einem engen und unmittelbaren Bezug zueinander.²

Die nachfolgende Ergebnisdarstellung orientiert sich an diesen Analysefokussen und fasst ausgewählte Befunde der Untersuchung zunächst zum Verhältnis von segregierter Schülerschaft und schulischer Praxis anhand der Aspekte Schulimage, schulische Ordnung und Schülerbilder zusammen. Ergebnisse der Studie in Bezug auf den Umgang mit steuerungspolitischen Impulsen und Hinweise auf Praktiken der Schulentwicklung an den Untersuchungsschulen werden in einem weiteren Schritt vorgestellt, bevor abschließend weitere Forschungsperspektiven skizziert werden.

1 Hinweise auf solche Differenzen finden sich bspw. in Studien zu differenziellen Lernmilieus in der Schulleistungsforschung (vgl. z. B. Stanat, Schwippert & Gröhlich 2010) sowie in der Schulklima- (vgl. Weishaupt, Böttcher & Plath 1999) oder Schulkulturforschung (vgl. z.B. Helsper et al. 2001).

2 Auf der Grundlage unserer Resultate wurden für die praktische Arbeit an Schulen in sozialstrukturell benachteiligten Stadtteilen Reflexionsimpulse zu den einzelnen Analyseschwerpunkten entwickelt, um den Transfer in die schulische Praxis anzuregen (vgl. www.brennpunkt-schule.de). Die online publizierten Ergebnisse sollen Diskussionsanreize für Lehrkräfte und schulische Gremien zum Umgang mit schulischer Segregation darstellen.

2. Schulische Ordnung und Schulimage an Schulen in benachteiligten Quartieren

„Die damalige Hauptschule, die Goethe Schule, hatte, wie ich angefangen bin, einen Ruf, der war ungefähr so wie der von der Willy Brandt Schule (...) nicht mehr zu beherrschen, keiner weiß mehr, wo's langgeht“ [Fallstudie 1: Gruppendiskussion Schulleitung, Schulgeschichte, 7–20].

„Was ich aber sagen will, is, dass das (...) unseren Ruf beispielsweise hier zum Stadtteil deutlich verbessert hat [Bw: ja]. Ähm weil deutlich war, und wenn wir 'ne Tage der offenen Tür, die wir machen, oder auch Feste, war immer das Thema Gewalt (...) wie ist es mit der Gewalt an Ihrer Schule (...), im letzten Jahr hab' ich es nicht mehr gehört“
[Fallstudie 2: Gruppendiskussion Schulleitung, 363–367].

In den abgedruckten Zitaten beschreiben die Schulleiter mit Verweis auf die Historie der Untersuchungsschulen diese als durch Ordnungsverstöße und Gewalt geprägt bzw. in der Öffentlichkeit als gewaltgeprägte Organisation wahrgenommen. Beide Dimensionen, die der Prekarität schulischer Ordnung sowie des öffentlichen Schulimages, bilden gemeinsame Erfahrungszusammenhänge der untersuchten Schulen.

Das überdurchschnittliche Vorkommen von Disziplinproblemen war bereits in anderen Studien zu segregierten Schulen beschrieben worden (vgl. z. B. McDougall 2006). Soziale Konflikte im Klassenzimmer und im Schulalltag sowie die Erfahrung massiver Unterrichtsstörungen beeinträchtigten die Gestaltung von Lehr-Lern-Prozessen in hohem Maße und stellten auch die Lehrkräfte der von uns untersuchten Schulen vor große Herausforderungen (vgl. Fölker et al. 2013). An beiden Untersuchungsschulen ergab sich hieraus eine Priorisierung der Herstellung von unterrichtlicher Ordnung gegenüber dem Fachunterricht (Fölker et al. 2013; mit Blick auf Schülerinnen und Schüler vgl. Hertel 2014). Unterschiede zwischen den Schulen bestehen in der grundsätzlichen Bezugnahme auf das Problem mangelnder Ordnung.

Während sich die Lehrkräfte im Rahmen der Fallstudie I als handlungssohnmächtig wahrnehmen und Störungen auf der Basis einer kollektiv geteilten Orientierung an Exklusion und Separierung bearbeiten, werden Störungen an der zweiten Fallstudie als Ausdruck von individuellen, sozialstrukturell bedingten Problemlagen verstanden und in professioneller Arbeitsteilung zum Gegenstand sozialpädagogischer Interventionen gemacht. Diese Praxis resultiert aus einer gemeinsamen handlungsleitenden Orientierung an der Wertschätzung von leistungs- und sozialverhaltensbezogener Heterogenität und an Individualisierung als Unterrichts- und Sozialprinzip (Fölker et al. 2013). Die Fragilität schulischer Ordnung als Erfahrungszusammenhang segregierter Schulen muss dabei vor dem Hintergrund strukturell verschärfter Bildungsungleichheit betrachtet werden. Niedrig qualifizierende, sozial und ethnisch segregierte Schulen arbeiten mit einer einerseits durch Leistungsselektion und andererseits durch sozialräumliche Segregation doppelt negativ segregierten Schülerschaft (vgl. Solga/Wagner 2005). Für sie verbinden sich mit der Institution Schule

einerseits Abwertungs- und Diskriminierungserfahrungen (vgl. z. B. Völcker 2014), andererseits bietet sie ihnen kaum adäquate Integrationsperspektiven.

Gleichzeitig bildet, wie das Eingangszitat aus der Fallstudie II zeigt, die Wahrnehmung von Gewalt an Schulen einen zentralen Zusammenhang der Konstruktion von Einzelschulen in der Öffentlichkeit. Auch einschlägige Studien verweisen auf ihre Bedeutung für das elterliche Schulwahlverhalten (vgl. Clausen 2006; Suter 2013). Daraus ergibt sich ein Erfahrungszusammenhang der schulischen Standortgefährdung durch niedrige Anmeldezahlen, der an beiden Schulen Krisenpotenzial entfaltet. Dies wird in unterschiedlicher Weise zum Ausgangspunkt von Schulentwicklungsprozessen. So veränderte sich eine der Untersuchungsschulen durch mehrfache systematische Öffnung der Organisation gegenüber neuen Schülerschaften, während an der zweiten Schule gemeinsame Praktiken der Schulentwicklung im Lehrerkollegium etabliert wurden (vgl. Fölker et al. 2015). Die Krisenerfahrung der Standortgefährdung nahm dabei an beiden Schulen zentrale Elemente und Auswirkungen der Wettbewerbssteuerung im Bildungssystem bereits vorweg. Wettbewerbssteuerung zielt im Rahmen aktueller Reformprozesse auf die Einführung von „Quasi-Märkten“ im Bildungswesen. In ihnen treten Einzelschulen verstärkt in Konkurrenz zueinander und sollen vor diesem Hintergrund zu Profilbildung und Schulentwicklung angeregt werden (vgl. z. B. Bellmann 2005). Die mit ihr verbundene Hoffnung auf eine Erhöhung der Chancengleichheit wurde bereits verschiedentlich in Zweifel gezogen (vgl. Bellmann 2007; Bellmann & Weiß 2009). Unsere Befunde liefern Hinweise auf einen hohen Bewährungs- und Scheiternsdruck, dem sich insbesondere segregierte Schulen aufgrund der oben beschriebenen Imagekrisen bereits lange vor der flächendeckenden Einführung schulischer Wettbewerbssteuerung ausgesetzt sahen. Ihre sozialstrukturelle Lagerung, das negative Schulimage und die Ausweichbewegungen der etablierten Mittelschicht sind Eckpfeiler dieser Entwicklung. Das legt den Schluss nahe, dass die Strukturen der Wettbewerbssteuerung insbesondere für solche Schulen jene bereits vorhandenen Krisenpotenziale zusätzlich erhöhen.

Vor dem Hintergrund eines durch diese strukturellen Problemlagen der Prekariät schulischer Ordnung und der schulischen Standortgefährdung gekennzeichneten Erfahrungsraums entfalten sich an den Schulen z. T. Selbstbeschreibungen der Akteurinnen und Akteure, in denen sich diese als Marginalisierte und Benachteiligte konstruieren (vgl. Richter & Pfaff 2014; Fölker et al. 2015; Hertel 2014). Andererseits bildet diese grundsätzliche sozial- und schulstrukturelle Einbettung auch die Grundlage pädagogischer und organisationaler Praxis an den untersuchten Schulen – dies wird im Folgenden an der Darstellung weiterführender Forschungsergebnisse zu Fragen der Klientelkonstruktion und der Schulentwicklung dargelegt.

3. „Klientelkonstruktionen“: Bilder und Vorstellungen über die Herkunftsmilieus der Jugendlichen und ihre Bedeutung für pädagogische Praktiken

„(...) und welche Sanktion wirkt überhaupt noch? Und das is, glaub ich, so'n Kernproblem der Klientel, die wir hier haben“ [Fallstudie 1: Gruppendiskussion AG Schulordnung, 589–590].

„(...) da wir in einem sozial schwachen Gebiet leben und die sozialen emotionalen Fähigkeiten zum Teil sehr eingeschränkt sind oder die Kinder auch 'n dicken Rucksack voll Schwierigkeiten mitbringen, müssen wir diesen Teil besonders auch unterstützen“ [Fallstudie 2: Gruppendiskussion Lehrende, 149–152].

Die im Rahmen unserer Studie rekonstruierten Orientierungen sind durchzogen von Bildern und Vorstellungen über die Schülerinnen und Schüler als „Klientel“ der pädagogischen Arbeit und als Teil des die Schule umgebenden sozialen Milieus. Die Ergebnisse der Rekonstruktionen liefern dabei Hinweise auf die ausgeprägte Wirkmächtigkeit professioneller Klientelkonstruktionen hinsichtlich der Ausgestaltung pädagogischer Praktiken etwa der Inklusion und Exklusion, aber auch von schulischen Betreuungsstrukturen. Einschlägige Studien haben dabei bereits auf die grundlegend ambivalente Verhältnisbildung zwischen Schule und Familie hingewiesen und betonen als ein Ergebnis die Tendenz zum defizitären Blick pädagogischer Professioneller auf die Jugendlichen und ihre Familien (vgl. z. B. Fritzsche & Rabenstein 2009; Scholz & Reh 2009). Für die Klientelbilder an den von uns untersuchten Schulen ist charakteristisch, dass bestehende Defizitkonstruktionen von Lernenden und ihren Familien zentral auf den Standort der Schule Bezug nehmen und die dort lebenden Familien als Sozialisationsraum und Erziehungszusammenhang tendenziell abwerten (vgl. Fölker & Hertel 2015).

Gleichsam zeigen sich im Fallvergleich der beiden Untersuchungsschulen grundlegende Differenzen auch in der Konstruktion des Schülerbildes. So werden die Lernenden von den pädagogischen Professionellen an der Fallstudie I vorwiegend als verhaltensunangepasst und deviant wahrgenommen und die ursächliche Wirkung dafür ihrem Herkunftsmilieu zugeschrieben. Das Milieu wird so zur Quelle der schülerseitigen Verhaltensunangepasstheit, die Schule letztlich durch milieubedingte Verhaltensweisen überwältigt und die Ausübung ihres Kerngeschäfts verunmöglicht. Das Kollegium begegnet der so wahrgenommenen Ausgangssituation durch Formen der Exklusion störender Schülerinnen und Schüler, die unter anderem durch einen Betreuungs- bzw. „Trainingsraum“ ermöglicht wird und stellenweise bis zu Versuchen reicht, bestimmte Lernende gezielt abzuschulen. Unterfüttert ist diese Vorgehensweise durch eine Klassifizierungspraxis, welche die Lernenden dichotom in „Beschulbare“ und „Unbeschulbare“ einteilt und anhand dieser Kategorien die besagten Ausschlussprozesse moderiert. In diesen Reaktionsweisen setzt sich damit die für das deutsche Schulsystem insgesamt charakteristische Tendenz zur Homogenisierung von Lerngruppen fort (vgl. Tillmann 2008), zielt hier jedoch zunächst auf die Kategorie des Verhaltens. Vor dem Hintergrund

des Erfahrungsraums sozialräumlicher Marginalisierung verbinden sich diese Orientierungen mit abwertenden Denkfiguren aus aktuellen Migrations- und Segregationsdiskursen (vgl. Yildiz 2014), zu deren Verstetigung sie im Umkehrschluss beitragen dürften. Nicht zuletzt sind somit Hinweise auf die Fortschreibung institutioneller Diskriminierungsmechanismen gegeben, wie sie anderenorts von Gommola und Radtke (2009) rekonstruiert wurden (vgl. Fölker et al. 2013, S. 107).

Auch an der Fallschule II wird das Milieu als Ausgangspunkt der wahrgenommenen Problematiken konstruiert. Die Schule wird hier jedoch gegenüber der Situation im Stadtteil als ein Schutz- und Schonraum für die Kinder entworfen. Zwar existieren auch hier Exklusionsarrangements für Störende, jedoch wird der Ausschluss aus dem Unterricht grundlegend mit Blick auf die Reintegration in selbigem betrieben. Dabei spielt die weiter oben bereits angesprochene Überführung der Kinder von unterrichtlichen in sozialpädagogische Betreuungssettings eine Schlüsselrolle. Dem zugrunde liegt eine im weitesten Sinne „pathologisierende“ Zuschreibung an die Jugendlichen. Diese und ihre Familien werden als „Bedürftige“ konstruiert, deren Problemlagen durch einzelfallbezogene Einwirkung innerhalb der Schule bearbeitet werden sollen. Diese pädagogischen Routinen haben ihren Ursprung in der ebenfalls oben dargestellten, schulkulturell tragenden Orientierung an Heterogenität als Bereicherung für Schule und Unterricht (Fölker et al. 2013, S. 94 ff.).

4. Zum Umgang mit bildungspolitischen Steuerungsimpulsen

„(...) unser Schulkonzept (...) beinhaltet schon sehr, sehr vieles oder fast alles (...), was die ‚integrierte Schulform‘ ahm, auch beinhaltet (...) und von daher war es für uns relativ leicht (...), diese neuen Sachen anzunehmen“ [Fallstudie 1: Interview, stellvertretender Schulleiter, 39–42].

„Insofern ist es heute für uns leichter als an anderen Schulen, weil wir uns nicht ändern müssen, weil wir da viel, ganz viele Dinge schon zu [Binnendifferenzierung: Anm. Aut.] gemacht haben“ [Fallstudie 2: Gruppendiskussion Schulleitung, Schulgeschichte, 430–432].

Aktuelle schulstrukturelle Reformen in vielen Bundesländern und Gemeinden der Bundesrepublik beziehen sich auf eine grundsätzliche Neuordnung der Schullandschaft. Dabei wird einerseits die Hauptschule aufgegeben und dem Gymnasium eine integrative Schulform gegenübergestellt (vgl. Tillmann 2012). Andererseits wird mit der gesetzlichen Verankerung von Inklusion und Individualisierung eine grundlegende Veränderung der Unterrichtspraxis gefordert (vgl. z. B. Wiechmann 2009). Hauptschulen in benachteiligten Stadtteilen bilden dabei einen besonderen Fokus, denn durch ihre Abschaffung und die gleichzeitige Etablierung eines integrierten zweigliedrigen Schulsystems soll der durch die Mehrgliedrigkeit begünstigten doppelt negativen Selektion Lernender (vgl. Solga & Wagner 2005) durch forcierte Durchmischung entgegengewirkt werden. Von den Schulen selbst verlangt die Re-

form, bestehende pädagogische Konzepte auf ihre Passung für eine heterogenere Schülerschaft hin zu befragen und gegebenenfalls zu modifizieren.

Die im Forschungsprojekt untersuchten Schulen standen im Zusammenhang mit regionalen Schulstrukturreformen vor verschiedenen Herausforderungen. Dies betraf einerseits die flächendeckende Einführung integrativer Schulformen mit mehreren Bildungsgängen, andererseits die damit verbundene inklusive Beschulung von Lernenden. Mit der Realisierung dieser strukturellen Veränderungen verbunden war für beide Schulen die Umsetzung von Schulfusionen. Dies eröffnete uns die Möglichkeit, intensive Prozesse der Schulentwicklung im Vollzug zu beobachten (Fölker 2013a; 2013b; 2015a; 2015b).

Gemeinsam ist beiden Untersuchungsschulen in der Bearbeitung der Steuerungsimpulse zunächst eine Priorisierung des Fusionsprozesses gegenüber der Umsetzung von Reformanforderungen (vgl. Fölker 2013a). Die Bearbeitung der Reformimpulse erfolgte an beiden Schulen in erster Linie über die Schaffung integrativer Lernumgebungen. Dabei kam es auf der Ebene handlungsleitender pädagogischer Orientierungen in dem Untersuchungszeitraum von zwei Jahren kaum zu Professionalisierungsprozessen (vgl. Fölker 2013b; 2015b).

Strukturell bestanden an beiden Untersuchungsschulen zunächst ähnliche Ausgangsbedingungen: So wurde an beiden Fallstudien eine integrierte Haupt- und Realschule mit einer Hauptschule zusammengelegt. Die jeweiligen Fusionschulen trugen dabei vergleichbare charakteristische Merkmale. Dazu gehört, dass in der jeweiligen Haupt- und Realschule durch einen vor Jahren realisierten Schulversuch bereits integrative Schulkonzepte etabliert worden waren. Demgegenüber hatten die Hauptschulen in Schulpreisen prämierte, klientelorientierte pädagogische Konzepte entwickelt. Unterschiede in den Ausgangslagen zur Fusion zeigten sich hinsichtlich materieller, räumlicher sowie kollegial-organisatorischer Rahmenbedingungen. Vor dem Hintergrund dieser Ausgangsbedingungen wurden sowohl die Schulfusionen wie auch die Umsetzung der Reformanforderungen in unterschiedlicher Form bewältigt.

Für die Fallstudie I zeigte sich bezogen auf die Bearbeitung des Fusionsprozesses ein langwieriger und konflikthafter Verlauf, in dem unterschiedliche Kulturen des Leitungshandelns und verschiedene pädagogische Orientierungen konfrontativ aufeinandertrafen. Beide Schulleitungen sowie zahlreiche Akteurinnen und Akteure beider Kollegien hielten dabei an den vor der Fusion etablierten Konzepten und Arbeitsformen fest, womit auch der Anspruch auf eine durch die jeweils andere Schule zu leistende Anpassung bzw. Unterordnung einherging (vgl. Fölker 2013a). In Gestalt der Legitimation eigener Handlungspraktiken wurde den Fusionsbedingungen gegenüber eigenlogisch argumentiert, sodass negative Zuschreibungen an die jeweils andere Schule und ihr Kollegium die Folge waren (ebd.). In diesem Prozess setzte sich eine Schulleitungspraxis organisatorisch durch, welche die Verantwortung für Schulentwicklungsprozesse an die an der Schule etablierte Organisationsebene der Jahrgangsteams überträgt. In diesem Sinne oblag auch die Umsetzung der Reformimpulse einem dafür zuständigen Jahrgangsteam. Dieses verhandelte die Reform vor dem Hintergrund eines defizitären Schülerbildes sowie einer daraus resultierenden handlungsleitenden Orientierung an der Reduktion von Heterogenität durch

Selektion primär im Sinne der Schaffung einer integrativen Lernumgebung (vgl. Fölker 2013b). Für das im Rahmen der Studie über einen Zeitraum von zwei Jahren bei der Umsetzung der Reform begleitete Team (vgl. ebd.; Fölker 2015a) bildet dabei das Strukturproblem der Gefährdung schulischer Ordnung (siehe Abschnitt 2) eine Rahmenbedingung des schulischen Handelns und damit die Grundlage ebenso wie die Grenze für die Umsetzung des Reformimpulses. So gelingt es den Teammitgliedern trotz intensiver Auseinandersetzung mit den Reformimpulsen nicht, die bislang bewährte Handlungspraxis, nämlich störende und nicht passformige Lernende zu separieren, aufzugeben (vgl. Fölker 2013b). Die Etablierung von Kontexten der parallel zum Regelunterricht erfolgenden Förderung von Lernenden mit Verhaltensauffälligkeiten und Leistungsdefiziten gerät vor diesem Hintergrund zur bestmöglichen Annäherung an den Reformimpuls (ebd.). Die gemeinsame Orientierung an der Reduktion von Heterogenität durch Selektion bleibt so stabil und die Umsetzung schulstruktureller Innovationen an die Herstellung einer schulischen und unterrichtlichen Ordnung als für schulische Segregation wesentliche Strukturproblematik gekoppelt.

Demgegenüber zeigen sich für die Fallstudie II sowohl bezogen auf die Fusion als auch auf die Bearbeitung der Reform der Integration der Bildungsgänge andere Praktiken der Schulentwicklung. Beide Innovationsimpulse werden an der Schule vor dem Hintergrund einer auf Schulentwicklungsprozesse bezogenen Verstetigung von Innovationen einerseits sowie der grundlegenden Wertschätzung von Heterogenität und Individualisierung als Unterrichtsprinzip andererseits bewältigt. So wurde die Realisierung der Schulfusion durch Rückgriff auf tradierte pädagogisch-kooperative Arbeitsroutinen und durch die Unterstützung einer externen Moderatorin in konsensuell verlaufenden Aushandlungsprozessen bearbeitet. Vor dem Hintergrund der zentralen pädagogischen Orientierungen an der Wertschätzung von Heterogenität (vgl. Fölker et al. 2013; Fölker 2015a) wurden bestehende Arbeitsweisen dabei kritisch reflektiert und an die neuen organisationalen Gegebenheiten angepasst (vgl. Fölker 2013a). Der Modus der Prüfung von Organisationsstrukturen der Schule wurde an der Fallstudie II in der Bewältigung vorangegangener Krisenprozesse entwickelt und wirkt auch in der Umsetzung der Schulstrukturreform (vgl. Fölker 2015a; 2015b). Der bildungspolitische Reformimpuls zur Etablierung einer inklusiven Schule mit mehreren Bildungsgängen wird von Schulleitung und in der Schulentwicklung aktiven Lehrkräften als an die bereits etablierten Handlungspraktiken der Anerkennung von Heterogenität und an die integrativen Lehr- und Lernformen anschlussfähig bewertet. Eine Veränderung schulischer Strukturen wurde vor dem Hintergrund dieser Einschätzung als nicht notwendig erachtet und dem Reformimpuls damit kein Innovationspotenzial zugesprochen (vgl. Fölker 2015b). Die Entwicklungsimpulse der Reform entfalten damit vor allem auf der schulorganisatorischen Ebene Innovationen, nicht aber auf der pädagogischen Interaktionsebene (ebd.).

Die Rekonstruktionen zu beiden Schulen zeigen, dass die pädagogischen Akteurinnen und Akteure eine intensive Auseinandersetzung mit den für sie relevanten Reformimpulsen führen, innerhalb welcher übergeordnete handlungsleitende Orientierungen sowohl auf der Ebene der pädagogischen Interaktion als auch auf der Ebene von Schulentwicklungspraktiken an beiden Schulen zunächst stabil bleiben.

5. Fazit und Ausblick

Der konjunktive Erfahrungsraum der von uns untersuchten Schulen ist durch die zwei strukturellen Problemlagen der Prekarität schulischer Ordnung sowie der Gefährdung des Schulimages gekennzeichnet. Sie manifestieren sich auf nahezu allen Ebenen des organisationalen Handelns, in Schulentwicklungsprozessen ebenso wie im Rahmen unterrichtlicher Praxis und hinsichtlich Fragen der Inklusion und Exklusion Lernender in den Unterricht. Sie moderieren diese schulischen Handlungszusammenhänge in bedeutender, jedoch auf Ebene der Einzelschulen unterschiedlicher Art und Weise und haben zusammen mit schulkulturell tradierten pädagogischen Orientierungen erheblichen Anteil an der Herausbildung von an den Schulen vorherrschenden Klientelbildern und pädagogischen Praktiken.

Dabei sind die angesprochenen Problemdimensionen in pädagogischen Zusammenhängen grundsätzlich keine unbekanntes Größen. Die sich aus der segregierten Lage der Schulen ergebende Spezifik scheint eher in der graduellen Ausprägung und umfassenden Dominanz dieser Problemdimensionen zu liegen. So ist die tendenzielle Brüchigkeit schulischer Ordnung auch in anderen Zusammenhängen beschrieben worden (vgl. z.B. Kalthoff & Kelle 2000). Unsere Rekonstruktionen zu segregierten Schulen legen den Schluss nahe, dass sich dort soziale Konflikte, die letztlich gesamtgesellschaftlichen Ursprungs sind, in verschärfter Form im schulischen Kontext niederschlagen und so die (Re-)Etablierung von Strukturen der (unterrichtlichen) Ordnung an den Einrichtungen nachhaltig irritieren. Ähnliches gilt für die Bestandsgefährdung durch Imageverlust, die im Zusammenhang der Einrichtung von Wettbewerbssteuerung und Quasi-Märkten im Bildungssystem als möglicher nichtintendierter Effekt diskutiert wird (vgl. z. B. Bellmann 2007; Bellmann & Weiß 2009; Radtke & Stošić 2009). Die Sorge um Imageverlust und sinkende Anmeldezahlen prägten das Handeln an den im Projekt untersuchten Schulen dabei bereits schon vor den Zeiten Neuer Steuerung – die segregierte Lage scheint die gesteigerte Einbindung der Schulen in ökonomische Zugzwänge bereits vorweggenommen zu haben. Die verstärkte Unterwerfung dieser Schulen unter ökonomische Zwänge dürfte diese Problemlage jedoch weiter verschärfen und wäre deshalb steuerungspolitisch kritisch zu reflektieren. Vor diesem Hintergrund müssen auch aktuelle Programme und Maßnahmen für segregierte Schulen, wie bspw. Standorttypenkonzepte (vgl. Racherbäumer et al. 2013), Strategien des „fairen Vergleichs“ (vgl. Fiege, Reuther & Nachtigall 2011) oder spezifische regionale Schulentwicklungsprogramme auf ihre latenten Wirkungen im Sinne der Stigmatisierung von Schulen und der einseitigen Delegation von Verantwortung für sozialstrukturelle Problemlagen an diese hinterfragt werden.

Die angesichts der beschriebenen schulischen Strukturprobleme herausgebildeten Orientierungen der Akteurinnen und Akteure bleiben dabei auch über verschiedene steuerungspolitische Impulse hinweg stabil. So wurden schulstrukturelle Reformimpulse an den untersuchten Schulen in erster Linie auf der Ebene der Schulorganisation aufgegriffen. Die in der Auseinandersetzung mit der jeweiligen Schülerklientel vor dem Hintergrund von defizitorientierten pädagogischen Konstruktionen entwickelten Orientierungen blieben stattdessen weitgehend stabil. In

diesem Zusammenhang konnten zwei zentrale Formen der schulischen Integration von Lernenden an segregierten Schulen aufgezeigt werden, die mit der Separierung von Lernenden im schulischen Raum einerseits und mit der Integration und sozialpädagogischen Intervention andererseits einhergehen. Die Dominanz der herausgearbeiteten struktureller Problemlagen weist darauf hin, dass Steuerungsimpulse an den von uns untersuchten Schulen zunächst zurückgestellt werden: die *Strukturprobleme* überlagern dort die *Strukturreform*.

Auf der Ebene von Leerstellen in der Forschung über schulische Segregation müssten Studien wie die hier vorgestellte durch Kontraste erweitert werden, die Datenmaterial auch aus nicht benachteiligten, sozialstrukturell besser gestellten Quartieren mit einbeziehen. Mit dem im Projekt gewählten minimalen Kontrast geht zwar das Potenzial einher, Schulen innerhalb von durch soziale und ethnische Segregation geprägten Erfahrungsräumen vergleichend zu untersuchen. Um die hier vorgestellten Ergebnisse jedoch weiter auszudifferenzieren, ist ein erweiterter Kontrast wünschenswert, den das Projekt aufgrund seiner Anlage schuldig bleiben musste. Drittens wäre weiterführende Forschung zur Verstrickung segregierter Schulen in gesellschaftlich tragende Marginalisierungs- und Stigmatisierungsdiskurse (vgl. z. B. Yildiz 2014) angezeigt, um der Frage nach ihrer Bedeutung für die symbolische Konstruktion und Praxis des schulischen Alltags unter Bedingungen sozialräumlicher Segregation weiter nachzugehen.

Literatur

- Ackeren, van, I. (2008): Schulentwicklung in benachteiligten Regionen. Eine exemplarische Bestandsaufnahme von Forschungsbefunden und Steuerungsstrategien. In W. Lohfeld (Hrsg.), *Gute Schulen in schlechter Gesellschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 47–58.
- Auernheimer, G. (Hrsg.) (2010): *Schieflagen im Bildungssystem. Die Benachteiligung der Migrantenkinder*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baur, Ch. & Häussermann, H. (2009): Ethnische Segregation in deutschen Schulen. *Leviathan* 37, S. 353–366.
- Bellmann, J. (2005). „Ökonomische Dimensionen der Bildungsreform. Unbeabsichtigte Folgen, perverse Effekte, Externalitäten.“ *Neue Sammlung* 45, 1, S. 15–31.
- Bellmann, J. (2007): Das Monopol des Marktes – Wettbewerbssteuerung im Schulsystem. *Berliner Debatte Initial* 18, S. 58–71.
- Bellmann, J. & Weiß, M. (2009): Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. Theoretische Konzeptualisierung und Erklärungsmodelle. *Zeitschrift für Pädagogik*, 55 (2), S. 286–308.
- Bohnsack, R. (2003): *Rekonstruktive Sozialforschung – Einführung in qualitative Methoden*. Opladen/Farmington Hills: Verlag Barbara Budrich, UTB.
- Clarke, P. (Hrsg.) (2005): *Improving Schools in Difficulty*. London: Continuum.
- 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), S. 69–90.

- Diefenbach, H. (2010): *Kinder und Jugendliche aus Migrantenfamilien im deutschen Bildungssystem. Erklärungen und empirische Befunde*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton, H. & Krüsken, J. (2007): Sozialräumliche Segregation und schulische Entwicklung. *Diskurs Kindheits- und Jugendforschung*, 2 (1), S. 23–38.
- Fiege, C., Reuther, F. & Nachtigall, C. (2011). Faire Vergleiche? – Berücksichtigung von Kontextbedingungen des Lernens beim Vergleich von Testergebnissen aus deutschen Vergleichsarbeiten. *Zeitschrift für Bildungsforschung*, 1, S. 133–149.
- Fölker, L. (2013a): „Da ist einfach ´ne gewisse Leistungs- und auch Schulentwicklungsgrenze erreicht!“ – Schulfusionen als zusätzliche Reformimpulse im Zuge der Integration von Haupt- und Realschulbildungsgängen. *Die Deutsche Schule DDS*, 12. Beiheft, S. 255–275.
- Fölker, L. (2013b): Organisationen im Wandel. *Zeitschrift für Qualitative Forschung ZQF*. 14 (1), S. 67–86.
- Fölker, L. (2015a): Schulstrukturereformen als Innovationsimpuls für Prozesse der Schulentwicklung an segregierten Schulen? In L. Fölker, T. Hertel & N. Pfaff (Hrsg.), *Brennpunkt(-)Schule: Analysen, Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren*. Opladen u. a.: Verlag Barbara Budrich, S. 211–230.
- Fölker, L. (2015b): Zur Konkretisierung von Reformimpulsen auf der Ebene der Einzelschule – Orientierungen von Lehrkräften bei der Integration von Haupt- und Realschulbildungsgang. In T.-S. Idel, F. Dietrich, K. Kunze, K. Rabenstein & A. Schütz (Hrsg.), *Professionsentwicklung und Schulstrukturreform. Zwischen Gymnasium und neuen Schulformen in der Sekundarstufe*. Bad Heilbrunn: Klinkhardt.
- Fölker, L. & Hertel, T. (2015): Differenz und Defizit. Rekonstruktionen zu pädagogischen Orientierungen und Praktiken der Klientelkonstruktion an zwei segregierten Großstadtschulen. In L. Fölker, T. Hertel & N. Pfaff (Hrsg.), *Brennpunkt(-) Schule: Analysen, Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren*. Opladen [u. a.]: Verlag Barbara Budrich, S. 107–124
- Fölker, L., Hertel, T. & Pfaff, N. (2015): Aberkennung von Erziehungsfähigkeit – Klientelkonstrukte als Ausdruck lokaler Bildungskulturen? In A. El Mafaalani & K. P. Strohmeier (Hrsg.), *Auf die Adresse kommt es an ... Segregierte Stadtteile als Problem- und Möglichkeitsräume*. Weinheim: Beltz-Juventa.
- Fölker, L., Hertel, T. & Pfaff, N. (Hrsg.) (2015): *Brennpunkt(-)Schule: Analysen, Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren*. Opladen u. a.: Verlag Barbara Budrich.
- Fölker, L., Hertel, T., Pfaff, N. & Wieneke, J. (2013): „Zahnlose Tiger“ und ihr Kerngeschäft – Die Abwesenheit schulischer Ordnung als Strukturproblem an Schulen in schwieriger Lage. *Zeitschrift für interpretative Schul- und Unterrichtsforschung ZISU*, 2 (2), S. 87–109.
- Friedrichs, J. & Triemer, S. (2008): *Gespaltene Städte? Soziale und ethnische Segregation in deutschen Großstädten*. VS Verlag für Sozialwissenschaften.
- Fritzsche, B. & Rabenstein, K. (2009): „Häusliches Elend“ und „Familienersatz“: Symbolische Konstruktionen in Legitimationsdiskursen von Ganztagschulen in der Gegenwart. In J. Ecarius, C. Groppe & H. Malmede (Hrsg.), *Familie und öffentliche*

- Erziehung. Theoretische Konzeptionen, historische und aktuelle Analysen.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 183–200.
- Gomolla, M. & Radtke, F.-O. (2009): *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule.* Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häusermann, H. (2008): Wohnen und Quartier: Ursachen sozialräumlicher Segregation. In E.-U. Huster, J. Boeckh & H. Mogge-Grotjahn (Hrsg.), *Handbuch Armut und soziale Ausgrenzung.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 335–349.
- Helsper, W., Böhme, J., Kramer, R.-T. & Lingkost, A. (2001): *Schulkultur und Schulmythos. Gymnasien zwischen elitärer Bildung und höherer Volksschule im Transformationsprozess. Rekonstruktionen zur Schulkultur.* Opladen: Leske und Budrich.
- Hertel, T. (2014): Zwischen Normbruch, Widerstand und ambivalenter Affirmation. Rekonstruktionen zu schulischer Ordnung und Subjektivierung an einer urbanen „Brennpunktschule“. In J. Hagedorn (Hrsg.), *Jugend, Schule und Identität. Selbstwerdung und Identitätskonstruktionen im Kontext Schule.* Wiesbaden: Springer-VS, S. 385–402.
- Hertel, T. (2015): Unterlaufen und Kontern statt „Überwachen und Strafen“? Rekonstruktionen zu pädagogischen Praktiken im „Trainingsraum“ einer segregierten Großstadtschule. In L. Fölker, T. Hertel & N. Pfaff (Hrsg.), *Brennpunkt(-)Schule: Analysen, Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren.* Opladen u. a.: Verlag Barbara Budrich, S. 155–170.
- Horstkemper, M. & Tillmann, K.-J. (2003): Schulformvergleiche und Studien zu Einzelschulen. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 287–324.
- Kalthoff, H. & Kelle, H. (2000): Pragmatik schulischer Ordnung. Zur Bedeutung von „Regeln“ im Schulalltag. *Zeitschrift für Pädagogik*, 46, S. 691–710.
- Lohfeld, W. (2008) (Hrsg.): *Gute Schulen in schlechter Gesellschaft.* Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mannheim, K. (1964): Beiträge zur Theorie der Weltanschauungs-interpretation. In K. Mannheim, *Wissenssoziologie.* Neuwied, S. 91–154.
- Maroy, Ch. & Zanten van, A. (2011): Steuerung und Wettbewerb zwischen Schulen in sechs europäischen Regionen. In H. Altrichter, M. Heinrich & K. Soukup-Altrichter (Hrsg.), *Schulentwicklung durch Schulprofilierung? Zur Veränderung von Koordinationsmechanismen im Schulsystem.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 195–213.
- McDougall, D., Gaskell, J., Flessa, J., Kugler, J., Jang, E. E., Herbert, M., Pollon, D., Russel, P. & Fantilli, R. (2006): *Improving Student Achievement in Schools Facing Challenging Circumstances: Draft Report.* Ontario Institute for Studies in Education: Toronto.
- Muijs, D., Harris, A., Chapman, Ch., Stoll, L. & Russ, J. (2005): Improving Schools in Socio-Economically Disadvantaged Areas: A Review of Research Evidence. In P. Clarke (Hrsg.), *Improving Schools in Difficulty.* London: Continuum, S. 95–114.
- Pfaff, N., Fölker, L. & Hertel, T. (2015): Schule als Gegenraum zum Quartier – Sozialräumliche Segregation und die Prekarität grundlegender schulischer Bedingungen. In L. Fölker, T. Hertel & N. Pfaff (Hrsg.), *Brennpunkt(-)Schule: Analysen,*

- Probleme und Perspektiven zur schulischen Arbeit in segregierten Quartieren.* Opladen u. a.: Verlag Barbara Budrich, S. 67-85.
- Racherbäumer, K., Funke, Ch., Ackeren, I. van & Clausen, M. (2013): Schuleffektivitätsforschung und die Frage nach guten Schulen in schwierigen Kontexten. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen.* Wiesbaden, VS Verlag für Sozialwissenschaften, S. 239-267.
- Radtke, F.-O. (2007): Segregation im deutschen Schulsystem. In W.-D. Bukow, C. Nikodem & E. Schulze (Hrsg.), *Was heißt hier Parallelgesellschaft? Zum Umgang mit Differenzen.* Wiesbaden: VS Verlag für Sozialwissenschaften.
- Radtke, F.-O. & Stošić, P. (2009): Lokale Bildungsräume: Ansatzpunkte für eine integrative Schulentwicklung. *Geographische Revue* 11, S. 34-51.
- Richter, E. & Pfaff, N. (2014): (Schulische) Bildung in ethnisch segregierten Stadtteilen – quantitative und qualitative Befunde im Zusammenhang. In W. Baros & W. Kempf (Hrsg.), *Erkenntnisinteresse, Methodologie und Methoden interkultureller Bildung.* Berlin: Regener-Verlag, S. 209-223.
- Scholz, J. & Reh, S. (2009): Verwahrloste Familien – Familiarisierte Schulen. Zum Verhältnis von Schule und Familie in den Diskursen der deutschen Schulgeschichte seit 1800. In F.-U. Kolbe (Hrsg.), *Ganztagsschule als symbolische Konstruktion: Fallanalysen zu Legitimationsdiskursen in schultheoretischer Perspektive.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 159-177.
- Solga, H. & Wagner, S. (2005): Die Zurückgelassenen – die soziale Verarmung der Lernumwelten von Hauptschülerinnen und Hauptschülern. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen von Bildungsungleichheit.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 191-220.
- Stanat, P. & Edele, A. (2011). Migration und soziale Ungleichheit. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 181-192.
- Stanat, P., Schwippert, K. & Gröhlich, C. (2010): Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb. Längsschnittliche Überprüfung eines umstrittenen Effekts. *Zeitschrift für Pädagogik*, Beiheft. 55 Weinheim u. a.: Beltz, S. 147-164.
- Suter, P. (2013): *Determinanten der Schulwahl. Elterliche Motive für oder gegen Privatschulen.* Wiesbaden: VS Verlag für Sozialwissenschaften.
- Terpoorten, T. (2014): *Räumliche Konfiguration der Bildungschancen. Segregation und Bildungsdisparitäten am Übergang in die weiterführende Schule im Agglomerationsraum Ruhrgebiet.* Bochum: ZEFIR.
- Tillmann, K.-J. (2008): Die homogene Lerngruppe – oder: System jagt Fiktion. In H.-U. Otto & U. Rauschenbach (Hrsg.), *Die andere Seite der Bildung. Zum Verhältnis von formellen und informellen Bildungsprozessen.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. 33-39.
- Tillmann, K.-J. (2012): Das Sekundarschulsystem auf dem Weg in die Zweigliedrigkeit. Historische Linien und aktuelle Verwirrungen. *PÄDAGOGIK* online. Abrufbar unter: <http://www.redaktion-paedagogik.de/2012/05/das-sekundarschulsystem-auf-dem-weg-in-die-zweigliedrigkeit/>.

- Völcker, M. (2014): „Und dann bin ich auch noch Hauptschule gekommen.“ Über die identitären Folgen der Hauptschulzugehörigkeit. UVK.
- Weishaupt, H., Böttcher, I. & Plath, M. (1999): *Gestaltung einer neuen Schulstruktur – Zur Entwicklung von Regelschulen und Gymnasien in Thüringen*. Münster u. a.: Waxmann.
- Wiechmann, J. (2009): Gemeinschaftsschule – ein neuer Begriff in der Bildungslandschaft. *Zeitschrift für Pädagogik*. 55 (3), S. 409–429.
- Yildiz, E. (2014): Migrationsfamilien: Vom hegemonialen Diskurs zur (transnationalen) Alltagspraxis. In T. Geisen, T. Studer & E. Yildiz (Hrsg.), *Migration, Familie und Gesellschaft. Beiträge zu Theorie, Kultur und Politik*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 59–71.

4. Schulleitungen, Fachkonferenzleitungen und Lehrkräfte als Akteure Neuer Steuerung

*Sebastian Wurster, Anabel Bach, Anna Schliesing,
Katja Thillmann, Hans Anand Pant, Felicitas Thiel*

Schulen als Steuerungsakteure im Bildungssystem – datenbasierte Schul- und Unterrichtsentwicklung aus der Perspektive von Schulleitungen, Fach- konferenzleitungen und Lehrkräften

1.1 Einleitung

Anfang der 2000er-Jahre setzten Veränderungen im deutschen Schulsystem ein. Diese umfassen u. a. die schrittweise Implementation von Qualitätsstandards (Bildungsstandards und Schulqualitätsrahmen) sowie darauf ausgerichtete neue Formen der Evaluation, wie Vergleichsarbeiten (VERA), Schulinspektionen und schulinterne Evaluationen. Zudem wurden in allen Bundesländern, in denen noch keine zentralen Prüfungen durchgeführt wurden, zentrale Abschlussprüfungen eingeführt.¹ Damit einhergehend sind Einzelschulen zunehmend zu eigenständigen Steuerungsakteuren im Bildungssystem geworden. Der Entwicklung evaluativer Potenziale kommt bei der Verlagerung von Steuerungskompetenzen auf die schulische Organisationsebene eine zentrale Bedeutung zu. Im BMBF-geförderten Projekt *Schulen als Steuerungsakteure im Bildungssystem* (StABil) wurden die schulische Praxis im Umgang mit evaluativem Wissen aus externer Evaluation (Schulinspektion), interner Evaluation und Ergebnissen aus Schulleistungstests (Vergleichsarbeiten und zentrale Prüfungen) sowie die organisationalen und individuellen Bedingungen einer Datennutzung in verschiedenen Teilstudien an Berliner, Brandenburger und nordrhein-westfälischen Schulen untersucht.

1.2 Schulen als Steuerungsakteure

Steuerung des Bildungssystems und Schulen als Steuerungsakteure

International ist eine Neuausrichtung der Bildungspolitik seit den 1980er-Jahren zu registrieren. Wenngleich die bildungspolitischen Reformstrategien in den einzelnen Ländern unterschiedlich begründet werden, adaptieren die Steuerungsmodelle doch wesentliche Prinzipien des New Public Management (vgl. Hentschke & Davies 1997; Dubs 2005), die mit der Trennung von strategischer und operativer Führung und der Implementation von Controlling-Instrumenten auf eine deutliche Verbesserung der

¹ Mit Ausnahme des Landes Rheinland-Pfalz.

Leistungserbringung bei effizienterem Ressourceneinsatz zielen. Eine Reihe von operativen Steuerungsentscheidungen wurde im Zuge einer schrittweisen Implementation des neuen Steuerungsmodells von der System- auf die Organisationsebene verlagert. Was Schulen betrifft, sind dies etwa Entscheidungen über das Profil und das Programm, curriculare Entscheidungen oder Entscheidungen über die Personalentwicklung. Allerdings variieren international die Freiheitsgrade für Schulen zur eigenständigen Umsetzung von Entwicklungsvorhaben, bei denen die deutschen Länder, z. B. hinsichtlich der Autonomie bei Personalentscheidungen, eher am unteren Ende rangieren (Brauckmann 2012; vbw 2010). Die Schule selbst wird zu einem zentralen Steuerungsakteur: „Current accountability mechanisms ... take the schools as the unit of accountability and seek to improve student learning by improving the functioning of the school organization“ (O’Day 2002, S. 2).

In der Formel von der Verpflichtung der Schule auf Ergebnisverantwortung bei gleichzeitiger Ausweitung schulischer Gestaltungsspielräume wird das neue Steuerungsmodell auf den Punkt gebracht (vgl. allgemein zu Steuerungsmechanismen: Wiesenthal 2005; Benz et al. 2007; zur Frage der Ablösung der Steuerungstheorie durch die Governance-Theorie: Kussau & Brüsemeister 2007 und kritisch: Mayntz, 2005). Im Unterschied zu der bis in die 1980er-Jahre hinein dominierenden bildungspolitischen Steuerungsstrategie des „Synoptic Planning and Bureaucratic Structuring“ und im Unterschied zu einer vorwiegenden Steuerung über den „Market Mechanism“ setzt das neue Steuerungsmodell, das mit dem Begriff der Outputsteuerung nur unzureichend beschrieben ist, auf „Retroactive Planning“ (Scheerens & Demeuse 2005), d. h. ein Planungshandeln, das primär auf der systematischen Auswertung zurückliegender Erfahrungen basiert. In diesem Steuerungsmodell hat Evaluation einen zentralen Stellenwert.

Die Rolle von Evaluationen bei der Selbststeuerung von Schulen

Beide Aspekte des neuen Steuerungsmodells, die Delegation von Entscheidungskompetenzen an die Einzelschule und die zentrale Bedeutung der Entwicklung evaluativer Potenziale, lassen sich aus den bildungspolitischen Programmpapieren (vgl. z. B. SenBWF 2007) und den neuen Schulgesetzen herauslesen. In ihnen ist als Ideal der neuen Schule, die *Schule als lernende Organisation*, beschrieben. Eine lernende Schule ist eine Schule, in der Zieldefinition und -operationalisierung, Maßnahmenentwicklung und Professionalisierung evidenzbasiert, d. h. auf der Basis verfügbarer Rückmeldungen aus Schülerleistungstests, den Ergebnissen zentraler Prüfungen, der externen Evaluation (Schulinspektion) und einer regelhaften Überprüfung der Ergebnis- und Prozessqualität durch interne Evaluation, erfolgen. Im Sinne der Schließung des Qualitätszyklus übersetzt die lernende Schule diese Daten kontinuierlich in Verfahren zur Prozessoptimierung (Silins & Mulford 2004). Kennzeichnend für lernende Schulen sind entsprechend spezifische „evaluation policies and practices“ (Scheerens & Bosker 1997, S. 307).

Die Entwicklung eines evaluativen Potenzials in Schulen setzt nicht lediglich die Akzeptanz, Bereitschaft und Kompetenz zur Generierung und Nutzung evaluativen

Wissens bei den einzelnen Lehrkräften voraus, sondern erfordert darüber hinaus die Ausbildung neuer Organisationsstrukturen, die Einführung neuer Abstimmungs- und Koordinationsprozesse, vor allem aber ein Personalmanagement, das Professionalisierung und Unterrichtsentwicklung im Dienste der Zielerreichung gewährleistet. Darüber hinaus wurde im Zuge der Einführung des neuen Steuerungsmodells die strategische Spitze, die Schulleitung, sowohl im Verhältnis zur Schulaufsicht als auch im Verhältnis zum Kollegium deutlich gestärkt. Mit der Implementation des neuen Steuerungsmodells werden demnach umfangreiche Reorganisationsforderungen an Schulen verknüpft (Silins & Mulford 2004). Dies findet seinen Ausdruck in der schulgesetzlichen Neudefinition der Rolle der Lehrkräfte und der Schulleitung sowie in einer neuen Zuständigkeits- und Aufgabenverteilung zwischen den Organen der Schulverfassung (Pedder & MacBeath 2008).

Unter welchen Bedingungen eine Implementation bildungspolitischer Reformstrategien in Schulen gelingen kann und welche systematischen Reibungsverluste entstehen, ist eine zentrale Frage der empirischen Bildungsforschung (vgl. Spillane et al. 2002; Coburn 2004; Gräsel et al. 2007; Pant et al. 2008). Das gilt insbesondere auch für die Implementation von Outputsteuerung und Accountability-Systemen (Elmore, Abelman & Fuhrman 1996; DeBray, Parsons & Woodworth 2001; O'Day 2002; Mintrop & Trujillo 2007) und den Umgang mit Evaluationsdaten. Mit Blick auf die Ausgestaltung von Accountability-Systemen lassen sich international bedeutende Unterschiede feststellen. Während im angloamerikanischen Raum die neuen Steuerungsparadigmen in der Regel mit bedeutsamen Konsequenzen (high-stakes) für die schulischen Akteure verbunden sind (z. B. Veröffentlichung der Ergebnisse, Schulschließung), ist das im deutschsprachigen Raum nicht der Fall (low-stakes) (vgl. z. B. Dederling 2012; Maag Merki 2010).

Faktoren gelingender Evaluationsnutzung

Die Nutzung von Daten ist sehr voraussetzungsvoll und hängt mit einer Vielzahl von Faktoren zusammen. Es liegen einige Modelle vor, die Barrieren und Gelingensbedingungen für die Nutzung von Informationen aus Evaluationen beschreiben. Das Modell von Visscher und Coe (2003) beschreibt beispielsweise Einflussfaktoren auf den Prozess schulinterner Nutzung von Daten aus „*school performance feedback systems*“ (SPFS). Die Datennutzung intendiert eine Veränderung schulischer Prozesse und Strukturen und kann so zu einer Steigerung der Schülerleistung beitragen. Neben diesen erwünschten Effekten können jedoch auch nicht intendierte Effekte, wie z. B. ein Teaching-to-the-Test, auftreten (z. B. Koretz 2005).

Die Einflussfaktoren lassen sich grob in zwei Bereiche unterteilen: die Eigenschaften des Evaluationsverfahrens und der schulische Kontext. Der schulische Kontext umfasst mögliche Faktoren, wie z. B. die Kooperation und Innovationsorientierung im Kollegium oder Ressourcen, die zur Datennutzung eingesetzt werden.

Das Modell von Visscher und Coe (2003) eignet sich gut als ein Ausgangspunkt zur Analyse von verschiedenen Evaluationsverfahren, da es die Eigenschaften der Evaluationsverfahren explizit benennt und als Einflussfaktoren operationalisiert. Zum einen wird der Schulsystemkontext genannt, der die zu erreichenden Ziele sowie die zu erfüllenden Funktionen des SPFS umfasst. Diese strategischen Entscheidungen auf Systemebene haben Auswirkungen auf die Ausgestaltung des SPFS, beispielsweise auf die Gestaltung der Rückmeldung der erfassten Informationen. Zudem wird auch die Implementation des SPFS ins Modell integriert, zu der u. a. ein Unterstützungssystem gezählt wird. Andere Modelle, wie z. B. von Schildkamp und Kuiper (2010), erweitern die von Visscher und Coe (2003) vorgeschlagenen Faktoren um die explizite Nennung von Einflussfaktoren auf Individualebene, wie z. B. die Akzeptanz des Verfahrens. Helmke (2004) beschreibt die Verarbeitung und Verwendung von Evaluationsergebnissen in einem Sequenzmodell. Nach dem Vorliegen der Daten müssen diese rezipiert und reflektiert werden, bevor Maßnahmen (Aktionen) abgeleitet und durchgeführt werden. Anschließend sollten die Maßnahmen evaluiert werden.

Grundsätzlich ist festzuhalten, dass in den Modellen die Ebene der Organisation zwar Berücksichtigung findet, jedoch nur auf Basis empirischer Befunde und nicht auf Basis differenzierter Theorien. Nur wenige Untersuchungen zur Implementation von Accountability-Systemen im Schulsystem adressieren die Organisationsebene explizit (wie etwa Mintrop & Trujillo 2007 oder Coburn 2004). Die zentralen Ansatzpunkte für die Selbststeuerung von Schulen sind, entsprechend einem auf der Grundlage der organisationstheoretischen Ansätze von Niklas Luhmann und Henry Mintzberg entwickelten organisationstheoretischen Modell der Schulentwicklung (Thiel 2008a, 2008b):

- Schulprogramme, die strategische Ziele und entsprechende Maßnahmen festschreiben,
- neue Formen der Aufgabenteilung zur Entwicklung und Umsetzung dieser Maßnahmen, wie die Stärkung der Schulleitung sowie der Fachkonferenzen und die Einrichtung von Steuergruppen,
- eine Personalentwicklung, die dafür Sorge trägt, dass Lehrpersonen diejenigen Kompetenzen erwerben oder weiterentwickeln, die für eine erfolgreiche Umsetzung der geplanten Maßnahmen notwendig sind, und
- eine innovations- und entwicklungsoffene Schulkultur.

Ob Zielformulierung, Maßnahmen- und Personalentwicklung datenbasiert erfolgen, ist entsprechend der Logik des *Retroactive Planning von entscheidender Bedeutung*.

1.3 Evaluationen als Ausgangspunkt für datenbasierte Schul- und Unterrichtsentwicklung

Schulen bekommen Informationen aus verschiedenen Evaluationsverfahren, die alle als Ausgangspunkt zur Qualitätsentwicklung dienen sollen. Idealtypisch verläuft eine schrittweise Veränderung in der Logik des Lernzirkels: „Measurement and assessment of performance; evaluative interpretation based on given or newly created norms; communication or feedback of this information to units that have the capacity to take corrective action; actual and sustained use (learning) of this information to improve organisational performance“ (Scheerens & Demeuse 2005, S. 378). Die Ergebnisse von Schulinspektionen, schulischer Selbstevaluation, Vergleichsarbeiten und zentralen Abschlussprüfungen stellen für Schulen vier Quellen evaluativen Wissens dar, die im Folgenden näher erläutert und in Beziehung zueinander und zum Schulentwicklungsmodell (Thiel 2008a, 2008b) gesetzt werden.

1.3.1 Schulbezogene Evaluationsverfahren im Überblick

Die *Schulinspektion* stellt ein externes Evaluationsverfahren dar, bei dem Merkmale der Prozessqualität im Fokus stehen. Im Rahmen einer Schulinspektion werden die Unterrichts- und Schulqualität durch ein Inspektionsteam auf Basis von Orientierungsrahmen zur Schulqualität (vgl. z. B. Brandenburger Orientierungsrahmen [MBJS, 2008]) analysiert (Döbert & Dederling 2008). Die Orientierungsrahmen beschreiben die (normativen) Zielvorstellungen von „guter Schule“ und operationalisieren diese in verschiedenen Qualitätsdimensionen, z. B. „Lehr- und Lernprozesse“, „Schulkultur“, „Schulmanagement“, die wiederum weiter aufgefächert werden (vgl. SenBWF 2007). Bei der Erfassung wird auf verschiedene Methoden, wie beispielsweise Interviews, Dokumentenanalysen, Surveys und Unterrichtsbeobachtungen, zurückgegriffen (Gärtner & Pant 2011). In den meisten dieser Handlungs- oder Orientierungsrahmen der 16 Bundesländer wird die Art des Einsatzes bzw. der Nutzung der anderen genannten Evaluationsverfahren zur Qualitätsentwicklung (Ergebnisse aus VERA und zentralen Prüfungen, interne Evaluation) als Qualitätsindikator benannt.

Interne Evaluation oder *Selbstevaluation von Schulen* stellt eher einen Sammelbegriff für zahlreiche Verfahren dar, die auf Initiative der Schule durchgeführt werden bzw. für deren Durchführung die Einzelschule verantwortlich ist und diese eigenständig gestaltet (Gärtner 2013). Es sollen Informationen über den Erfolg der eigenen Arbeit (Berkemeyer & Müller 2010) bzw. zur Bestandsaufnahme der eigenen Ressourcen gewonnen werden (Thiel & Thillmann 2012). Im Rahmen des Ansatzes der Schule als lernender Organisation fällt der internen Evaluation deshalb eine entscheidende Rolle zu (Thiel & Thillmann 2012). Interne Evaluation umfasst jede Form der systematischen Informationsbeschaffung und -bewertung in der Verantwortung der Schule.

Vergleichsarbeiten (VERA) und *zentrale Abschlussprüfungen* bzw. die Ergebnismeldungen nach deren Durchführung können in Schulen als Informati-

onsquellen zu den Leistungserträgen der Schule verwendet werden bzw. Entwicklungsbedarfe auf Klassen- und Schulebene identifizieren. Beiden Verfahren liegen für die schulischen Kernfächer – zumindest mittelbar – die ländergemeinsamen Bildungsstandards der Kultusministerkonferenz (z. B. KMK 2004) als Referenzrahmen zugrunde. Für zentrale Abschlussprüfungen wie den mittleren Schulabschluss (MSA) dienen Rahmenlehrpläne als Referenzrahmen, die an den Bildungsstandards orientiert sind (LISUM 2008). Die Bildungsstandards sind für VERA ebenfalls der Referenzrahmen (Köller 2009). Die Ergebnisse der VERA-Tests sollen als Ausgangspunkt für Schul- und Unterrichtsentwicklungsmaßnahmen genutzt werden, indem das Erreichen der in den Bildungsstandards formulierten Kompetenzziele bereits zu einem früheren Zeitpunkt überprüft wird. Diese Informationen über den Leistungsstand der Schülerschaft werden z. B. in Berlin und Brandenburg über Rückmeldungen für verschiedene Ebenen (Individualebene, Klassenebene und Schulebene, vgl. Pant, Emmrich, Harych & Kuhl 2011) den Lehrkräften und Schulleitungen zur Verfügung gestellt. Zentrale Abschlussprüfungen sind dadurch gekennzeichnet, dass in ihnen „alle Schülerinnen und Schüler eines Bundeslandes in bestimmten Fächern zur gleichen Zeit – zum Abschluss eines Bildungsganges – die gleichen schriftlichen, zentral vorgegebenen Aufgabenstellungen“ bearbeiten (Kühn 2010, S. 41). Sie dienen im Unterschied zu den Vergleichsarbeiten primär der Zertifizierung der Schülerleistung (Maag Merki 2010).

Bezieht man die durch die verschiedenen Evaluationsverfahren gewonnenen Informationen auf die vier beschriebenen Faktoren der Schulentwicklung (Schulprogramme, Aufgabenteilung, Personalentwicklung, Schulkultur), so wird deutlich, dass die Schulinspektion auf Basis der jeweiligen Operationalisierungen und Schwerpunktsetzungen der Schulqualitätsrahmen der einzelnen Bundesländer Informationen zu allen Faktoren erfasst. Interne Evaluationen können ebenfalls alle Merkmale betreffen. Im Idealfall sollte sich die interne Evaluation auf das Schulprogramm beziehen. Der Bereich des Unterrichts ist im Rahmen der Inspektion ein sehr bedeutsamer Bestandteil, bei interner Evaluation kann er es ebenso sein. Im Vergleich zur Inspektion und zur internen Evaluation erfassen VERA und zentrale Prüfungen ausschließlich Informationen zur Schülerleistung und betreffen so indirekt Bestandteile des Modells (Programme und Curricula sowie Lern-/Leistungsziele), über die ein Bezug zur Leistung hergestellt werden kann. Zur Veranschaulichung werden die Funktionen und die Ausgestaltung der unterschiedlichen Verfahren in Tabelle 1 gegenübergestellt. Zur Systematisierung werden auf Basis des oben skizzierten Modells von Visscher und Coe (2003) die vorgeschlagenen potenziellen verfahrensbezogenen Einflussfaktoren herangezogen.

Tabelle 1: Vergleich der Evaluationsverfahren am Beispiel des Landes Brandenburg				
	VERA 8	Zentrale Prüfungen: Mittlerer Schulabschluss (MSA) und Zentralabitur (ZA)	Schulinspektion	Interne Evaluation am Beispiel Selbstevaluationsportal (Gärtner 2010)
A) Schulsystemkontext				
Primäre Funktion	Schul- und Unterrichtsentwicklung	Zertifizierung und Selektion der Schülerschaft	Schul- und Unterrichtsentwicklung	Schul- und Unterrichtsentwicklung
Rechenschaftspflicht	<i>Low-stakes</i> : Schulen, Lehrkräfte, bildungspolitisch Verantwortliche (Pant 2013)	<i>High-stakes</i> : Schülerschaft, <i>low-stakes</i> : Schulen, Lehrkräfte, bildungspolitisch Verantwortliche	<i>Low-stakes</i> : Schulen, bildungspolitisch Verantwortliche	<i>Low-stakes</i> : Schulleitungen, Lehrkräfte (Schulaufsicht)
B) Eigenschaften des Testsystems				
Evaluationsgegenstand	Bildungsstandards; Fächer: Deutsch, Mathematik, 1. Fremdsprache	Rahmenlehrpläne/ orientiert an Bildungsstandards (bisher nur MSA); Fächer: Deutsch, Mathematik, Englisch, weitere durch Schüler/innen gewählte Fächer	Erfüllung der Kriterien im Orientierungsrahmen Schulqualität (MBJS 2008)	Selbst-/Fremdeinschätzung der Qualität des Schulmanagements und der Unterrichtsqualität; eigene Inhalte
Turnus	Jährlich	Jährlich	Ca. alle 5 Jahre	Entscheidung durch Schule
Datenerhebung	Standardisierte Leistungstests (Vollerhebung in Klasse 8; Testadministration & Auswertung durch Lehrkräfte); z. T. Verpflichtung zur Teilnahme (nicht alle Fächer)	Standardisierte Leistungstests (Vollerhebung zum Ende der Lernperiode [MSA: 10. Klasse; ZA: 12./13.Klasse]; Testadministration & Auswertung durch Lehrkräfte); Verpflichtung zur Teilnahme	Beobachtung, standardisierte Fragebögen, Interviews, Dokumentenanalyse; Verpflichtung zur Teilnahme	Online-Befragung mit standardisierten Fragebögen; freiwillig
Ausgestaltung der Informationen	Schul-, Klassen- und Individualrückmeldung auf Aufgaben- und Kompetenzebene (Lösungshäufigkeiten, Kompetenzstufen)	Nur durch Schulen selbst eingegebene Prüfungsergebnisse in Form von Noten	Ausführlicher schriftlicher Bericht über die erhobenen Merkmale, Stärken-Schwächen-Profil	Ergebnisbericht, inhaltliche Erläuterungen, Tipps zur Weiterarbeit; Profilgrafik mit Vergleich der Selbst- und Fremdeinschätzung

Tabelle 1 (Fortsetzung)				
	VERA 8	Zentrale Prüfungen	Schulinspektion	Interne Evaluation
Vergleichsmöglichkeiten	Sozial (Schulform als Vergleichsgruppe), kriterial (Kompetenzstufen)	Sozial (Bundesland als Vergleichsgruppe), kriterial (bestanden/nicht bestanden)	Sozial (Vergleich mit Werten in Landesberichten), kriterial (erfasste Bereiche Qualitätsrahmen)	Sozial (Schulform, Fach und Jahrgangsstufe; bei genügend großer Vergleichsgruppe), kriterial (erfasste Qualitätsaspekte)
C) Merkmale der Implementation				
Unterstützungssystem	Aufgabendatenbank, didaktische Handreichungen, Informationen zur Weiterarbeit, Informationsveranstaltungen		Nicht flächendeckend vorgesehen	Informationen und Anleitungen auf Webseite
Quelle: in Anlehnung an Wurster & Richter (eingereicht); Wurster, Richter, Schliesing & Pant 2013				

1.3.2 Forschungsstand zur Verwendung von Evaluationen für datenbasierte Schul- und Unterrichtsentwicklung

Die Forschung im deutschsprachigen Raum zum Umgang schulischer Akteure mit Daten aus den neu eingeführten Evaluationsverfahren widmete sich bislang vor allem der Schulinspektion und VERA, während zur internen Evaluation wenig bekannt ist (Thiel & Thillmann 2012). Niederländische und belgische Studien zur internen *Evaluation* (Schildkamp, Vanhoof, van Petegem & Visscher 2012; Schildkamp & Visscher 2009) zeigen, dass in etwa einem Viertel der Schulen die Ergebnisse kommuniziert und für Entwicklungsmaßnahmen verwendet werden. Maßnahmen werden von der Hälfte der Schulleitungen und einem Viertel der Lehrkräfte aus den Ergebnissen abgeleitet. Einen Einfluss auf die Nutzung hat eine positive Einschätzung des Verfahrens hinsichtlich der Nützlichkeit.

Zur Nutzung von Ergebnissen aus *zentralen Abschlussprüfungen* liegen ebenfalls keine deutschsprachigen Studien vor (Klein, Krüger, Kühn & van Ackeren 2014). Dennoch kann davon ausgegangen werden, dass die Ergebnisse zentraler Abschlussprüfungen trotz ihrer primären Zertifizierungsfunktion in Schulen zum Teil auch zur Ableitung von Entwicklungsaktivitäten verwendet werden, wie die internationale Forschung zeigt (Klein 2013; Schildkamp, Rekers-Mombarg & Harms 2012).

Die Ergebnisse aus *Vergleichsarbeiten* (zusammenfassend: Maier & Kuper 2012; Nachtigall & Hellrung 2013; international: Hellrung & Hartig 2013) werden in Schulen ebenfalls rezipiert, insbesondere durch die Lehrkräfte. Insgesamt ist die Wahrnehmung des Verfahrens überwiegend positiv, jedoch nicht bei allen Befragten. Die Schulleitungen stehen den Vergleichsarbeiten positiver gegenüber als die Lehrkräfte. Aus VERA-Ergebnissen abgeleitete Maßnahmen beziehen sich vornehmlich auf den

Unterricht einzelner Lehrkräfte und auf die Bereiche Aufgaben und Leistungsbeurteilungen sowie Wiederholung von Bereichen, in denen Schwächen festgestellt wurden. Umfassende Maßnahmen der Schul- und Unterrichtsentwicklung sind selten, ebenso die Veränderung didaktischer Strategien. Die wahrgenommene Nützlichkeit erweist sich als bedeutsamer Prädiktor für die Ergebnisnutzung. Weiterhin ist auch eine kriteriale Bezugsnormorientierung, z. B. an den Kompetenzmodellen der Bildungsstandards, für die Nutzung von VERA-Ergebnissen förderlich (Groß Ophoff 2013).

Forschungsbefunde aus dem deutschsprachigen Raum zur *Schulinspektion* (zusammenfassend: Dederling 2012; Husfeldt 2010) zeigen, dass Inspektionsergebnisse in der Mehrheit der Schulen unter Einbezug verschiedener Akteure rezipiert werden. Sie werden vor allem im Kollegium und in entsprechenden Konferenzen kommuniziert. Neben der nahezu flächendeckenden innerschulischen Rezeption der Ergebnisse werden in der Mehrheit der Schulen Entwicklungsmaßnahmen, z. B. in den Bereichen Schulprogrammarbeit, interne Evaluation und Unterrichtsentwicklung, abgeleitet. Die Einstellung diesem Verfahren gegenüber ist weitgehend positiv, wobei die Schulleitungen die Inspektion positiver einschätzen als die Lehrkräfte.

Einige Arbeiten stellen heraus, dass sich Schulen in der Reaktion auf Evaluationsverfahren (Stamm 2003) und die Verarbeitung von Schulinspektionen (Wurster & Gärtner 2013), VERA (Groß Ophoff, Hosenfeld & Koch 2007) und zentralen Abschlussprüfungen (Maag Merki, Klieme & Holmeier 2008) unterscheiden und verdeutlichen, dass Faktoren auf Einzelschulebene eine wichtige Rolle bei der Evaluationsnutzung spielen.

Im Verbundprojekt „Schulen als Steuerungsakteure im Bildungssystem“ entstanden die ersten Studien, die die verschiedenen Evaluationsverfahren im Gesamtzusammenhang untersuchen.

2. Das Projekt Schulen als Steuerungsakteure im Bildungssystem

2.1 Fragestellungen und Vorgehen

Ziel des Verbundvorhabens „Schulen als Steuerungsakteure im Bildungssystem“ war es, die Praxis interner Evaluation sowie die Nutzung von Daten aus externer Evaluation auf der Schulebene zu beschreiben, Bedingungen und Barrieren der Nutzung zu identifizieren sowie ein optimiertes Rückmeldeformat für Daten aus Vergleichsarbeiten zu entwickeln. Im Fokus stand dabei nicht nur die Analyse aller im Schulsystem relevanten Evaluationsformen, sondern auch eine multiperspektivische Betrachtung der verschiedenen involvierten Akteure der Schule: die Schul- und Fachkonferenzleitungen sowie Lehrkräfte. In mehreren Teilstudien wurden folgende Kernfragen untersucht:

- wie interne Evaluation, Vergleichsarbeitsdaten und Daten aus zentralen Abschlussprüfungen sowie der externen Evaluation in den einzelnen Schulen gehandhabt werden,

- für welche Maßnahmen die unterschiedlichen Daten vorwiegend genutzt werden,
- welche individuellen und schulischen (organisatorischen) Bedingungen die Nutzung der Daten für die Personal- und Schulprogrammentwicklung beeinflussen und
- wie eine evaluative Rückmeldung gestaltet werden muss, die die Nutzungswahrscheinlichkeit von Daten aus Vergleichsarbeiten erhöht.

Es werden ausgewählte Resultate des Projektes präsentiert.

2.2 Datengrundlage

Als Grundlage der ersten und zweiten Teilstudie wurde eine Online-Befragung von Schul- und Fachkonferenzleitungen (der Fächer Deutsch und Mathematik) aller Berliner und Brandenburger allgemeinbildenden Schulen durchgeführt. Während die Schulleitungen als Experten für die Abläufe auf der Schulebene befragt wurden, wurden die Leitungen der Fachkonferenzen als Experten für die Abläufe auf der Fach- und Unterrichtsebene befragt. Mit der Befragung konnte insgesamt auf der Schulebene ein Rücklauf von 27 Prozent ($N = 327$) erreicht werden (Tabelle 2). Auf der Ebene der Fachkonferenzen beteiligten sich insgesamt $N = 337$ Fachkonferenzleitungen der Fächer Deutsch ($N = 155$), Mathematik ($N = 173$) und Andere ($N = 9$).

	Brandenburg		Berlin		Insgesamt	
	N	in %	N	in %	N	in %
Gesamt	216	66	111	34	327	100
Grundschule	141	65	63	57	204	62
Gymnasium	26	12	22	20	48	15
Andere weiterf. Schulen	49	23	26	23	75	23

Ergänzend dazu erfolgte in einer Substichprobe der Teilnehmerschulen eine Befragung von Lehrkräften, um differenzierte Informationen über die individuelle Nutzung von Evaluationsergebnissen für die Unterrichtsentwicklung zu bekommen. Die Lehrkräftestichprobe umfasst insgesamt $N = 51$ Schulen (Berlin: $N = 14$ Grundschulen; Brandenburg: $N = 6$ Grundschulen, $N = 31$ weiterführende Schulen). Insgesamt haben $N = 662$ Lehrkräfte an der Befragung teilgenommen. Durchschnittlich haben 42 % der Lehrkräfte einer Schule bei einer durchschnittlichen Anzahl von $N = 31$ Lehrkräften pro Schule die Fragebögen beantwortet.

Die beschriebenen Stichproben liegen den meisten im Folgenden präsentierten Ergebnissen zugrunde. Für bestimmte Detailanalysen wurden weitere Erhebungen durchgeführt, die im Rahmen der jeweiligen Ergebnisdarstellung dargestellt werden.

3. Ergebnisse

Ausgehend vom theoretischen Modell der Schulentwicklung (Thiel 2008a, 2008b) werden die Nutzung von Evaluationsergebnissen für die Unterrichtsentwicklung, Maßnahmen der Lernförderung und verschiedene Maßnahmen der Schulentwicklung (Entwicklung von Zielen und Strategien, Schulmanagement, Qualitätsentwicklung, Personalentwicklung, Strukturentwicklung, Schulprogrammentwicklung) vorgestellt. Weiterhin werden die Akzeptanz der Evaluationen und die schulinterne Kommunikation der Ergebnisse betrachtet und der Einfluss dieser Faktoren auf die Nutzung analysiert. Ergänzend dazu werden Ergebnisse für die vor dem Hintergrund der schulischen Selbststeuerung bedeutsamen Bereiche der Personalentwicklung und der Schulprogramme beschrieben. Abschließend werden Ergebnisse einer Studie zur Optimierung von Rückmeldeformaten vorgestellt.

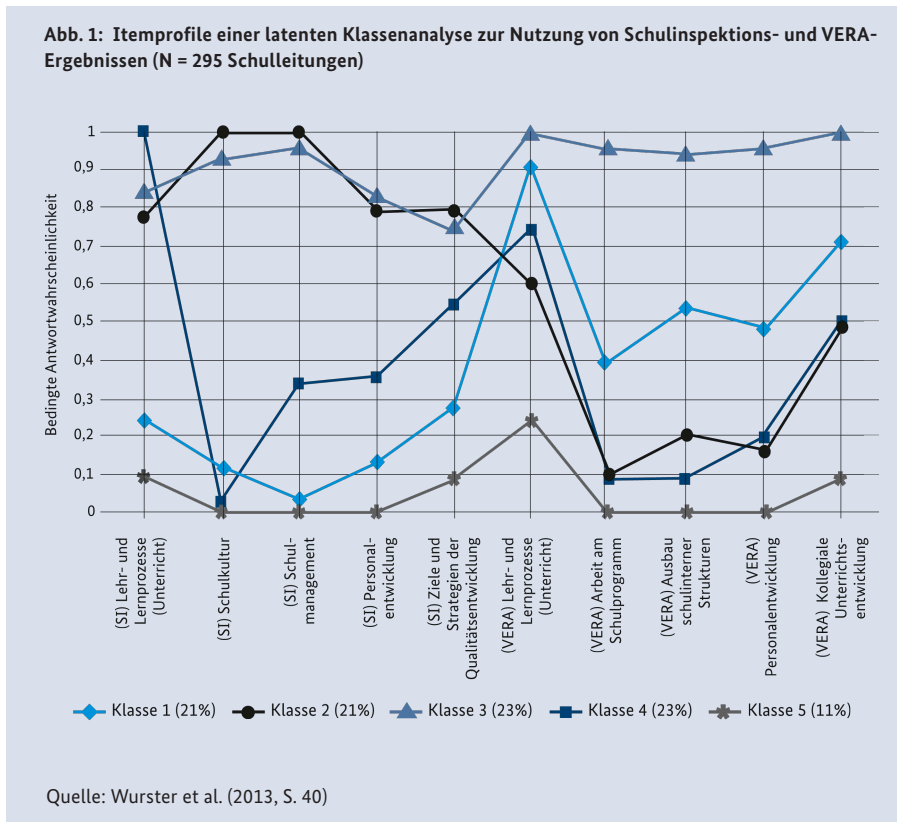
3.1 Nutzung von Ergebnissen aus Schulinspektionen, interner Evaluation, Vergleichsarbeiten und zentralen Prüfungen für die Schul- und Unterrichtsentwicklung im Vergleich

Bislang wurden die verschiedenen Evaluationsinstrumente weitgehend getrennt voneinander empirisch analysiert, ohne diese miteinander in Beziehung zu setzen. Mittels deskriptiver Analysen konnte zunächst ein differenziertes Bild hinsichtlich der Nutzung verschiedener Evaluationsinstrumente aus Sicht der *Schulleitungen* gewonnen werden (Tabelle 3). Eine Nutzung von Schulinspektions- und VERA-Ergebnissen findet vor allem im Bereich des Unterrichts statt, wobei die prominenteste Maßnahme die Binnendifferenzierung ist. Es wird ebenfalls deutlich, dass es eine große Bandbreite an abgeleiteten Maßnahmen gibt, wie die offenen Antworten zeigen. Doppelnennungen in verschiedenen Bereichen zeigen, dass die Befragten manche Maßnahmen (Fortbildungen, schulinternes Curriculum) auch nach den Inhalten der Maßnahme zuordnen. So wird z. B. Fortbildung von einem Teil als Personalentwicklungsmaßnahme verstanden, von einem anderen Teil über die Inhalte der Fortbildung dem Bereich Unterricht zugeordnet. In etwa der Hälfte der Schulen werden Inspektionsergebnisse auch für unterschiedliche Maßnahmen der Schulentwicklung verwendet. VERA-Ergebnisse werden nach den Angaben der Schulleitungen in ca. 30 % bis 40 % der Schulen für diese Zwecke eingesetzt.

Tabelle 3: Nutzung von VERA- und Schulinspektionsergebnissen nach Angabe der Schulleitungen

Bereich	%	Beispielnennungen offene Antworten
Schulinspektion		
Lehr- und Lernprozesse (Unterricht)	66	Binnendifferenzierung, Arbeit am schulinternen Curriculum, Fortbildungen
Ziele und Strategien Qualitätsentwicklung	54	Arbeit am schulinternen Curriculum und Schulprogramm, Einführung Selbstevaluation
Schulmanagement	52	Hospitation vom Schulleiter, bessere Aufgabenverteilung
Personalentwicklung	47	Fortbildungen, kollegiale Hospitationen, Personalgespräche
Schulkultur	47	Neue Strukturen bei Schüler- und Elternbeteiligung, Kooperationen, Etablierung schulischer Höhepunkte
Vergleichsarbeiten		
Lehr- und Lernprozesse (Unterricht)	75	Binnendifferenzierung, Festlegung von Lernschwerpunkten
Kollegiale Unterrichtsentwicklung	61	Unterrichtsbesuche, fächerverbindendes Arbeiten
Personalentwicklung	40	Fortbildungen, Hospitationen
Ausbau schulinterner Strukturen	40	Intensivierung der Fachkonferenzarbeit, Arbeitsgruppe zur Erarbeitung des Übungsmaterials
Arbeit am Schulprogramm	34	Überarbeitung schulinterner Ziele und Maßnahmen
Anmerkungen: Antwortmöglichkeiten bei der Erfragung der Bereiche der Nutzung dichotom (ja/nein); die Befragten konnten den einzelnen Bereichen ergänzend offene Antworten zuordnen. Quelle: Wurster et al. (2013, S. 37)		

Neben der Betrachtung einzelner Aktivitäten wurde auch die verfahrensübergreifende Kombination verschiedener Evaluationsverfahren analysiert. Mittels Latent-Class-Analysen wurden auf Basis der in Tabelle 3 dargestellten dichotom erfragten *Nutzungsbereiche* fünf unterschiedliche Nutzungskonfigurationen ermittelt. Sie unterscheiden sich hinsichtlich der bevorzugten Evaluationsverfahren sowie in der Intensität und Art der Nutzung (Wurster et al. 2013). Die Nutzungsprofile in Abbildung 1 zeigen einerseits Schulen, die VERA- und Inspektionsergebnisse nutzen, wobei 23 % der Schulen vor allem datenbasierte Unterrichtsentwicklung betreiben (Klasse 4) und weitere 23 % der Schulen in allen Bereichen eine umfangreiche Nutzung aufweisen (Klasse 3). Daneben gibt es auch Schulen, die nur ein Verfahren (Klasse 1 und 2 jeweils 21 %) für Entwicklungsmaßnahmen nutzen. Nur eine kleine Minderheit der befragten Schulleitungen gibt an, keine Ergebnisse aus VERA oder der Inspektion zu nutzen (Klasse 5, 11 %).



Innerhalb der Schule spielt die Fachkonferenz eine Schlüsselrolle zur Verarbeitung der Ergebnisse und zur Einleitung von Entwicklungsmaßnahmen auf der Basis von Vergleichsarbeiten (VERA) und von zentralen Abschlussprüfungen wie dem mittleren Schulabschluss (MSA). Eine Befragung der *Fachkonferenzleitungen* zeigt, dass diese VERA-Ergebnisse (54 %) deutlich seltener für die Entwicklung des eigenen Unterrichts nutzen als MSA-Ergebnisse (87 %). Dies lässt sich auch in der kollektiven Ergebnisnutzung in Fachkonferenzen feststellen (Tabelle 4). MSA-Ergebnisse sind häufiger Ausgangspunkt für Unterrichtsentwicklungsaktivitäten als VERA-Ergebnisse. Unterrichtsentwicklung manifestiert sich am häufigsten in der Entwicklung oder zielgerichteten Veränderung von Aufgaben. Weniger häufig werden Materialien und Methoden weiterentwickelt oder Fortbildungen konzipiert (Wurster & Richter, eingereicht).

Tabelle 4: Nutzung von VERA- und MSA-Ergebnissen für die Unterrichtsentwicklung

	VERA 8	Mittlerer Schulabschluss
Entwicklung von Aufgaben	51 % _a	86 % _b
Einführung von Unterrichtsmethoden	29 % _a	45 % _b
Entwicklung neuer Materialien	38 % _a	62 % _b
Entwicklung von Maßnahmen zur Einzelförderung	43 % _a	51 % _a
Konzeption von Fortbildungsmaßnahmen	16 % _a	35 % _b

Anmerkungen: N = 93; Antwortmöglichkeiten: dichotom ja/nein; .a/b: gleicher Buchstabe in einer Zeile bedeutet, dass es keine signifikanten Unterschiede zwischen beiden Verfahren gibt, ungleiche Buchstaben zeigen bedeutsame Unterschiede ($p < .05$; $d > .30$) an.
Quelle: Wurster und Richter (eingereicht)

Ergänzend zu den Angaben der Schul- und Fachkonferenzleitungen wurden Lehrkräfte gefragt, für welche Zwecke sie VERA-Ergebnisse verwenden (Tabelle 5). Sie werden von der überwiegenden Mehrheit der Lehrkräfte zur Identifikation von Unterrichtsinhalten genutzt, die weiter geübt werden müssen. Außerdem werden Stärken und Schwächen einzelner Schülerinnen und Schüler identifiziert und bei Bedarf auch die Leistungsanforderungen angepasst. Vergleichsweise seltener werden neue Unterrichtsmaterialien erstellt.

Tabelle 5: Angaben von Lehrkräften zur Nutzung von VERA-Ergebnissen

Um festzustellen, welche Aufgaben verstärkt geübt werden müssen.	71 %
Um festzustellen, welche Inhalte ich ausführlicher erklären muss.	68 %
Um zu prüfen, ob meine Leistungsanforderungen zu hoch oder zu niedrig sind.	64 %
Zur Identifikation von Stärken und Schwächen einzelner Schüler/-innen.	62 %
Zur Beratung von Eltern hinsichtlich des Leistungsstands ihres Kindes.	59 %
Um neue Aufgaben auszusuchen bzw. Aufgaben weiterzuentwickeln.	55 %
Für die Planung von Maßnahmen zur gezielten Förderung einzelner Schüler/-innen.	50 %
Um neue Lernmaterialien auszuwählen bzw. zu entwickeln.	36 %

Anmerkungen: $N_{\min/\max} = 317/321$; Antwortmöglichkeiten: dichotom ja/nein

Vergleicht man die Nutzung von Evaluationsdaten durch die verschiedenen schulischen Akteure am Beispiel der Vergleichsarbeiten, so zeigt sich, dass in Schulen die Daten eher selten (30 % bis 40 %) für strategische Entscheidungen im Bereich der Personalentwicklung und Schulprogramme verwendet werden. Auf Ebene der

Fachkonferenzen zeigte sich ebenfalls, dass ausgehend von VERA-Ergebnissen eher unterrichtsnahe Maßnahmen, wie die Entwicklung von Aufgaben oder die Einzelförderung, durchgeführt werden als umfangreichere Vorhaben, wie die Änderung von Methoden oder Fortbildungen. Lehrkräfte nutzen VERA-Ergebnisse deutlich stärker, vor allem für die Diagnose und Entwicklung des eigenen Unterrichts, wie z. B. die Identifikation von Aufgaben, die geübt werden müssen.

Bevor Entwicklungsmaßnahmen umgesetzt werden können, müssen zunächst die Evaluationsergebnisse ausgewertet werden. Bezüglich der *schulinternen Rezeption und Reflexion der Ergebnisse* gaben die befragten Schulleitungen an, dass Ergebnisse aus der Inspektion, der internen Evaluation und VERA in der in Schul- und Lehrerkonferenzen institutionalisierten Kommunikation in der Regel mindestens einmalig besprochen werden. Unterschiede zwischen den Verfahren bestehen vor allem darin, dass Schulinspektionsergebnisse in der Mehrheit der Schulen mehrfach thematisiert werden, während dies bei internen Evaluationen häufig nur einmal der Fall ist (Wurster et al. 2013). Eine Befragung Brandenburger Fachkonferenzleitungen zeigt, dass Ergebnisse aus VERA und dem mittleren Schulabschluss schulweit kommuniziert werden. Die befragten Fachkonferenzleitungen selbst setzen sich nahezu alle mit den Ergebnissen aus VERA (91 %) und dem MSA (96 %) auseinander (Wurster & Richter, eingereicht).

3.2 Einstellungen von Schulleitungen, Lehrkräften und Fachkonferenzleitungen gegenüber den Evaluationsverfahren

Die befragten *Schulleitungen* nehmen im Durchschnitt die Inspektion, die interne Evaluation und VERA insgesamt eher positiv wahr (Tabelle 6). Die Nützlichkeit sowie die Diagnosegüte werden als eher hoch bewertet, während der Aufwand als eher gering angesehen wird. Ein Vergleich der Verfahren zeigt allerdings Unterschiede in der wahrgenommenen Nützlichkeit und dem Aufwand. Der internen Evaluation wird von den befragten Schulleitungen ein vergleichsweise hoher Stellenwert für die Schul- und Unterrichtsentwicklung beigemessen (Wurster et al. 2013). Der Aufwand wird im Vergleich jedoch auch als höher eingeschätzt. VERA und die Schulinspektion unterscheiden sich in der Einschätzung der Nützlichkeit und des Aufwandes aus Sicht der Schulleitungen nicht bedeutsam.

Vergleicht man die beiden auf Schülerleistungen bezogenen Verfahren, so zeigt sich, dass die Fachkonferenzleitungen hinsichtlich der wahrgenommenen Nützlichkeit, des Aufwands und der Diagnosegüte dem mittleren Schulabschluss gegenüber deutlich positiver eingestellt sind (Wurster & Richter, eingereicht). Die befragten Lehrkräfte bestätigen, dass VERA als eher weniger nützlich und als eher zu aufwendig eingeschätzt wird. Insgesamt zeigt sich, dass die Schulleitungen VERA gegenüber deutlich positiver eingestellt sind als die Fachkonferenzleitungen und diese wiederum positiver als die Lehrkräfte. Der mit dem Verfahren verbundene Aufwand wird prinzipiell ähnlich eingeschätzt. Je intensiver die Befragten an der Durchführung beteiligt ist, desto schlechter wird das Aufwand-Nutzen-Verhältnis wahrgenommen.

Tabelle 6: Einschätzung verschiedener Evaluationsverfahren durch Schulleitungen, Fachkonferenzleitungen und Lehrkräfte im Vergleich

	Schulinspektion M (SD)	Interne Evaluation M (SD)	VERA M (SD)	Mittlerer Schulabschluss M (SD)
Schulleitungen (N = 235–244)				
Aufwand	2.99 _a (1.46)	3.64 _b (1.39)	3.08 _a (1.47)	
Nützlichkeit	3.55 _a (1.33)	4.09 _b (1.21)	3.45 _a (1.39)	
Diagnosegüte	4.29 _b (1.21)	4.37 _b (1.26)	3.76 _a (1.34)	
Fachkonferenzleitungen (N = 125-127)				
Aufwand			4.38 _a (1.40)	3.12 _b (1.40)
Nützlichkeit			2.88 _a (1.29)	4.00 _b (1.40)
Diagnosegüte			3.50 _a (1.32)	4.21 _b (1.22)
Lehrkräfte (N = 226)				
Aufwand			4.27 (1.51)	
Nützlichkeit			2.95 (1.38)	
Anmerkungen: Antwortmöglichkeiten: sechsstufig 1 = „Stimme überhaupt nicht zu“ bis 6 = „Stimme vollständig zu“; .a/b: gleicher Buchstabe in einer Zeile bedeutet, dass es keine signifikanten Unterschiede zwischen beiden Mittelwerten gibt, ungleiche Buchstaben zeigen signifikante Unterschiede ($p < .05$) an (Schulleitungen: Unterschiede mit ANOVA ermittelt; Fachkonferenzleitungen: Unterschiede alle mind. $d > .60$ und mit t-Tests ermittelt). Quelle: Wurster et al. (2013, S. 35), Wurster und Richter (eingereicht), weitere Berechnungen				

3.3 Evaluationsverfahrensbezogene Einflussfaktoren auf die Nutzung am Beispiel der Vergleichsarbeiten

Zur Analyse förderlicher und hinderlicher Bedingungen für die Nutzung von Evaluationsergebnissen werden zunächst auf das Verfahren bezogene Faktoren betrachtet. VERA dient als Beispiel, da hier Daten auf verschiedenen Ebenen für die Akteursgruppen Schul- und Fachkonferenzleitungen sowie Lehrkräfte vorliegen und somit ein Vergleich möglich ist. Mittels Regressionsanalysen wird die Nutzung für verschiedene Schul- und Unterrichtsentwicklungsmaßnahmen analysiert (Tabelle 7). Es werden zum einen mit *Einstellungen* dem Verfahren gegenüber (Nützlichkeit, Diagnosegüte) und zum anderen mit der *Kommunikation* über die Ergebnisse primär verfahrensnaher Einflussfaktoren untersucht. Auf Ebene der individuellen Lehrkräfte werden Nutzungsaktivitäten untersucht und weitere Einflussfaktoren, wie die individuellen Bedingungen, die Arbeitsbelastung und Selbstwirksamkeit, aber auch die Rolle der Schulleitung, ergänzt.

Die Ergebnisse zeigen, dass über alle drei Ebenen bzw. schulischen Akteursgruppen hinweg die wahrgenommene Nützlichkeit der relevanteste Prädiktor für alle erfassten Entwicklungsaktivitäten ist. Die Kommunikation über die VERA-Ergebnisse ist ebenfalls für die Ebene der Lehrkräfte und fast alle Aktivitäten auf Fachebene sowie auch einige Aktivitäten auf Schulebene ein wichtiger Prädiktor. Weiterhin zeigt sich, dass die Schulleitung für die Nutzung von VERA-Ergebnissen durch einzelne Lehrkräfte eine wichtige Rolle spielt. Die Arbeitsbelastung und das Selbstwirksamkeitserleben der Lehrkräfte haben hingegen keinen signifikanten Einfluss.

Tabelle 7: Regressionsanalysen zur Vorhersage der VERA-Nutzung						
Schulleitungen (logistische Regressionsanalyse; dargestellt sind die adjustierten Odds Ratios; N = 240; Nagelkerke-R ² = .13 bis .20)						
	Lehr- & Lernprozesse	Schulprogramm	Ausbau Strukturen	Personalentwicklung	Unterrichtsentw.	
Nützlichkeit	1.48*	1.52**	1.33*	1.52**	1.39*	
Diagnosegüte	0.98	1.22	1.52**	1.05	1.11	
Kommunikation Konferenz einmalig	1.83	0.61	0.63	1.84	1.94	
Kommunikation Konferenz mehrfach	3.94*	1.29	0.76	3.05	4.21*	
Fachkonferenzleitungen (logistische Regressionsanalyse; dargestellt sind die adjustierten Odds Ratios; N = 322 bis 323; Nagelkerke-R ² = .13 bis .39)						
	Aufgaben	Methoden	Materialien	Einzelförderung	Fortbildung	Elternberatung
Nützlichkeit	2.14***	1.88***	1.83***	1.66***	1.42*	1.83**
Diagnosegüte	1.11	1.09	1.06	1.12	1.35	0.94
Kommunikation	1.76***	1.17	1.32*	1.79***	1.11	1.97***
Lehrkräfte (lineare Regressionsanalyse; dargestellt sind die standardisierten Regressionskoeffizienten; N = 180; R ² = .55)						
	VERA-Nutzung					
Data-wise leadership	0.14*					
Nützlichkeit	0.64***					
Kollaborative Ergebnisauswertung	0.13*					
Arbeitsbelastung	0.08					
Selbstwirksamkeit	0.05					
Anmerkungen: Informationen zu den Instrumenten: Schulleitungen: Wurster et al. (2013); Fachkonferenzleitungen: Wurster und Richter (eingereicht); Lehrkräfte: Bach, Wurster, Thillmann, Pant und Thiel (2014).						

3.4 Vergleichsarbeiten und schulische Personalentwicklung

Die Nutzung von VERA-Daten für eine systematische Personalentwicklung soll in der Logik des neuen Steuerungskonzepts sowohl durch Schulleitungen als auch durch Fachkonferenzen erfolgen. Die Rückmeldungen von VERA bieten den Schulleitungen die Möglichkeit, die Daten in Mitarbeitergesprächen, zur Vor- und Nachbereitung von Unterrichtsbesuchen und für eine systematische Fortbildungsplanung zu nutzen. Fachkonferenzen tragen ebenfalls die Verantwortung für eine systematische, datenbasierte Personalentwicklung. Sie sind für die Planung der fachspezifischen Fortbildungen zuständig und gelten neben den einzelnen Lehrkräften zugleich als wichtigster Adressat von VERA-Rückmeldungen (KMK 2010).

Auf der Basis einer Substichprobe von Schulleitungen ($N = 163$) und Lehrkräften aus Brandenburger weiterführenden Schulen ($N = 175$) wurde untersucht, inwieweit die VERA-Ergebnisse für gezielte Maßnahmen der Personalentwicklung genutzt werden und welche Faktoren eine datenbasierte Personalentwicklung beeinflussen (Bach et al. 2014).

Unseren Ergebnissen zufolge werden die VERA-Rückmeldungen von 48 % der Schulleitungen für Maßnahmen der Personalentwicklung genutzt. Dies entspricht auch in etwa dem Anteil in der Gesamtstichprobe (vgl. Abschnitt 3.1). Auch die Nutzung von VERA-Ergebnissen zur Fortbildungsplanung in den Fachkonferenzen ist nach Angaben der Lehrkräfte nur wenig verbreitet. Der Mittelwert liegt mit 2.41 ($SD = 1.40$) deutlich unter dem Skalenmittelwert der sechsstufigen Antwortskala. Für eine systematische Personalentwicklung scheinen VERA-Daten offenbar nur eine untergeordnete Rolle zu spielen, auch wenn in dem von uns untersuchten Bundesland Brandenburg den Schulen im Hinblick auf die Personalentwicklung hohe Gestaltungsspielräume eingeräumt (vbw 2010) und entsprechende Steuerungsimpulse gesetzt wurden (z. B. MBJS 2012).

Auf der Schulleitungsebene wurde untersucht, ob die Organisation (Vorhandensein einer erweiterten Schulleitung), die Qualifikation, das Alter und die Einstellung (im Sinne der subjektiv wahrgenommenen Nützlichkeit von VERA) von Schulleitungen mit der Verwendung der Daten für Personalentwicklungsmaßnahmen in Zusammenhang stehen. Unsere Analysen zeigen, dass ausschließlich die Einschätzung der Nützlichkeit des Evaluationsinstruments damit zusammenhängt, ob aus den VERA-Ergebnissen Personalentwicklungsmaßnahmen abgeleitet werden (Tabelle 8).

Zur Vorhersage der Nutzung der VERA-Ergebnisse zur fachkonferenzbezogenen Fortbildungsplanung wurde der Einfluss theoretisch relevanter Variablen untersucht.² Die gemeinsame Auswertung der Ergebnisse erweist sich als stärkster Prädiktor: Je intensiver in den Fachkonferenzen ein Austausch über die VERA-Rückmeldungen stattfindet, desto eher werden die Ergebnisse in den Fachkonferenzen zur Fortbildungsplanung genutzt. Zudem steht auch das Verhalten der Schulleitung mit datenbasierter Personalentwicklung in den Fachkonferenzen in Zusammenhang: Je stärker eine Schulleitung den Umgang mit den Daten insgesamt anregt und unterstützt (sog. *data-wise leadership*), desto eher findet nach unseren Ergebnissen in den Fachkonferenzen eine datenbasierte Fortbildungsplanung statt.

2 Details zu Ergebnissen und Skalen finden sich bei Bach et al. (2014).

Tabelle 8: Regressionsanalysen zur Vorhersage der VERA-Nutzung für die Personalentwicklung	
Schulleitungen (logistische Regressionsanalyse; dargestellt sind die adjustierten Odds Ratios; N = 163; Nagelkerke-R ² = .21)	
Nützlichkeit	1.82**
Qualifikation der Schulleitung (ja)	1.05
Erweiterte Schulleitung (ja)	1.38
Alter (bis 40 Jahre)	10.29
Alter (41–50 Jahre)	1.11
Alter (51–60 Jahre)	2.43
Fortbildungsplanung in den Fachkonferenzen (auf Schulebene aggregierte Daten; lineare Regressionsanalyse; dargestellt sind die standardisierten Regressionskoeffizienten; N = 29; R ² = .56)	
Data-wise leadership	0.39***
Gemeinsame Auswertung	0.52***
Nützlichkeit	-0.00
Lehrersebstwirksamkeitserwartung	-0.11
Konsens über pädagogische Ziele	-0.27**
Quelle: Bach et al. (2014)	

Die Ergebnisse liefern erste Hinweise auf das Ausmaß und die Bedingungen einer datenbasierten Personalentwicklung. Um die Generalisierbarkeit der Befunde zu erhöhen, müssten weitere Bundesländer sowie auf der Ebene der Lehrkräfte auch Grundschulen mit einbezogen werden. Zudem wäre in Folgestudien zu überprüfen, welchen Einfluss weitere, hier nicht berücksichtigte Faktoren (z. B. Verständlichkeit der Rückmeldungen, schulisches Innovationsklima) auf die datenbasierte Personalentwicklung haben. Um kausale Schlüsse über die untersuchten Zusammenhänge ziehen zu können, sollten in folgenden Analysen entsprechende Forschungsdesigns eingesetzt werden.

3.5 Schulische Selbststeuerung durch interne Evaluation

Als ein weiteres Instrument der Selbststeuerung steht den Schulen die interne Evaluation zur Verfügung, die im Rahmen der Schulprogrammarbeit in den Schulgesetzen bzw. Ausführungsverordnungen der meisten Bundesländer verankert ist. Der Einsatz von interner Evaluation als Instrument schulischer Selbststeuerung setzt voraus, dass Daten nicht nur mit dem Ziel der Beschreibung und Beurteilung erhoben werden, sondern darüber hinaus Daten für eine kontinuierliche Optimierung von

Maßnahmen mit Bezug auf die schuleigenen Entwicklungsziele generiert werden (Thiel & Thillmann 2012).

Ausgehend von einer Analyse von $N = 31$ Schulprogrammen Berliner und Brandenburger Gymnasien, deren Schulleitungen bereits an der Online-Befragung teilgenommen hatten, haben wir die Qualität der in den Schulprogrammen enthaltenen Evaluationsprogramme untersucht. Durch die Verknüpfung mit den Daten der Online-Befragung konnte außerdem analysiert werden, inwiefern die Qualität der Programme von bestimmten organisationalen Voraussetzungen abhängt (Thillmann, Bach, Gerl & Thiel 2013).

Auf der Grundlage theoretischer Vorüberlegungen (Thiel & Thillmann 2012; Holtappels & Müller 2002; SenBWF 2008) wurde ein Analyseraster für das Rating entwickelt.³ Alle verfügbaren Evaluationsprogramme wurden von einer Raterin, ein Drittel dieser Programme von einer zweiten Raterin analysiert. Die Übereinstimmungsrate lag bei durchschnittlich 88 % (Thillmann et al. 2013, S. 211).

Zur Beurteilung der Qualität wurden die Programme im Hinblick auf zwei übergeordnete Kriterien analysiert: Die *methodische Fundierung* bemisst sich daran, ob und wie konkret für die benannten Evaluationsvorhaben Indikatoren zur Überprüfung der Zielerreichung festgelegt und der geplante Einsatz von Evaluationsverfahren beschrieben wurden; das Kriterium der *Kohärenz* bezieht sich auf das Ausmaß der Zusammenhänge zwischen der Bestandsaufnahme, den genannten Entwicklungszielen und -maßnahmen sowie den Evaluationsverfahren. Unsere Ergebnisse zeigen, dass vor allem die Programme der Brandenburger Gymnasien hinsichtlich der beiden Qualitätskriterien deutliche Defizite aufweisen.

Betrachtet man die unterschiedlichen Implementationsbedingungen der beiden Bundesländer, so lässt sich feststellen, dass in Berlin eine Steuerungsstrategie verfolgt wurde, die auf externe Steuerungsimpulse in Form von konkreten Vorgaben, gezielte Unterstützung und regelmäßige Rechenschaftspflicht setzt. Unsere Ergebnisse deuten darauf hin, dass diese klare Steuerungsstrategie die Qualität der Programme positiv beeinflussen kann.

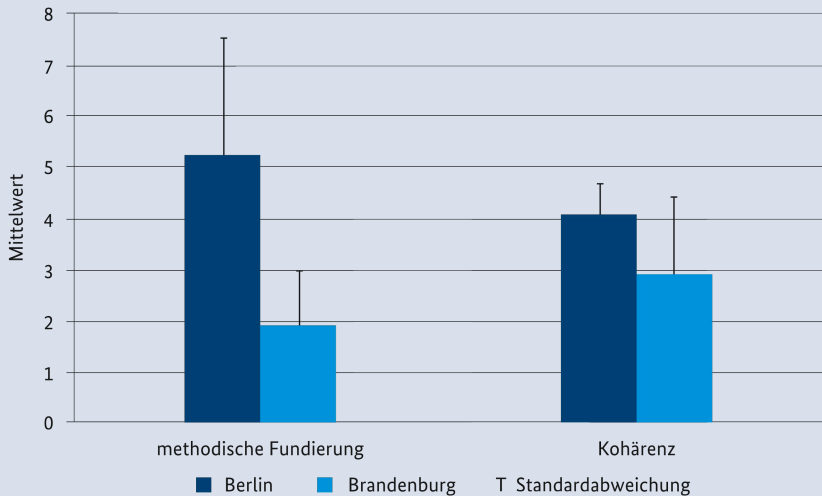
Insgesamt gelingt es in beiden Bundesländern jedoch den wenigsten Schulen, ein Evaluationsprogramm zu entwickeln, das als Instrument der strategischen Planung gelten kann: Eine breite Bestandsaufnahme, die unterschiedliche Bereiche in den Blick nimmt, eine konsequente Beurteilung der Bestandsaufnahme im Hinblick auf einen definierten Soll-Zustand und eine entsprechende Ableitung von Zielen erfolgt nur in drei der untersuchten Programme. Dabei handelt es sich bei allen drei Fällen um Evaluationsprogramme von Berliner Schulen.

Was die Zusammenhänge der Programmqualität mit schulorganisationalen Merkmalen anbelangt, zeigen sich bei der Kontrolle des Bundeslandes deutliche Effekte für die Schulleiterqualifizierung. Die Evaluationsprogramme der Schulen, deren Leitungen eine Fort- oder Weiterbildung für die Schulleitungstätigkeit absolviert haben, weisen tendenziell eine höhere Programmqualität auf.

Auf den Stellenwert einer breiten Beteiligung der Lehrkräfte weisen unsere Befun-

3 Das Analyseraster beinhaltet insgesamt 19 Items (z. B.: „Wurden konkrete Entwicklungsziele aus der Bestandsaufnahme abgeleitet?“, „Wie konkret wurden die geplanten Evaluationsverfahren und -instrumente beschrieben?“); zudem wurden mittels Indizes übergeordnete Kriterien gebildet (Thillmann et al. 2013).

Abb. 2: Übergreifende Qualitätsmerkmale der Evaluationsprogramme (dargestellt sind die Mittelwerte der N = 31 Evaluationsprogramme, Darstellung getrennt nach Bundesländern)



Anmerkungen: Der Index methodische Fundierung kann die Ausprägungen 0 (überhaupt nicht fundiert) bis 8 (ausgesprochen fundiert) annehmen, der Index Kohärenz die Ausprägungen 0 (gar nicht kohärent) bis 6 (durchweg kohärent).

Quelle: Thillmann et al. (2013)

de zur Einrichtung von Steuergruppen hin: Steuergruppen haben eher Einfluss auf die Qualität der Evaluationsprogramme, wenn sie eine koordinierende Funktion übernehmen, die inhaltliche Arbeit aber in den Händen des Gesamtkollegiums bleibt. Auch das Merkmal der funktionalen Spezialisierung korreliert schwach mit der Qualität der Evaluationsprogramme: Schulen, an denen Evaluationsverantwortliche bestimmt wurden, weisen tendenziell etwas bessere Evaluationsprogramme auf.

Insgesamt geben die Ergebnisse der Untersuchung wichtige Hinweise darauf, dass sowohl landesspezifische Implementationsbedingungen als auch innerorganisationale Bedingungen die Qualität von Evaluationsprogrammen und damit auch ihre Eignung als Instrument schulischer Selbststeuerung beeinflussen. Dieses Ergebnis ist insofern von Bedeutung für die Schulentwicklung, weil ein qualitativ hochwertiges Evaluationsprogramm eine zentrale Grundlage für eine datenbasierte schulische Selbststeuerung darstellt.

3.6 Die Gestaltung optimaler Rückmeldeformate für Vergleichsarbeiten

Aus der letzten hier vorgestellten Studie des Forschungsverbunds StABil (Schliesing, in Vorbereitung), die der Frage nachgeht, wie Rückmeldungen aus Vergleichsarbeiten gestaltet sein sollten, damit Lehrkräfte sie verstehen und nutzen können, werden erste Ergebnisse präsentiert. In einem methodenintegrativen Design wurden drei

Teilstudien durchgeführt. Zunächst wurden Gruppendiskussionen durchgeführt, um Optimierungsmöglichkeiten einer Beispielmeldung von Ergebnissen einer Vergleichsarbeit zu erörtern. In einer zweiten Erhebung wurden verschiedene Rückmeldevarianten in einem experimentellen Design auf ihre Verständlichkeit und Nutzbarkeit hin untersucht. Die dritte Erhebung mittels Fokusinterviews, die jedoch noch nicht abschließend analysiert ist, diente der Identifikation von rezeptionsfördernden und -hemmenden Bedingungen.

Die Gruppendiskussionen zur Optimierung der Beispielmeldung wurden an drei Grundschulen in Brandenburg in Gruppen mit zwei bis vier Lehrkräften durchgeführt. Die Ergebnisse der inhaltsanalytischen Auswertung weisen darauf hin, dass sich Lehrkräfte in Ergebnismeldungen sehr konkrete Hinweise zur Weiterarbeit wünschen. So formulieren alle Lehrkräfte den Wunsch nach Lernaufgaben, die sie im Unterricht einsetzen können. Bezüglich der Darstellung der zurückgemeldeten Schülerleistungen präferieren die Lehrkräfte sowohl eine tabellarische Darstellung der einzelnen Schülerergebnisse und die Lösungshäufigkeiten der Klasse für die einzelnen Aufgaben als auch eine grafische Darstellung der Kompetenzstufenverteilung.

Mit einem experimentellen Design wurde untersucht, wie verschiedene VERA-Beispielmeldungen von Lehrkräften verstanden werden und wie nützlich sie für die Weiterarbeit mit den Ergebnissen sind. Vor dem Hintergrund der Ergebnisse der Gruppendiskussionen sowie weiterer theoretischer Erwägungen wurde die VERA-Beispielmeldung weiterentwickelt und in acht verschiedene Rückmeldevarianten überführt. Dabei wurden der Textumfang (mit vs. ohne), die Anzahl an dargebotenen Grafiken (fünf vs. drei) und die Darbietung von Lernaufgaben (mit vs. ohne) variiert, woraus sich ein $2 \times 2 \times 2$ -faktorielles Design ergab. Als abhängige Variable diente das via Verständnistest erfasste Verstehen der Beispielmeldungen. Das Experiment wurde als Online-Befragung in Nordrhein-Westfalen realisiert. Die Stichprobe umfasste $N = 1143$ Lehrkräfte.

Die Ergebnisse eines entlang der Ziele von VERA konzipierten Wissenstests legen nahe, dass die teilnehmenden Lehrkräfte insgesamt gut über VERA informiert sind. Es zeigt sich, dass sie insbesondere über die Durchführung und die Nutzungsvorschriften (z. B. Beratung über die Ergebnisse auf der Fach- und Schulkonferenz) informiert sind, weniger gut hingegen über die Ziele der Vergleichsarbeiten.

Während die VERA-Beispielmeldungen insgesamt als eher verständlich eingeschätzt werden, fällt die globale Einschätzung der Nützlichkeit der Beispielmeldungen niedriger aus. Analog zu den Ergebnissen der Gruppendiskussionen zeigt die Analyse der Einschätzungen der Nützlichkeit einzelner Text- und Grafikbausteine jedoch, dass die Schülerleistungstabelle und eine Grafik zu den erreichten Kompetenzstufen einzelner Schülerinnen und Schüler von den Lehrkräften als nützlich empfunden werden. Auch die Darbietung von Lernaufgaben wird von den Lehrkräften als eher nützlich empfunden.

Varianzanalytische Vergleiche ergeben keine bedeutsamen Unterschiede im Verstehen der Rückmeldungen zwischen den Bedingungen mit versus ohne erklärende schriftliche Informationen. Damit weisen die Ergebnisse darauf hin, dass die für diese Studie ausgewählten Grafiken von Lehrkräften auch ohne erklärende schriftliche Informationen verstanden werden. Im varianzanalytischen Vergleich der Bedingun-

gen mit drei versus fünf Grafiken zeige sich darüber hinaus, dass Rückmeldungen mit drei Grafiken besser verstanden werden als Rückmeldungen mit fünf Grafiken.

Die inhaltsanalytische Auswertung offener Antworten zur Nutzung der Beispielmeldungen weist bedingungsübergreifend darauf hin, dass sich Maßnahmen zur Weiterarbeit mit den Ergebnismeldungen insbesondere auf die gezielte Förderung von Schülerinnen und Schülern, die Besprechung von VERA-Ergebnissen mit den Kolleginnen und Kollegen, das gezielte Aufgreifen von VERA-Aufgaben sowie das Aufgreifen und Wiederholen bestimmter Unterrichtsinhalte beziehen.

4. Diskussion und Zusammenfassung zentraler Ergebnisse

Schulen sind im Zuge der Implementierung des neuen Steuerungsmodells zu einem zentralen Steuerungsakteur geworden. Selbststeuerung setzt voraus, auf Basis evaluativen Wissens Entscheidungen zu treffen und die schulische Qualität zu sichern und weiterzuentwickeln. Mit den Evaluationsverfahren VERA, Schulinspektion, interne Evaluation und zentrale Abschlussprüfungen stehen den Schulen hierfür verschiedene Informationsquellen zur Verfügung. Das StABil-Projekt hat untersucht, wie Schulen mit evaluativem Wissen umgehen, wie sie es zur Qualitätsentwicklung verwenden und welche Faktoren darauf Einfluss haben. Neben der Akzeptanz und Nutzung der unterschiedlichen Daten wurde erfasst, in welchem Umfang Schulen Daten auch im Bereich des Schulmanagements für die Schulprogramm- und Personalentwicklung nutzen.

Die wesentlichen bislang ermittelten Befunde aus dem Projekt können wie folgt zusammengefasst werden:

- Informationen aus Evaluationen werden in Schulen vor allem im Bereich der Unterrichtsentwicklung genutzt, weniger häufig werden Maßnahmen der Schulentwicklung durchgeführt.
- Kontrastierende Analysen zeigen Unterschiede zwischen der Nutzung unterschiedlicher Daten: Zwar werden alle Evaluationsergebnisse schulintern von vielen Akteursgruppen rezipiert und kommuniziert, im Vergleich mit zentralen Abschlussprüfungen, mit Schulinspektion und interner Evaluation wird VERA allerdings deutlich negativer beurteilt und als weniger nützlich wahrgenommen. Zentrale Abschlussprüfungen, Schulinspektionen und interne Evaluationen werden hinsichtlich ihrer Nützlichkeit eher positiv wahrgenommen.
- Es bestehen Interdependenzen zwischen den unterschiedlichen Evaluationsverfahren: Ergebnisse aus VERA und der Schulinspektion werden als Anlass genommen, verstärkt interne Evaluationen durchzuführen.
- In etwa der Hälfte der Schulen werden die Ergebnisse aus verschiedenen Evaluationsverfahren (VERA und Schulinspektion) im Zusammenhang für die Ableitung von Entwicklungsmaßnahmen verwendet.
- Die wahrgenommene Nützlichkeit und die Kommunikation bzw. gemeinsame Auswertung von Daten beeinflussen die Nutzung der VERA-Daten für die Entwicklung von Maßnahmen signifikant.

- Die Schulleitung spielt eine Schlüsselrolle für die Nutzung der Daten durch Lehrkräfte und Fachkonferenzen.
- Die Gestaltung der Rückmeldung und die Bereitstellung von Lernaufgaben können die Nutzung der VERA-Daten deutlich beeinflussen.
- Die Nutzung von Daten für die Personalentwicklung bleibt sowohl auf der Ebene der Schulleitungen als auch auf der Ebene der Fachkonferenzen hinter den durch die Bildungspolitik formulierten Erwartungen zurück.
- Evaluationsprogramme als wesentliche Bestandteile von Schulprogrammen werden sowohl durch politische Rahmenbedingungen (konkrete Vorgaben, gezielte Unterstützung und regelmäßige Rechenschaftspflicht) als auch durch organisationale Faktoren (qualifizierte Schulleitungen, breite Beteiligung der Lehrkräfte, koordiniert durch Steuergruppen) beeinflusst.

Die bisherigen Ergebnisse zeigen, dass in Schulen Evaluationsdaten zwar für einige Entwicklungsaktivitäten genutzt werden, allerdings kann nicht von einer flächendeckenden Nutzung gesprochen werden. Mögliche Gründe liegen einerseits in fehlenden Kompetenzen der Verarbeitung von datenbasierten Informationen sowie an grundsätzlichen Vorbehalten gegenüber der Evaluation bei den Schulleitungen und Lehrkräften; auf der anderen Seite sind mögliche Überforderungen durch ein hohes Tempo der Reformimplementation in Rechnung zu stellen (Thiel, Cortina & Pant 2014). Offensichtliche Defizite in der Verzahnung von Evaluationsverfahren und Unterstützungssystem sind hierbei ein zentrales Problem. Weiterhin stellt sich z. B. hinsichtlich der Nutzbarkeit von Evaluationen zur datenbasierten Unterrichtsentwicklung die Frage, ob beispielsweise die in VERA zurückgemeldete Schülerleistung allein ausreichende Informationen liefert oder erst durch die systematische Verknüpfung mit Informationen zu den Lehr- und Lernprozessen der betreffenden Klasse praktisch relevant werden kann. Das Gleiche trifft auf die oberflächlichen Unterrichtsbeobachtungen der Inspektion zu, die zudem lediglich auf Schulebene aggregiert zurückgemeldet werden. Die zum Teil geringe Akzeptanz der Evaluationsinstrumente und mangelnde Kompetenzen in der Datenanalyse kommen in der durch die Ergebnisse dokumentierten als teilweise gering wahrgenommenen Nützlichkeit zum Ausdruck. Weil, wie oben beschrieben, die wahrgenommene Nützlichkeit einen zentralen Stellenwert für die datenbasierte Praxis sowohl auf der Ebene des Unterrichts als auch auf der Ebene der Schule hat, verweisen die beschriebenen Befunde auf Handlungsbedarf. Die gegenwärtigen Unterstützungssysteme, Informationen und Qualifizierungen der schulischen Akteure sind vor diesem Hintergrund als nicht ausreichend zu betrachten.

Der Befund, dass die Kommunikation und kollegiale Auswertung der Daten die Nutzung erhöhen, legt es nahe, in den Schulen Anreizstrukturen und ein praktikables Prozedere für Datenauswertungen zu entwickeln, das einen intensiven Austausch über Evaluationsdaten, insbesondere auch in Fachkonferenzen, vorsieht.

Dass ein kleiner Teil der Schulen Daten aus VERA und Schulinspektionen gar nicht nutzt, obwohl sie gesetzlich dazu verpflichtet sind, wirft Fragen hinsichtlich des Verhältnisses von Schulen und Schulaufsicht auf. Ob die Schulaufsicht über die Nichtnutzung von Daten hinwegsieht oder ob etablierte Accountability-Mecha-

nismen, wie Zielvereinbarungen, nicht greifen, kann auf der Basis der Befunde aus den StABil-Teilstudien allerdings nicht beantwortet werden.

Unterschiedliche Befunde aus den StABil-Teilstudien verweisen auf die zentrale Rolle der Schulleitung für eine systematische Nutzung von Daten für die Unterrichts- und Schulentwicklung. Eine Schulleitung, die Lehrkräfte zur Datennutzung animiert und unterstützt, nimmt damit wesentlichen Einfluss auf die Datennutzung. Allerdings muss durch die politischen und bildungsadministrativen Rahmenbedingungen auch sichergestellt werden, dass die Schulleitungen ausreichende Entscheidungsspielräume für die Entwicklung von Maßnahmen haben. Dies ist gegenwärtig für den Bereich der Personalentwicklung zu bezweifeln (Brauckmann 2012).

Die Projektergebnisse verweisen außerdem darauf, dass das größte Potenzial einer datengestützten Unterrichts- und Schulentwicklung in einer koordinierten Nutzung der Daten liegt. So können beispielsweise Maßnahmen der internen Evaluation aus Inspektions- und VERA-Ergebnissen abgeleitet werden.

Einen besonderen Stellenwert haben Daten aus interner Evaluation. Die interne Evaluation ermöglicht eine fokussierte Beurteilung entwickelter Maßnahmen. Allerdings stehen der hohe Aufwand und die erforderlichen Kompetenzen einer Generierung von belastbaren Daten oft entgegen. Die Bereitstellung valider und reliabler Instrumente zur Selbstevaluation (vgl. Gärtner 2010) reduziert den Aufwand der Datengewinnung und steigert die Belastbarkeit der Daten beträchtlich, weil Schulen von der Instrumentenentwicklung entlastet werden.

Eine wichtige Möglichkeit der Unterstützung der Nutzung von Evaluationsdaten durch Schulen liegt, wie am Beispiel von VERA gezeigt werden konnte, in der Gestaltung von Ergebnismrückmeldungen, um die Nutzbarkeit zu erhöhen.

Die Ergebnisse werfen eine Reihe von Fragen für Anschlussstudien auf: Unter welchen Bedingungen sind Schulen in der Lage, selbstständig als Steuerungsakteure im Sinne des Modells des *Retroactive Planning* zu agieren? Welche Rolle spielen hierbei organisationale Faktoren und Fragen der Systemgestaltung beispielsweise hinsichtlich der Schulautonomie? Wie können unterschiedliche Akteure im Prozess evidenzbasierter Entscheidung effektiv zusammenwirken? Wie muss eine geeignete Unterstützung aussehen? Und führt Evaluationsnutzung tatsächlich (langfristig) zu einer Verbesserung der Schülerleistungen oder zu anderen wichtigen Erziehungszielen von Schulen?

Weiterführende Analysen im Kontext des StABil-Projekts werden zum einen die Zusammenhänge zwischen der Evaluationsdatennutzung und den Schülerleistungen sowie den kodeterminierenden Einfluss organisationaler Merkmale untersuchen. Zum anderen wird das Zusammenspiel verschiedener Ebenen und Akteure innerhalb der Schule durch zusätzliche Mehrebenenmodellierungen und die Auswertung von Fallstudien in den Blick genommen. Die Fallstudien ermöglichen zudem weitere differenziertere Einblicke in die tatsächlich durchgeführten Entwicklungsmaßnahmen. Außerdem werden Analysen zu möglichen Fächerunterschieden bei der Nutzung von Schülerleistungsdaten aus VERA und zentralen Prüfungen durchgeführt.

Literaturverzeichnis

- Bach, A., Wurster, S., Thillmann, K., Pant, H. A. & Thiel, F. (2014). Vergleichsarbeiten und schulische Personalentwicklung – Ausmaß und Voraussetzungen der Datennutzung. *Zeitschrift für Erziehungswissenschaft*, 17 (1), 61–84. doi: 10.1007/s11618-014-0486-5
- Benz, A., Lütz, S. & Simonis, G. (Hrsg.) (2007). *Handbuch Governance. Theoretische Grundlagen und empirische Anwendungsfelder*. Wiesbaden: Verlag für Sozialwissenschaften.
- Berkemeyer, N. & Müller, S. (2010). Schulinterne Evaluation – nur ein Instrument zur Selbststeuerung von Schulen? In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 195–218. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brauckmann, S. (2012). Schulleitungshandeln zwischen deconcentration, devolution und delegation (3D) – empirische Annäherungen aus internationaler Perspektive. *Empirische Pädagogik*, 26 (1), S. 78–102. URN: urn:nbn:de:0111-opus-58723
- Coburn, C. E. (2004). Beyond Decoupling. Rethinking the Relationship between the institutional environment and the classroom. *Sociology of Education*, 77, 211–244.
- DeBray, E., Parsons, G. & Woodworth, K. (2001). Patterns of response in four high schools under state accountability policies in Vermont and New York. In S. Fuhrman (Hrsg.), *From the capitol to the classroom: Standard based reform in the states*, S. 170–192. Chicago: University of Chicago Press.
- Dedering, K. (2012). Schulinspektion als wirksamer Weg der Systemsteuerung? *Zeitschrift für Pädagogik*, 58 (1), S. 69–88.
- Döbert, H. & Dedering, K. (Hrsg.) (2008). *Externe Evaluation von Schulen. Historische, rechtliche und vergleichende Aspekte*. Münster: Waxmann.
- Dubs, R. (2005). New Public Management und die Führung der Schule. In M. Sigrist, T. Wehner & A. Legler (Hrsg.), *Schule als Arbeitsplatz. Mitarbeiterbeurteilung zwischen Absicht, Leistungsfähigkeit und Akzeptanz*, S. 39–60. Zürich: Verlag Pestalozzianum.
- Elmore, R. F., Abelman, C. H. & Fuhrman, S. H. (1996). The new accountability in state educational reform: From process to performance. In H. F. Ladd (Hrsg.), *Holding schools accountable. Performance based reform in education*, S. 65–98. Washington, D.C.: Brookings Institution.
- Gärtner, H. (2010). Das ISQ-Selbstevaluationsportal. Konzepte eines Online-Angebots, um die Selbstevaluation in Schule und Unterricht zu unterstützen. *Die Deutsche Schule*, 102 (2), S. 163–175.
- Gärtner, H. (2013). Praxis und Verhältnis interner und externer Evaluation im Schulsystem im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 16 (4), S. 693–712.
- 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), S. 85–93.

- Gräsel, C., Stark, R., Sparka, A. & Herzmann, P. (2007). Schulische Kooperationsmuster und die Implementation eines Programms zur Förderung der Lesekompetenz. *Beiheft der Zeitschrift für Berufs- und Wirtschaftspädagogik*, 21, S. 93-107.
- Groß Ophoff, J. (2013). Der Effekt der Bezugsnormorientierung auf die Reflexion und Nutzung von Rückmeldungen aus Vergleichsarbeiten. *Empirische Pädagogik*, 27 (4), S. 442-458.
- Groß Ophoff, J., Hosenfeld, I. & Koch, U. (2007). Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. *Empirische Pädagogik*, 21 (4), S. 411-427.
- Hellrung, K. & Hartig, J. (2013). Understanding and using feedback – A review of empirical studies concerning feedback from external evaluations to teachers. *Educational Research Review*, 9, 174-190. doi: j.edurev.2012.09.001
- Helmke, A. (2004). Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule. *Das Seminar*, 2, S. 90-112.
- Hentschke, G. & Davies, B. (1997). Reconceptualising the Nature of Self-managing Schools. In B. Davies & L. Ellison, L. (Hrsg.), *School Leadership for the 21st Century*. London, New York, S. 23-35.
- Holtappels, H. G. & Müller, S. (2002). Inhaltsanalyse der Schulprogrammtexte Hamburger Schulen. In H.-G. Rolff, H. G. Holtappels, K. Klemm & H. Pfeiffer (Hrsg.), *Jahrbuch der Schulentwicklung, Daten, Beispiele und Perspektiven*, S. 209-231. Weinheim/München: Juventa.
- Husfeldt, V. (2010). Wirkungen und Wirksamkeit der externen Schulevaluation. Überblick zum Stand der Forschung. *Zeitschrift für Erziehungswissenschaft*, 14, 259-282. doi: 10.1007/s11618-011-0204-5
- Klein, E. D. (2013). *Statewide Exit Exams, Governance, and School Development. An International Comparison*. Münster: Waxmann.
- Klein, E. D., Krüger, M., Kühn, S. M. & van Ackeren, I. (2014). Wirkungen zentraler Abschlussprüfungen im Mehrebenensystem Schule. Eine Zwischenbilanz internationaler und nationaler Befunde und Forschungsdesiderata. *Zeitschrift für Erziehungswissenschaft*, S. 7-33. doi: 10.1007/s11618-014-0479-4
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Wolters Kluwer.
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529-548). Wiesbaden: Verlag für Sozialwissenschaften.
- Koretz, D. (2005). Alignment, high stakes, and the inflation of test scores. *Yearbook of the National Society for the Study of Education*, 104 (2), S. 99-118.
- Kühn, S. M. (2010). *Steuerung und Innovation durch Abschlussprüfungen?* Wiesbaden: Verlag für Sozialwissenschaften.
- Kussau, J. & Brüsemeister, T. (2007). Educational Governance: Zur Analyse der Handlungskoordination im Mehrebenensystem der Schule. In H. Altricher, T. Brüse-

- meister & J. Wissinger (Hrsg.), *Educational Governance. Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: Verlag für Sozialwissenschaften.
- LISUM (Hrsg.) (2008). *Rahmenlehrplan für die Sekundarstufe I. Mathematik*: Landesinstitut für Schule und Medien Berlin-Brandenburg.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, S. 145–169. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maag Merki, K., Klieme, E. & Holmeier, M. (2008). Unterrichtsgestaltung unter den Bedingungen zentraler Abiturprüfungen. Differenzielle Analysen auf Schulebene mittels Latent Class Analysen. *Zeitschrift für Pädagogik*, 54 (6), S. 791–808.
- Maier, U. & Kuper, H. (2012). Vergleichsarbeiten als Instrumente der Qualitätsentwicklung an Schulen – Überblick zum Forschungsstand. *Die Deutsche Schule*, 104 (1), S. 88–99.
- Mayntz, R. (2005). Governancetheorie als fortentwickelte Steuerungstheorie? In G. F. Schupper (Hrsg.), *Governance Forschung. Vergewisserung über Stand und Entwicklungslinien*. Baden-Baden: Nomos.
- MBJS (2008). *Orientierungsrahmen Schulqualität in Brandenburg. Qualitätsbereiche und Qualitätsmerkmale guter Schulen/Version 2.0*. Potsdam: Ministerium für Bildung, Sport und Jugend.
- MBJS (2012). Verwaltungsvorschriften über die Führung eines Leistungs- und Entwicklungsgesprächs mit Lehrkräften an öffentlichen Schulen. *Amtsblatt des Ministeriums für Bildung, Jugend und Sport*, 21. Jahrgang (Nr. 7), S. 286–290. <http://www.brandenburg.de/media/1172/mitarbeitergespraech.pdf>
- Mintrop, H. & Trujillo, T. (2007). The practical relevance of accountability systems for school improvement: a descriptive analysis of California schools. *Educational Evaluation and Policy Analysis*, 29 (4), S. 319–352.
- Nachtigall, C. & Hellrung, K. (2013). Zur zeitlichen Entwicklung der Reflexion von Vergleichsarbeiten. *Empirische Pädagogik*, 27 (4), S. 423–441.
- O'Day, J. A. (2002). Complexity, accountability and school improvement. *Harvard Educational Review*, 72 (3), S. 293–329.
- Pant, H. A. (2013). Wer hat einen Nutzen von Kompetenzmodellen? *Zeitschrift für Erziehungswissenschaft*, 16 (1), 71–79. doi: 10.1007/s11618-013-0388-y
- Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008). Offenheit für Innovationen: Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, 54, S. 827–845.
- Pant, H. A., Emmrich, R., Harych, P. & Kuhl, P. (2011). Leistungsüberprüfungen durch Schulleistungsstudien und Vergleichsarbeiten. In W. Sacher & F. Winter (Hrsg.), *Professionswissen für Lehrerinnen und Lehrer. Bd. 4: Diagnose und Beurteilung von Schülerleistungen. Grundlagen und Reformansätze*, S. 123–142. Baltmannsweiler: Schneider-Verlag Hohengehren.
- Pedder, D. & MacBeath, J. (2008). Organisational learning approaches to school leadership and management: teachers' values and perceptions of practice. *School Effectiveness and School Improvement*, 19 (2), S. 207–224.

- Scheerens, J. & Bosker, R. (1997). *The Foundations of Educational Effectiveness*. Oxford, UK: Pergamon Press.
- Scheerens, J. & Demeuse, M. (2005). The Theoretical Basis of the Effective School Improvement Model (ESI). *School Effectiveness and School Improvement*, 16 (4), 373–385. doi: 10.1080/09243450500234567
- Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, S. 482–496.
- 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), S. 229–255. doi: 10.1080/09243453.2011.652123
- Schildkamp, K., Vanhoof, J., van Petegem, P. & Visscher, A. (2012). The use of school self-evaluation results in the Netherlands and Flanders. *British Educational Research Journal*, 38 (1), S. 125–152. doi: 10.1080/01411926.2010.528556
- Schildkamp, K. & Visscher, A. (2009). Factors influencing the utilisation of a school self-evaluation instrument. *Studies In Educational Evaluation*, 35 (4), S. 150–159. doi: 10.1016/j.stueduc.2009.12.001
- Schliesing, A. (in Vorbereitung). *Entwicklung und Evaluation eines elektronischen Rückmeldeformats für Daten aus Vergleichsarbeiten*.
- SenBWF (2005). *Qualitätssicherung und Qualitätsentwicklung in der Berliner Schule – Rahmenkonzeption*. Hrsg. von der Senatsverwaltung für Bildung und Wissenschaft, Berlin.
- SenBWF (2008). *Ausführungsvorschriften zur Erstellung der Schulprogramme und zur internen Evaluation (AV Schulprogramm)*. http://www.berlin.de/imperia/md/content/sen-bildung/rechtsvorschriften/av_schulprogramm.pdf?start&ts=1422020679&file=av_schulprogramm.pdf
- Silins, H. & Mulford, B. (2004). Schools as Learning Organisations – Effects on Teacher Leadership and Student Outcomes. *School Effectiveness and School Improvement*, 15 (3–4), 43–466. doi: 10.1080/09243450512331383272
- Spillane, J. et al. (2002). Local theories of teacher change: The pedagogy of district policies and programs. *Teacher College Records*, 104, S. 377–420.
- Stamm, M. (2003). *Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung*. Münster: Waxmann.
- Thiel, F. (2008a). Die Organisation der Bildung – eine Zumutung für die Profession? In Y. Ehrenspeck, G. de Haan & F. Thiel (Hrsg.), *Bildung: Angebot oder Zumutung?*, S. 211–228. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Thiel, F. (2008b). Organisationssoziologische Vorarbeiten zu einer Theorie der Schulentwicklung. *Journal für Schulentwicklung*, 1, S. 31–39.
- Thiel, F., Cortina, K. S. & Pant, H. A. (2014). Steuerung im Bildungssystem im internationalen Vergleich. *Zeitschrift für Pädagogik*, 60. Beiheft, S. 123–138.
- Thiel, F. & Thillmann, K. (2012). Interne Evaluation als Instrument der Selbststeuerung von Schulen. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungs-*

- methodische Implikation*, S. 35–55. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Thillmann, K., Bach, A., Gerl, C. & Thiel, F. (2013). Schulische Selbststeuerung durch interne Evaluation. Eine Analyse von Evaluationsprogrammen Berliner und Brandenburger Gymnasien. *Die Deutsche Schule*, 12. Beiheft, S. 197–225.
- vbw – Vereinigung der Bayerischen Wirtschaft e.V. (2010). *Bildungsautonomie: Zwischen Regulierung und Eigenverantwortung – die Bundesländer im Vergleich. Expertenrating der Schul- und Hochschulgesetze der Länder zum Jahresgutachten 2010*. München: Verlag für Sozialwissenschaften.
- Visscher, A. J. & Coe, R. (2003). School Performance Feedback Systems: Conceptualisation, Analysis, and Reflection. *School Effectiveness and School Improvement*, 14 (3), S. 321–349.
- Wiesenthal, H. (2005). Markt, Organisation und Gemeinschaft als „zweitbeste“ Verfahren sozialer Koordination: In W. Jäger, W. & U. Schimank (Hrsg.), *Organisationsgesellschaft. Facetten und Perspektiven*, S. 223–264. Wiesbaden.
- Wurster, S. & Gärtner, H. (2013). Schulen im Umgang mit Schulinspektion und deren Ergebnissen. *Zeitschrift für Pädagogik*, 59 (3), S. 425–445.
- Wurster, S. & Richter, D. (eingereicht). *Nutzung von Schülerleistungsdaten aus Vergleichsarbeiten und zentrale Abschlussprüfungen für Unterrichtsentwicklung in Fachkonferenzen*.
- Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2013). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. *DDS – Die Deutsche Schule*, 12, S. 19–49.

Johannes Bellmann, Sebastian Schweizer, Corrie Thiel

Nebenfolgen Neuer Steuerung unter Bedingungen von „low-stakes“ und „no-stakes“ – Qualitative und quantitative Befunde einer Untersuchung in vier Bundesländern

1. Einleitung

Bei der Frage nach Effekten Neuer Steuerung im Schulsystem hat sich die Forschung bislang, von Ausnahmen abgesehen (vgl. van Ackeren et al. 2013), auf die Evaluation einzelner Instrumente der Neuen Steuerung und deren Nutzung konzentriert, wobei meist die Frage nach den intendierten Effekten des jeweiligen Instruments im Mittelpunkt steht. Die Evaluationslogik der Instrumente spiegelt sich somit auch in der Forschung wider. So hat sich beispielsweise in Deutschland die Forschung über zentrale Vergleichsarbeiten bislang auf die Frage konzentriert, inwiefern etwa die Rückmeldungen aus VERA von Lehrkräften tatsächlich, wie intendiert, für die Verbesserung des eigenen Unterrichts genutzt werden (vgl. Wacker et al. 2012). Die Ergebnisse sind – wie Ewald Terhart jüngst resümierte – „zum größten Teil sehr ernüchternd“ (2012, S. 10). Die Lehrkräfte nutzen Testergebnisse eher zur Abrundung ihrer Notengebung, aber nur zu einem geringen Teil zur Verbesserung des eigenen Unterrichts.

Mit der Beantwortung der Frage nach dem Eintreffen bzw. Nichteintreffen der intendierten Effekte ist freilich die Frage nach den nicht intendierten Effekten oder, allgemeiner gesagt, den Nebenfolgen Neuer Steuerung noch nicht beantwortet. Nebenfolgen können dabei weitreichender sein als fehlgeleitete Formen der Nutzung von Steuerungsinstrumenten; sie können auch „externe Effekte“ auf aggregierter Ebene einschließen, die in einer auf Nutzung und Implementation fokussierten Forschung zumeist nicht erfasst werden. Die Nebenfolgenproblematik wurde bislang in der deutschsprachigen Forschungsliteratur eher am Rande und im Zusammenhang mit der Evaluation von Einzelinstrumenten behandelt (vgl. z. B. Maier 2010; Quesel et al. 2011; Wacker & Kramer 2012). Eine systematische empirische Untersuchung zur ganzen Bandbreite möglicher Nebenfolgen neuer Steuerungsinstrumente fehlte bislang.

Dies steht in Kontrast zum englischsprachigen Raum, wo es eine breite Forschungsliteratur und lebhafte Diskussion über die nicht intendierten Effekte (Jones et al. 2003) und Kollateralschäden (Nichols & Berliner 2007) der Neuen Steuerung gibt. Eine Aufarbeitung der internationalen Nebenfolgendiskussion findet sich bei Bellmann & Weiß (2009) und de Wolf & Janssens (2007). Auffällig ist, dass solche Nebenfolgen wie das berühmte-berüchtigte „teaching to the test“ in der Regel mit „high-stakes“ in Verbindung gebracht werden, d. h. dem Umstand, dass in vielen Ländern Ergebnisse zentraler Ver-

gleichsarbeiten mit ernsthaften Konsequenzen für Schulen und Lehrkräfte verknüpft sind. Einzelne Forschungsprojekte haben sogar versucht, die Verbreitung von Nebenfolgen in einen direkten Zusammenhang mit dem Grad von „high-stakes pressure“ in bestimmten Bundesstaaten zu stellen (vgl. Nichols et al. 2005).

Die insgesamt deutlich geringere Aufmerksamkeit für die Nebenfolgenproblematik im deutschsprachigen Raum ist auch mit dem Umstand verbunden, dass hier die Einführung neuer Steuerungsinstrumente bislang nicht systematisch mit „high-stakes“ verknüpft worden ist. So führt Blömeke (2004, S. 82) unerwünschte Nebenwirkungen von zentralen Vergleichsarbeiten „in erster Linie auf die unmittelbare Verknüpfung von Testergebnissen und Konsequenzen“ zurück. Die Erfahrungen aus den USA und Großbritannien mit „high-stakes accountability“ dienen dann im deutschen Diskurs immer wieder als „Warnung vor dem falschen Vorbild“ (Böttcher 2012). Was sich allerdings nicht belegen ließe, so Blömeke (2004, S. 83), seien „Befürchtungen, bereits die Einführung von Tests – beispielsweise zu Rückmelde- und Vergleichszwecken ohne unmittelbare Konsequenzen – würde unerwünschte Nebenwirkungen (z. B. teaching to the test) mit sich bringen, ja solche würden sogar die Regel sein“. Dementsprechend hat die Frage, ob Nebenfolgen Neuer Steuerung auch unter Bedingungen von „low-stakes“ oder „no-stakes“ zu beobachten sind, bislang nur geringe Aufmerksamkeit erfahren (vgl. Jäger 2012).

Das BMBF-Projekt „Nicht-intendierte Effekte Neuer Steuerung im Schulsystem“ mit dem Akronym „Nefo“ (für Nebenfolgen) hat nun versucht, dieses Forschungsdesiderat auf einer breiten empirischen Basis aufzuarbeiten. Entgegen der in der Forschungsliteratur verbreiteten Erwartung weisen die Befunde des Nefo-Projekts darauf hin, dass schulische Akteure offensichtlich auch unter den in Deutschland herrschenden Bedingungen von „low-stakes“ oder „no-stakes“ eine ganze Bandbreite von Nebenfolgen berichten bzw. beobachten.

Um diese Nebenfolgen in einem ersten explorativen Zugriff zu untersuchen, wurde ein triangulatives Design gewählt. In der ersten Projektphase wurde eine Interviewstudie mit Schulleitungen und Lehrkräften in Berlin, Brandenburg, Rheinland-Pfalz und Thüringen durchgeführt. Hieran schloss sich in der zweiten Projektphase eine quantitative Fragebogenerhebung an. Im Folgenden wird zunächst das dem Nefo-Projekt zugrunde liegende Verständnis von Nebenfolgen erläutert, bevor nach einer kurzen Vorstellung der Steuerungsregime der vier untersuchten Bundesländer die Ergebnisse der qualitativen Erhebung dargestellt werden. Danach werden Ergebnisse der Fragebogenerhebung anhand ausgewählter Nebenfolgenitems präsentiert.

Insgesamt zeigten sich differenzielle Effekte im Spektrum der Nebenfolgen. Dies hängt, wie im Folgenden gezeigt wird, damit zusammen, dass es unterschiedliche Formen von Nebenfolgen gibt, für deren Entstehen auch unterschiedliche Erklärungsansätze erforderlich sind. Grundsätzlich lassen sich Formen eines anreizkonformen Anpassungsverhaltens identifizieren, das vom Steuerungsmodell selbst nahegelegt oder auf der Ebene spezifischer Steuerungsregime aktiv forciert wird. Davon lassen sich Formen eines anreizwidrigen Ausweichverhaltens unterscheiden, das vor allem mit Druckmomenten spezifisch ausgestalteter staatlich-administrativer und marktlich-wettbewerblicher Steuerungsmechanismen in Verbindung steht. Von diesen Nebenfolgen auf der Verhaltensebene lassen sich wiederum Nebenfolgen auf der

Einstellungsebene unterscheiden, die Verarbeitungsformen entweder von Druckmomenten spezifisch ausgestalteter Steuerungsmechanismen sein können oder von Deutungsangeboten, die bereits im Modell Neuer Steuerung selbst angelegt sind.

2. Nebenfolgen

Nebenfolgen sozialen Handelns lassen sich aus unterschiedlichen Perspektiven in den Blick nehmen. In der Rede von *nicht intendierten Effekten* (z. B. Bellmann & Weiß 2009) sind Intentionen von Akteuren Ausgangspunkt der Betrachtung. Entscheidend ist, ob beobachtbare Effekte auf konkrete Absichten rückgeführt werden können. Ein solcher Zugriff auf Nebenfolgen steht vor dem Problem, zunächst Intentionen rekonstruieren zu müssen, bevor der Blick auf Nebenfolgen gerichtet werden kann. Zudem stellt sich das Problem der Reintentionalisierung von Effekten.

Nebenfolgen müssen jedoch nicht notwendig mit Rekurs auf die (vorgängige) Intentionalität von Akteuren gefasst werden. Sie können auch als *externe Effekte* von Handlungen beschrieben werden, die unbeteiligte Dritte betreffen (Holzer 2006, S. 50). Dabei ist es nicht entscheidend, ob diese externen Effekte *ex ante* intendiert oder nicht-intendiert waren oder ob sich die Akteure *ex post* der externen Effekte ihres Handelns bewusst sind. Entscheidend ist, dass bestimmten Akteuren externe Effekte als Nebenfolgen ihres Handelns *zugeschrieben* werden, sei es von anderen betroffenen oder beobachtenden Akteuren oder aus der wissenschaftlichen Beobachterperspektive. Dies impliziert Konflikte darüber, wie Folgen zu bewerten sind und wem sie überhaupt zugerechnet werden können und müssen (vgl. ebd., S. 52).

Sowohl nicht intendierte Effekte wie auch externe Effekte können positiv oder negativ bewertet werden. Ausgehend von der Annahme, dass neue Steuerungsinstrumente als Instrumente der Qualitätsentwicklung eingeführt wurden, richtet sich der Fokus des Nefo-Projektes aber auf *diejenigen externen Effekte, die zum Zweck der Qualitätsentwicklung in einem negativen oder zumindest ambivalenten Verhältnis stehen*. Hierbei ist in Rechnung zu stellen, dass es sich bei der von Neuer Steuerung bezweckten Qualitätsentwicklung um eine mehrdimensionale Größe handelt, die Fortschritte im Hinblick auf Effizienz und Gerechtigkeit des Schulsystems verspricht (vgl. Fuchs 2009; Bellmann 2006). Positive externe Effekte wie etwa verstärkte Kooperation unter Lehrkräften können ebenfalls festgestellt werden, bilden aber nicht den Untersuchungsgegenstand des Projekts.

Jede Klassifikation beobachtbarer Phänomene als Nebenfolgen Neuer Steuerung beinhaltet eine Zuschreibung und eine Bewertung. Erstens müssen beobachtbare Phänomene der Neuen Steuerung zugeschrieben werden. Im Nefo-Projekt wurde dies methodisch durch die pragmatische Entscheidung umgesetzt, den Zuschreibungen der interviewten Lehrkräfte zu folgen (vgl. Kapitel 4). Zweitens müssen die identifizierten Effekte Neuer Steuerung als Nebenfolgen bewertet werden. Im Nefo-Projekt wurde jedes Phänomen, das im Rahmen der Interviews von Lehrkräften und Schulleitern als Effekt Neuer Steuerung benannt wurde, vonseiten des Forscherteams unter Rückgriff auf Literatur in Erziehungswissenschaft und Bildungsforschung daraufhin beurteilt, ob es zum Zweck schulischer Qualitätsentwicklung in einem nega-

tiven oder zumindest ambivalenten Verhältnis steht. Entsprechend unterscheiden etwa auch Ehren und Visscher (2006, S. 63) „effects“ und „side-effects“ von Schulinspektionen, indem sie zunächst von den bezweckten Veränderungen ausgehen, die als Folge des neuen Steuerungsinstruments erwartet werden. Nebenfolgen werden dann als „dysfunctional effects“ (S. 57) verstanden, die den bezweckten Veränderungen zuwiderlaufen. Die Einordnung beobachteter Phänomene als Nebenfolgen erfolgte im Nefo-Projekt unabhängig von den Einschätzungen der Lehrkräfte, wenngleich die Einschätzungen des Forscherteams und der Lehrkräfte z. T. übereinstimmten.

Für die Nebenfolgen Neuer Steuerung lässt sich weder eine finite Liste vorlegen, noch sind die als Nebenfolgen gefassten Phänomene beliebiger Art. Nebenfolgen stellen Reaktionsformen von Akteuren auf Steuerungsinstrumente und Steuerungsversuche dar, bei denen man als wissenschaftlicher Beobachter auch mit bislang unbekanntem „kreativen“ Neuerfindungen rechnen muss. Zugleich gibt es im Spektrum der Phänomene typische wiederkehrende Nebenfolgen, die in der Literatur inzwischen gut dokumentiert sind. Was diese typischen wiederkehrenden Nebenfolgenphänomene gemeinsam haben, ist Gegenstand einer Theorie der Nebenfolgen Neuer Steuerung, für die es bislang nur erste Ansätze gibt.

Eine Theorie der Nebenfolgen Neuer Steuerung kann nur im Zusammenhang mit einer Theorie Neuer Steuerung selbst entwickelt werden. Im Kern lässt sich die Neue Steuerung im Schulsystem und in anderen Bereichen des öffentlichen Sektors als *datengetriebene Steuerung* auffassen, die auf der Etablierung eines Systems von Leistungsindikatoren und entsprechenden Rückmeldesystemen beruht (vgl. Bellmann 2014).

Zur Forschung über Nebenfolgen indikatorengestützter Systeme von „performance measurement“ gibt es eine breite sozialwissenschaftliche Literaturgrundlage. Bereits Etzioni (1964) verweist auf die Probleme der Messung der Effektivität und Effizienz von Organisationen: „Most organizations under pressure to be rational are eager to measure their efficiency. Curiously, the very effort – the desire to establish how we are doing and to find ways of improving if we are not doing as well as we ought to – often has quite undesired effects from the point of view of the organizational goals. Frequent measuring can distort the organizational efforts because, as a rule, some aspects of its output are more measurable than others. Frequent measuring tends to encourage over-production of highly measurable items and neglect of less measurable ones“ (S. 9). Nach Etzioni sind die Nebenfolgen von Leistungsindikatoren eher gering, wenn es sich um Organisationen mit eng gefassten und konkreten Zielstellungen handelt. Anders ist dies bei Organisationen mit vagen Zielstellungen: „The distortion consequences of over-measuring are larger when it is impossible or impractical to quantify the more central, substantive output of an organization“ (ebd.).

Anknüpfend an Etzioni können drei Einsichten für eine Theorie der Nebenfolgen Neuer Steuerung festgehalten werden: (1) Nebenfolgen indikatorengestützter Rückmeldesysteme entstehen nicht erst durch eine fehlende oder falsche Nutzung dieser Systeme durch die Akteure; solche Systeme sind vielmehr systematisch mit einem Risiko verbunden, neben dem gewünschten Verhalten auch die Produktion von Nebenfolgen anzureizen. (2) Nebenfolgen kommen nicht erst dann ins Spiel, wenn Akteure dem Druck der Leistungsmessung auszuweichen versuchen (etwa durch Formen des Betrugs, Window Dressing oder Klientenselektion), sondern auch,

wenn sie lernen, sich dem System der Leistungsindikatoren anzupassen, indem sie ihr Handeln verstärkt an den gemessenen Größen orientieren, während nicht gemessene Größen an Bedeutung verlieren (siehe Formen der Reallokation, Coaching, Kurzsichtigkeit) (vgl. auch Thiel/van Leeuw 2002) (3) Es muss davon ausgegangen werden, dass die Einführung von Leistungsindikatoren in Organisationen mit sog. „ill-defined problems“ ein besonderes hohes Risiko birgt. Dies gilt insbesondere für professionelle Handlungskontexte, die durch eine Orientierung an vage definierten gesellschaftlichen Zentralwerten und einen hohen Grad von Unsicherheit gekennzeichnet sind (vgl. Kurtz 2009; Hopmann 2003, S. 436).

Ein wesentlicher Beitrag zu einer Theorie der Nebenfolgen datengetriebener Steuerung über Leistungsindikatoren und Rückmeldesysteme stammt von Donald T. Campbell (1975), der die in der Literatur vielfach dokumentierten Nebenfolgen als Korruptionsphänomene beschreibt. Der später als „Campbell's law bezeichnete soziale Mechanismus lautet: „The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor“ (S. 35).

Von Korruption ist hier in einem doppelten Sinne die Rede: Einerseits im Sinne einer potenziellen *Korruption der Indikatoren*, die als Qualitätsmaße an Validität verlieren, sobald die Akteure ihr Leistungsspektrum auf ebendiese Maße verengen. Genau in diesem Sinne analysiert auch Koretz (2005) das Phänomen von „score inflation“ im Kontext von standards-based accountability. Andererseits im Sinne einer potenziellen *Korruption der Akteure* selbst, deren Motivationsstruktur durch die Einführung eines Systems von Leistungsindikatoren so umgebaut werden kann, dass sie bestimmte Leistungen, die sie vormals aus freien Stücken bzw. auf der Basis ihrer professionellen Überzeugung erbracht haben, nun nicht mehr (im gleichen Maße) erbringen. In diesem Fall kann von einem Effekt des „crowding out“ gesprochen werden, der zur Verdrängung bestehender Motivationsstrukturen führt (Frey & Jegen 2002; Jäger 2012).

Nichols und Berliner (2005), die an Campbell anschließen, beschreiben Nebenfolgen Neuer Steuerung im Schulsystem als „Corruption of Indicators and Educators“, wobei sie – anders als Campbell selbst – die mit Leistungsergebnissen verbundenen „high-stakes“ als den entscheidenden Faktor betrachten, der für die Produktion von Nebenfolgen verantwortlich ist. Dies führt dazu, dass die bei Etzioni und Campbell vorhandene Einsicht in den Hintergrund tritt, dass Nebenfolgen mit Systemen indikatorengestützter Leistungsmessung nicht nur in kontingenter, d. h. über die zusätzliche Einführung von high-stakes vermittelter, sondern in systematischer Weise verbunden sein können. Die sowohl in der englischsprachigen als auch deutschsprachigen Literatur zu verzeichnende Fixierung der Nebenfolgendiskussion auf die Problematik von „high-stakes“ steht deshalb der Entwicklung einer Theorie der Nebenfolgen eher im Wege.

Im Nefo-Projekt wurde deshalb versucht, an die ältere (organisations-)soziologische Literatur über Nebenfolgen indikatorengestützter Leistungsmessung anzuschließen. So wies bereits Ridgway (1956) darauf hin, dass mit „Dysfunctional Consequences of Performance Measurements“ auch dann gerechnet werden müsse, wenn diese bloß der Rückmeldung und Information dienen: „Even where performance

measures are instituted purely for purposes of information, they are probably interpreted as definitions of the important aspects of that job activity and hence have important implications for the motivation of behavior” (S. 247).

Die Beschreibung von Nebenfolgen als Korruptionsphänomene in dem oben skizzierten Sinne erscheint im Blick auf die Bandbreite der im Projekt dokumentierten Nebenfolgen die höchste theoretische Integrationskraft zu besitzen. Diese Theoriperspektive erscheint insofern auch für eine Erklärung von Nebenfolgen Neuer Steuerung unter Bedingungen von low-stakes und no-stakes geeignet zu sein.

3. Bundesländersample

Im Rahmen des Nefo-Projektes wurden Nebenfolgen Neuer Steuerung in den Bundesländern Berlin, Brandenburg, Rheinland-Pfalz und Thüringen untersucht. In allen vier Ländern wurden in den letzten Jahren Instrumente der Neuen Steuerung wie zentrale Vergleichsarbeiten und Schulinspektionen sowie erweiterte Schulautonomie und erweiterte Schulwahlmöglichkeiten eingeführt.

Als Kennzeichen des neuen Steuerungsmodells kann die Verknüpfung dreier Steuerungsmechanismen angesehen werden: staatlich-administrative Steuerung, marktlich-wettbewerbliche Steuerung und professionelle Selbststeuerung (vgl. Thiel et al. 2014, S. 127). Bei der Verknüpfung der drei Steuerungsmechanismen spielen die Erhebung und Rückmeldung von Leistungs- und Prozessdaten von Schulen eine zentrale Rolle: Sie ermöglichen Vergleiche zwischen den staatlich-administrativ gesetzten normativen Erwartungen in Form von Standards und Qualitätsrahmen mit dem Ist-Zustand schulischer Ergebnisse und Prozesse; sie schaffen zugleich Möglichkeiten für Vergleiche zwischen Schulen und eröffnen so einen zwischenschulischen Wettbewerb im Rahmen einer gemeinsamen Währung; schließlich bieten Rückmeldungen von Ergebnis- und Prozessdaten den schulischen Akteuren eine Grundlage für die Reflexion und Verbesserung ihrer Arbeit.

Die Länderauswahl erfolgte zunächst literaturbasiert nach den Kriterien Einführungszeitpunkt und Verbreitungsstärke neuer Steuerungsinstrumente. Ziel war es, ein möglichst kontrastives Sample zusammenzustellen. Für beide Dimensionen lässt sich zeigen, dass Brandenburg (neben Niedersachsen und Hessen) und mit einem gewissen Abstand auch Berlin eine führende Position im Ländervergleich einnehmen, Thüringen sich bislang eher zurückhaltend verhalten hat und Rheinland-Pfalz im Mittelfeld anzusiedeln ist (Rürup 2007, S. 266).

Nach der Auswahl wurde für jedes der vier Bundesländer eine detaillierte Beschreibung der landesspezifischen Ausgestaltung Neuer Steuerung angefertigt auf Basis der Analyse einer Vielzahl bildungspolitischer Dokumente, wie z. B. Gesetzen, Verordnungen, Rundbriefen an die Schulen und öffentlich zugänglichen Begleitmaterialien zu den eingeführten neuen Steuerungsinstrumenten.¹ Für jedes Bundesland wurde gefragt, wie stark jeweils Evaluationsdaten in die staatlich-administrati-

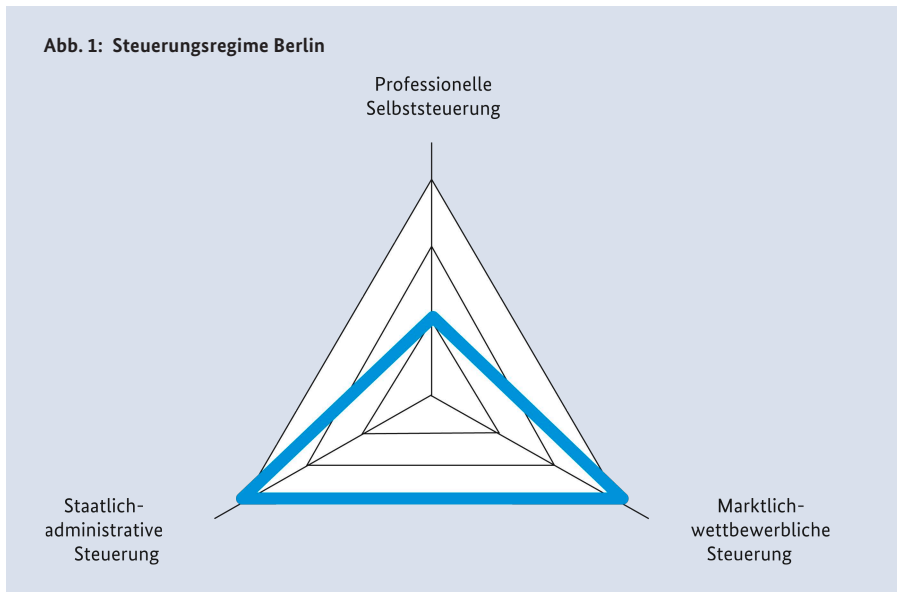
1 Eine detaillierte Liste der analysierten Dokumente zur Charakterisierung der Steuerungsregime findet sich auf der Projekthomepage unter <http://www.uni-muenster.de/EW/forschung/projekte/nefo/materialien.html>.

ve, marktlich-wettbewerbliche und professionelle Steuerung eingebunden sind. Die Ausprägung der Stärke des marktlich-wettbewerblichen Mechanismus wurde festgemacht an der Existenz flächendeckend freier Schulwahlmöglichkeiten für Eltern, den erweiterten Schülerwahlmöglichkeiten für Schulen und dem Vorhandensein einer Schuldatenbank, über die schulische Prozess- und Leistungsdaten öffentlich zugänglich sind. Die Einschätzung der Stärke des Mechanismus staatlich-administrativer Steuerung wurde anhand der Anzahl von Steuerungsinstrumenten vorgenommen, die die Interaktion von Staat und Schule hinsichtlich des Umgangs mit Daten regeln. Dazu zählt die Kommunikation der Ergebnisse interner und externer Evaluation der Schule an die Schulaufsicht, Zielvereinbarungen zwischen der Schule und der Administration in Anschluss an die Verfahren interner und externer Evaluation und die Verkürzung des Inspektionsintervalls im Falle schlechter Ergebnisse von Schulinspektionen. Staatlich-administrative und marktlich-wettbewerbliche Steuerung stellen dabei insofern funktionale Äquivalente dar, als dass sie Voraussetzungen darstellen, dass Druck zur Veränderung von Schule und Unterricht ausgeübt werden kann. Die Einschätzung der Ausprägung professioneller Selbststeuerung ergibt sich aus der Anzahl an Evaluationsverfahren, die nicht in marktlich-wettbewerbliche und staatlich-administrative Steuerung eingebunden sind. Dazu zählt etwa, wenn die Formulierung von Zielen nach einem Evaluationsprozess rein auf der schulischen Ebene verbleibt und nicht der administrativen Kontrolle unterliegt. Die Einschätzung der Stärke der drei Mechanismen verbleibt methodisch bedingt durch die Dokumentenanalyse auf der Ebene der „policy action“ (Tyack & Cuban 1995, S. 40). Die Ebene der „policy implementation“ (ebd.) bleibt hier eingeklammert. Beispielsweise können die bildungspolitischen Voraussetzungen für die Entstehung von Wettbewerb zwar gegeben sein („policy action“), möglicherweise entsteht aber faktisch kein Wettbewerb zwischen Schulen („policy implementation“), weil der Besuch weit entfernter Schulen keine praktikable Option darstellt.

Im Berliner Steuerungsregime kommt Daten sowohl im Rahmen staatlich-administrativer als auch marktlich-wettbewerblicher Steuerung von Schulen maßgebliche Bedeutung zu. Entsprechend der programmatischen Konzeption kann die Schulaufsicht z. B. über regelmäßig abzuschließende Zielvereinbarungen Einfluss auf die Einzelschule nehmen. Daneben sind in Berlin die Voraussetzungen für marktlich-wettbewerbliche Steuerung geschaffen worden. Über das Schulporträt wird den „Kunden“ des Schulsystems eine Vielfalt von Informationen über Einzelschulen zur Verfügung gestellt, die den Vergleich von infrage kommenden Schulen ermöglicht und Schulwahlentscheidungen beeinflussen kann. Professioneller Selbststeuerung kommt dementsprechend nur eine untergeordnete Rolle zu. Es findet sich mit der Selbstevaluation der Lehrkräfte lediglich ein Evaluationsinstrument, das nicht der administrativen Kontrolle unterliegt und das auch im Wettbewerb zwischen Schulen keine Rolle spielen kann, da die Ergebnisse gegenwärtig nur an die betreffende Lehrkraft zurückgemeldet werden. In den folgenden Abbildungen repräsentiert der äußere Ring eine starke, der mittlere eine mittlere und der innere Ring eine schwache Ausprägung des Steuerungsmechanismus.

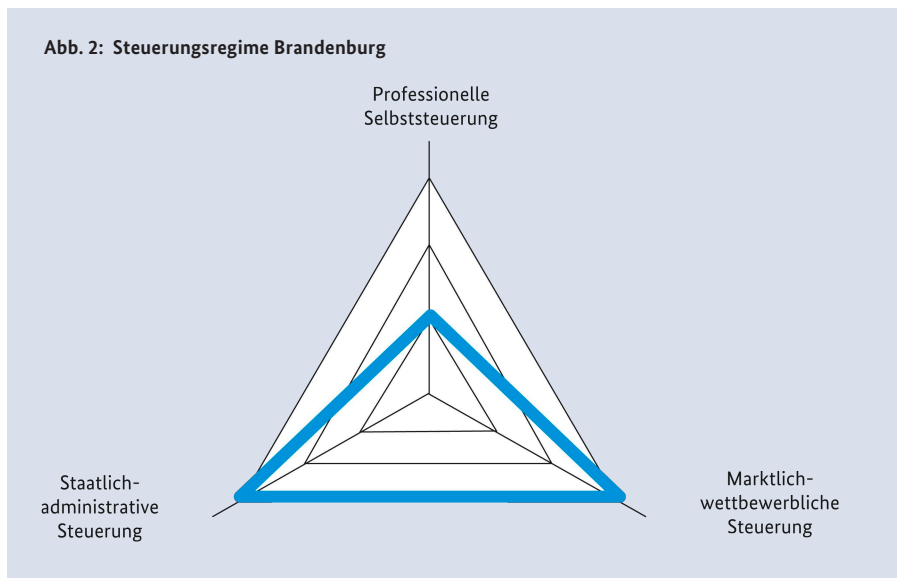
Das brandenburgische Steuerungsregime ist ähnlich ausgestaltet wie das Berliner Regime. Auch hier sind schulische Leistungs- und Prozessdaten Teil der staatlich-ad-

Abb. 1: Steuerungsregime Berlin

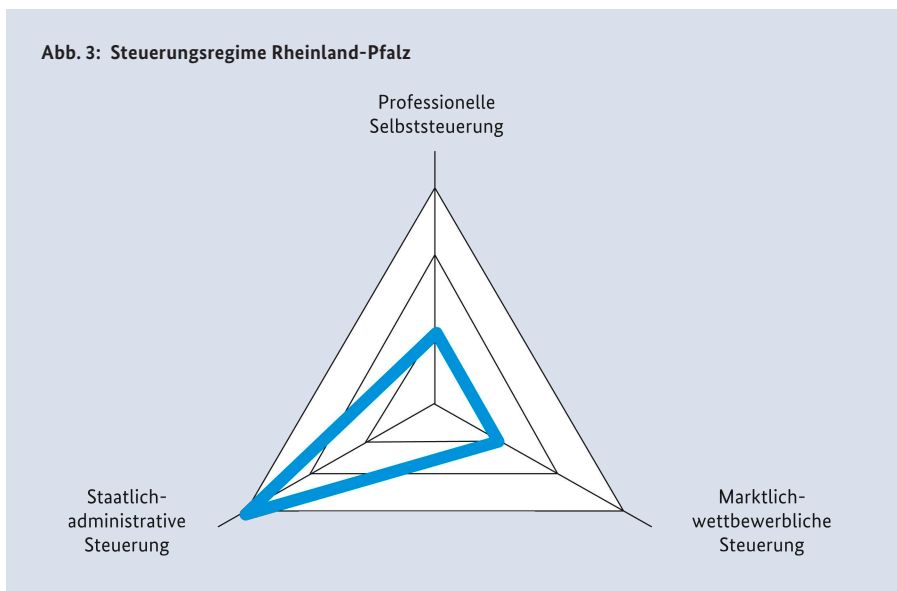


ministrativen Einflussnahme auf Schule, z. B. im Rahmen der sog. „datengestützten Qualitätsgespräche“ (DAQ) zwischen einem Schulaufsichtsbeamten und der Schulleitung. Auch ist in Brandenburg mittlerweile mit dem „Schulporträt“ eine Datenbank eingerichtet worden, über die schulische Leistungsdaten öffentlich zugänglich sind, die von Eltern in die Schulwahlentscheidung einbezogen werden können (marktlich-wettbewerbliche Steuerung). Im Vergleich zum Stadtstaat Berlin ist für das Flächenland Brandenburg allerdings zu erwarten, dass der Druck des Marktes geringer ausfällt. Dem Mechanismus der professionellen Selbststeuerung kommt analog zu Berlin ebenfalls eine nur marginale Rolle zu.

Abb. 2: Steuerungsregime Brandenburg

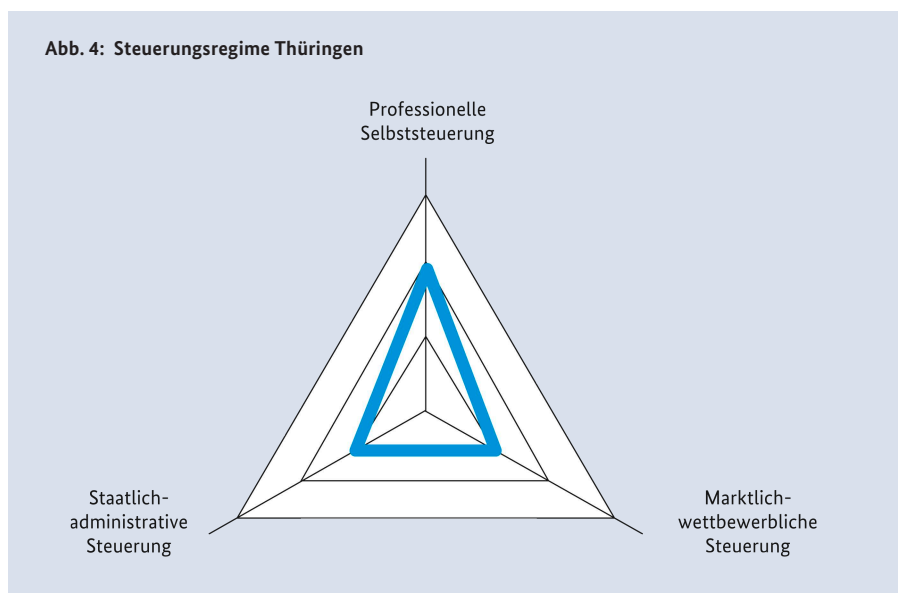


In Rheinland-Pfalz ist die staatlich-administrative Steuerung von Schule durch Kontrolle primärer Steuerungsmechanismus. Die Schulen legen der Administration im Anschluss an die verschiedenen Verfahren interner wie externer Evaluation schulintern ausgearbeitete Zielvereinbarungen vor. Da sowohl die schulintern gezogenen Konsequenzen aus der internen wie der externen Evaluation an die Schulaufsicht mittels eines Formblattes kommuniziert werden sollen, lässt sich in Bezug auf das rheinland-pfälzische Steuerungsregime nicht von professioneller Selbststeuerung im engen Sinne sprechen. Marktlich-wettbewerblicher Steuerung kommt aufgrund des Fehlens einer öffentlich zugänglichen Datenbank mit Leistungsdaten aller Schulen geringere Bedeutung zu als in Berlin und Brandenburg. Gleichzeitig steht den Eltern die Wahl der weiterführenden Schule offen, sodass auf programmatischer Ebene die Möglichkeit für das Entstehen marktähnlicher Strukturen geschaffen ist.



In Thüringen hingegen ist der Anstoß innerschulischer und schulübergreifender professioneller Selbststeuerungsprozesse ein zentralerer Steuerungsmechanismus. Daten spielen lediglich im Rahmen von Zielvereinbarungen zwischen Schulaufsicht und Einzelschule in Anschluss an die externe Evaluation eine Rolle. Auch marktlich-wettbewerblicher Steuerung kommt in Thüringen nur eine untergeordnete Bedeutung zu, da ebenso wie in Rheinland-Pfalz keine Datenbank mit schulischen Leistungsdaten existiert. Zudem sind die Schulwahlmöglichkeiten der Eltern in Thüringen nicht flächendeckend eingeführt.

Das Vergleichsdesign lässt hypothetische Schlüsse darüber zu, inwiefern die auftretenden Nebenfolgen durch das Steuerungsmodell selbst systematisch erzeugt worden sind und welchen Einfluss die Merkmale der konkret ausgestalteten Steuerungsregime auf das Auftreten von Nebenfolgen haben. Gemeinsamkeiten hinsicht-



lich der Art und Verbreitung von Nebenfolgen der vier Steuerungsregime sprechen für Ersteres. Unterschiede hinsichtlich der Art und Verbreitung von Nebenfolgen zwischen den Bundesländern könnten dagegen eher mit den Charakteristika der spezifischen Steuerungsregime im Zusammenhang stehen.

4. Qualitative Erhebung

In der ersten Phase des Projekts wurden von März bis September 2012 101 Interviews² mit Schulleitungen und Lehrkräften an Schulen aller allgemeinbildenden Schulformen geführt. Angestrebt war es, je Schulform an drei Schulen Interviews mit der Schulleitung und zwei Lehrkräften zu führen, wobei eine der Lehrkräfte aktiv in Schulentwicklungsprozesse involviert sein sollte und die andere nicht. Aufgrund von Schulstrukturreformen in den Bundesländern wurden keine Interviews an Haupt- und Realschulen geführt. Außerdem gelang es nicht, Integrierte Gesamtschulen für die Teilnahme an der Studie zu gewinnen. In allen vier Bundesländern wurden aber Interviews an Grundschulen, Schulen mit mehreren Bildungsgängen, wie z. B. der Realschule Plus in Rheinland-Pfalz, und Gymnasien geführt. Alle geführten Interviews wurden zunächst inhaltsanalytisch auf Basis von Abhören unter der Fragestellung ausgewertet, welche Nebenfolgen von den Interviewpartnern benannt wurden. Relevante Interviewpassagen wurden in paraphrasierter Form aufgeschrieben. Die Textdateien wurden anschließend mithilfe der Analysesoftware MAXQDA kodiert. Paraphrasiert wurden zunächst alle Interviewpassagen, in denen die Interviewten Ergebnisse als Effekte Neuer Steuerungsinstrumente darstellten. Die Klassifikation der

2 Der Interviewleitfaden und eine Auswahl an Interviewtranskripten sind unter <http://www.uni-muenster.de/EW/forschung/projekte/nefo/materialien.html> veröffentlicht.

jeweiligen Phänomene als Nebenfolgen erfolgte diskursiv im Projektteam entsprechend der Nebenfolgendefinition (siehe Kapitel 2). Ziel dieses Auswertungsschrittes war es, die von Bellmann & Weiß (2009) auf Basis der internationalen Literatur entwickelte Nebenfolgenphänomenologie für den deutschen Kontext zu überprüfen und ggf. zu ergänzen. In einem zweiten Schritt wurde pro Bundesland eine Schule jeder Schulform für eine vertiefende *rekonstruktive* Analyse ausgewählt. Alle Interviews der ausgewählten Schulen wurden vollständig transkribiert und mithilfe der dokumentarischen Methode ausgewertet.

Bereits die *inhaltsanalytische* Auswertung der Interviews kann zeigen, dass nahezu alle aus der internationalen Literatur bekannten „unintended consequences“ der Neuen Steuerung auch von den interviewten Akteuren in den Schulsystemen der untersuchten Bundesländer berichtet bzw. beobachtet werden.

So wurde uns im Hinblick auf die Ebene des Unterrichts berichtet, dass zum Teil über Wochen und Monate auf Aufgaben und Aufgabenformate von zentralen Vergleichsarbeiten vorbereitet wird oder in Einzelfällen sogar Testergebnisse geschönt werden. Dies führt dann dazu, dass die Testergebnisse nicht den tatsächlichen Leistungsstand der Schülerschaft widerspiegeln. Einige Akteure berichten auch davon, dass das Unterrichtsniveau sinkt, da eine Anpassung an das Niveau der Testaufgaben vorgenommen wird.

Auch im Hinblick auf die Ebene der Schule gibt es Risiken und Nebenwirkungen der Neuen Steuerung. So hat die zunehmende Bedeutung von Leistungsvergleichen und Wettbewerb zwischen Schulen dazu geführt, dass Schulen vermehrt versuchen, eine besonders starke Schülerschaft zu rekrutieren und schwache Schülerinnen und Schüler, soweit möglich, abzulehnen. So kann zwischen Schulen statt des gewünschten Qualitätswettbewerbs auch ein Verdrängungswettbewerb um Schülerinnen und Schüler entstehen. Auch wird berichtet, dass die zunehmenden Verpflichtungen zur Dokumentation von Leistungen sowie zur Erarbeitung von Förderprogrammen und Schulentwicklungskonzepten oftmals zulasten des Unterrichts gehen. Es wird berichtet, dass in der Vorbereitung auf eine Schulinspektion das Erscheinungsbild der Schule kurzfristig optimiert wird, um nach außen gut dazustehen, obwohl dieses Bild nicht dem Alltag entspricht. Schulleitungen bringen zudem zum Ausdruck, dass ihre Rolle zunehmend mit der Ausübung von Kontrolle und Druck verbunden ist. Zudem geht es nach Einschätzung mancher Schulleiterinnen und Schulleiter und Lehrerinnen und Lehrer im Umgang mit den neuen Steuerungsinstrumenten oftmals gar nicht darum, Schule besser zu machen, sondern nur das zu erfüllen, was vorgeschrieben ist.

Sehr häufig wird auch von Nebenfolgen berichtet, die primär die Einstellungen von Akteuren betreffen. Neu hinzukommende Aufgaben für Lehrkräfte (Dokumentationsaufgaben, Statistiken etc.) und Schulleitungen (Öffentlichkeitsarbeit und repräsentative Tätigkeiten, Management- und Verwaltungstätigkeiten, Anbahnung und Ausgestaltung von Kooperationen) werden teilweise als nicht kompatibel mit den Anforderungen des Lehrerberufs erachtet. Damit kann ein Gefühl der eingeschränkten Selbstbestimmung durch verstärkte Kontrollen verbunden sein. Ein Teil der Lehrerschaft vergleicht sich anhand von Testergebnissen, vermeidet aber die Diskussion im Kollegium. So besteht die Gefahr, dass Kooperation verhindert und Kon-

kurrenzdenken verschärft wird. Sehr häufig wird auch geschildert, dass die neuen Instrumente der Qualitätsentwicklung mit Überforderung, Angst, Stress, Frustration und Verunsicherung verbunden sind. Einzelne Reforminstrumente werden teilweise auch als Misstrauensvotum vonseiten der Bildungspolitik und -administration betrachtet. Kontrolle, Bevormundung und Auflagen können die Freiheit der Akteure einschränken und so ihrerseits ein Misstrauen gegenüber der Bildungspolitik und -administration hervorrufen.

Im *rekonstruktiven* Teil der Interviewstudie wurde nun ein Ansatz entwickelt, der das Auftreten von Nebenfolgen Neuer Steuerung auch unter Bedingungen von „low-stakes“ oder „no-stakes“ erklären kann. Im Mittelpunkt dieses Ansatzes stehen habituelle Faktoren, nämlich sechs mithilfe der dokumentarischen Methode (vgl. Nohl 2009) rekonstruierte sog. Reformrezeptionstypen, die wir als Typus der Profession, der Pragmatik, der Kritik, der Innovation, der Vorgabe und der Performanz bezeichnet haben. Grundsätzlich gilt, dass die Neue Steuerung im Feld der Schule auf differente Deutungsmuster der Lehrkräfte und Schulleitungen trifft, die nicht nur die Wahrnehmung der Neuen Steuerung vorstrukturieren, sondern auch den Umgang mit dieser, einschließlich der möglichen „Produktion“ von Nebenfolgen. Diese Typen wurden themen- und interviewübergreifend rekonstruiert und sind dementsprechend nicht deckungsgleich mit einzelnen Lehrerinnen und Lehrern. Aufgrund des nur beschränkten Umfangs des Beitrags können die einzelnen Typen an dieser Stelle nicht umfassend dargestellt werden. Eine ausführlichere Darstellung des Typus der Profession und der Performanz findet sich aber bei Bellmann et al. (angenommen).

Der Typus der *Profession* ist gekennzeichnet durch eine starke Abgrenzung des eigenen pädagogisch-professionellen Kompetenzbereiches gegenüber der Einflussnahme von Personen, die selbst keine Lehrkräfte sind. Auf Grundlage der durch langjährige Praxiserfahrung erlangten Kompetenz werden selbstbewusst Ziele und Standards für die eigene unterrichtliche Arbeit gesetzt. Als Erfolgsindikator für die eigene Arbeit gilt der erfolgreiche Übergang der Schülerinnen und Schüler in weiterführende Schulen oder das Erwerbsleben. Einher gehen diese Handlungsorientierungen mit einer geringen Empfänglichkeit für die Anreizstrukturen Neuer Steuerung. Zentrale Vergleichsarbeiten und Schulinspektionen werden zwar durchgeführt, für die Veränderung von Unterricht bleiben sie aber ohne Bedeutung. Auch die Produktion von Nebenfolgen bleibt weitgehend aus.

Für den Typ der *Pragmatik* ist die pragmatische, effiziente und effektive Bearbeitung situativer Handlungsanforderungen, u. a. auch der neuen Steuerungsinstrumente, kennzeichnend. Dabei wird zwischen Anforderungen sortiert, die aktiv bearbeitet werden können, und solchen, auf die nicht reagiert werden muss. Hierbei sind Überlegungen zur Umsetzbarkeit von Maßnahmen unter den Bedingungen alltäglichen schulischen Handelns von zentraler Bedeutung. Während im Typus der Profession die Ergebnisse externer Evaluation vor dem Hintergrund der eigenen professionellen Handlungskompetenz als überflüssig erfahren werden, entscheidet im Muster der Pragmatik der Abgleich der gestellten Anforderungen mit den Gegebenheiten schulischen Handelns über die Bedeutung, die den Instrumenten Neuer Steuerung im Schulalltag zugewiesen wird. Ob und inwiefern diese Instrumente

zur Unterrichtsveränderung führen, ist abhängig von den Reflexionsleistungen von Schulleitungen und Lehrkräften. Diese Vermittlungsleistung des pädagogisch-professionellen Personals scheint auch hier die Produktion von Nebenfolgen weitestgehend zu verhindern.

Zentral für den Typus der *Kritik* ist eine starke Orientierung an einzelfallspezifischen Besonderheiten, sei es einzelner Schülerinnen und Schüler oder der Einzelschule. Die Instrumente der Outputsteuerung werden als Versuche rezipiert, die Schülerschaft und schulisches Handeln im Sinne einer Normierung zu standardisieren, was der Ausrichtung dieses Typus am Einzelfall diametral gegenübersteht. Anders als im Typus der Profession gelingt es jedoch hier nicht, einen eigenen pädagogisch-professionellen Kompetenzbereich selbstbewusst abzugrenzen, sodass der Typus der Kritik insbesondere Bildungsstandards, zentrale Vergleichsarbeiten und Schulinspektionen als problematische Übergriffe erlebt. Weder Anpassungs- noch Ausweichverhalten können mit diesem Typus systematisch in Verbindung gebracht werden. Auf Einstellungsebene finden sich aber Nebenfolgen wie Deprofessionalisierung, ein Gefühl der Verletzung professioneller Integrität und psychische Kosten.

Für den Typus der *Innovation* ist eine generelle Orientierung an Schul- und Unterrichtsentwicklung charakteristisch. Die Instrumente Neuer Steuerung fügen sich unproblematisch in diese Orientierung ein. Sie gewinnen in diesem Rezeptionsmuster vorrangig als ein Mittel an Bedeutung, durch das diejenigen Akteure, die als weniger engagiert und innovativ wahrgenommen werden, unter Druck geraten. Auf Basis der Darstellungen der interviewten Lehrkräfte und Schulleitungen zu Phänomenen, die im Rahmen des Nefo-Projektes als Nebenfolgen klassifiziert wurden, lässt sich sagen, dass solche Nebenfolgen, die sich als Anpassungsverhalten beschreiben lassen, im Typus der Innovation aktiv produziert und kaum als problematisch wahrgenommen werden. Demgegenüber besteht ein hohes Problembewusstsein in Bezug auf Nebenfolgen, die sich als Ausweichverhalten klassifizieren lassen. Nebenfolgen auf Einstellungsebene treten in Zusammenhang mit diesem Reformrezeptionsmuster eher nicht auf.

Kennzeichnend für den Typus der *Vorgabe* ist ein grundsätzlicher Vertrauensvorschuss gegenüber bildungspolitischem Handeln. Administrative und bildungspolitische Vorgaben werden genau befolgt, da davon ausgegangen wird, dass diese dem Wohl der Schülerschaft dienen. Da Handeln hier erst auf spezifische Vorgaben und Aufforderungen erfolgt, kommen Rückmeldungen aus zentralen Vergleichsarbeiten und Schulinspektionen besondere Bedeutung zu. Entsprechend der in diesen Rückmeldungen aufgelisteten Defizite wird das schulische Handeln angepasst. Unabhängig von konkreten und punktuellen Vorgaben und Aufforderungen wird – anders als im Typus der Innovation – keine Unterrichtsentwicklung betrieben. Da hier davon ausgegangen wird, dass die Anpassung von Unterricht in Reaktion auf Evaluationsergebnisse grundsätzlich zu einer Verbesserung von Unterricht führt, bleiben zwar die Formen des Ausweichverhaltens weitestgehend aus, Anpassungsverhalten ist dagegen erwartbar. Auf Einstellungsebene steht das Reformrezeptionsmuster der Vorgabe in enger Verbindung mit der Entwicklung einer Abhängigkeit vom Expertenurteil.

Im Typus der *Performanz* werden die Instrumente Neuer Steuerung schließlich als Instrumente bedeutsam, durch die Leistung intersubjektiv sichtbar wird.

Die Instrumente werden dementsprechend in erster Linie als Vergleichsinstrumente und nicht als Instrumente der Qualitätsentwicklung rezipiert. Was zählt, ist das gute Abschneiden im Vergleich mit anderen. Demgegenüber spielt die Ausrichtung des eigenen Handelns an den Erfordernissen konkreter Unterrichtssituationen eine untergeordnete Rolle. Diesem Reformrezeptionsmuster lassen sich auf Basis der Interviewdarstellungen Formen des Anpassungs- wie auch des Ausweichverhaltens zuordnen. Dies lässt sich beispielsweise an den Textpassagen zum sog. „Window Dressing“ zeigen. An Stellen, an denen Interviewpartner darlegen, während der Schulinspektion den eigenen Unterricht anders zu gestalten als im Schulalltag, lässt sich auch die Rahmung der Steuerungsinstrumente als Instrumente, über die Leistung für andere sichtbar wird, aufzeigen. Insofern das Sichtbarwerden der eigenen Leistung als problematisch erlebt wird, ist diese Art der Reformrezeption auf Einstellungsebene mit psychischen Kosten verknüpft.

In Anschluss an die Rekonstruktion der Reformrezeptionstypen lässt sich ein Erklärungsansatz für das Auftreten von Nebenfolgen formulieren:

Erstens lassen sich die sechs Reformrezeptionsmuster danach sortieren, inwiefern zwischen innerschulischen Handlungsanforderungen einerseits und Anforderungen Neuer Steuerung andererseits unterschieden wird. Wir sprechen hier von „Differenzorientierung“. Jene Muster, denen eine solche Differenzorientierung zugesprochen werden kann (Typus der Profession, Typus der Pragmatik, Typus der Kritik), stehen kaum mit der Produktion anreizkonformen Anpassungsverhaltens in Verbindung.

Zur Erklärung des Auftretens eines anreizwidrigen Ausweichverhaltens gilt es, zweitens, die Rede von „stakes“ im Rahmen der Neuen Steuerung zu differenzieren. Sowohl Output- als auch Wettbewerbssteuerung bringen spezifische Anreize ins Spiel, die in einem grundsätzlichen Sinne als „stakes“ verstanden werden können. Gemeinsam ist diesen Anreizen, dass sie aus der Umwelt der Schule kommen, entweder in Form von Rückmeldungen aus zentralen Vergleichsarbeiten und Schulinspektionen oder in Form von Signalen des Marktumfelds der Schule. Obwohl diese „stakes“ durchaus einen „objektiven“ Charakter haben, sind sie nicht für alle Reformrezeptionstypen gleichermaßen handlungsrelevant. Es erscheint deshalb sinnvoll, ähnlich wie in der Forschung zu Wettbewerb im Bildungsbereich (vgl. Levacic & Woods 2000), zwischen objektiven (strukturellen) und subjektiven (wahrgenommenen) „stakes“ zu unterscheiden. Auf Grundlage dieser Unterscheidung wird erklärbar, warum ein und dieselbe Anreizstruktur von unterschiedlichen Akteuren unterschiedlich entziffert wird. Während neue Steuerungsinstrumente für den Typus der Profession möglicherweise gar keine „stakes“ ins Spiel bringen, kann der Typus der Performanz diese sehr wohl als „stakes“ betrachten. Die Reaktion auf diese subjektiv wahrgenommenen „stakes“ können dann Verhaltensweisen sein, die als anreizwidriges Ausweichverhalten beschrieben werden können.

5. Quantitative Erhebung

In der zweiten Projektphase wurde von Mai bis Oktober 2013 eine quantitative Fragebogenerhebung durchgeführt. Ziel war es zu erfassen, wie weit die aus der Forschungsliteratur und den geführten Interviews bekannten Nebenfolgen Neuer Steuerung verbreitet sind. Basis der Untersuchung bildete eine vom Data Processing Center (Hamburg) gezogene geschichtete Zufallsstichprobe. Angestrebt war es, 5 % aller Grundschulen und 10 % aller weiterführenden Schulen in die Untersuchung einzubeziehen. Anders als im Fall der qualitativen Erhebung wurden in die quantitative Datenerhebung auch Haupt-, Real- und Gesamtschulen einbezogen. An allen gezogenen Schulen sollte das gesamte Kollegium mittels des Kernfragebogens befragt werden. Für jede Samplingschule hat das DPC vier Ersatzschulen nachgezogen, auf die zurückgegriffen werden konnte, verweigerten die gewünschten Samplingschulen die Teilnahme. Angaben zur Anzahl rückgesendeter Fragebögen wie der Rücklaufquote auf Schul- und Lehrkräfteebene nach Replacement finden sich in Tabelle 1:

Bundesland	Anzahl ausgefüllt rückgesendeter Kontextfragebögen	Anzahl ausgefüllt rückgesendeter Kernfragebögen	Rücklaufquote auf Schulebene nach Replacement	Rücklaufquote auf Lehrkräfteebene	Teilnehmerrate nach Replacement
Berlin	26	513	64,6 %	39,4 %	25,7 %
Brandenburg	40	578	83,3 %	50,3 %	41,9 %
Rheinland-Pfalz	77	1031	75,9 %	55,8 %	42,3 %
Thüringen	42	515	83,7 %	44,9 %	37,6 %

Der Rücklauf auf Schul- und Lehrkräfteebene kann für eine auf Freiwilligkeit basierende Befragung im Bereich der Schulforschung durchaus als zufriedenstellend eingeschätzt werden. Keine der bundeslandspezifischen Teilnehmerraten entspricht jedoch internationalen Qualitätsstandards³, sodass von erheblichen Verzerrungseffekten der Analyseergebnisse ausgegangen werden muss. Besonders problematisch erscheint die Teilnehmerrate des Bundeslandes Berlin, die mit 25,7 % weitaus geringer ist als die Teilnehmerraten der Vergleichsländer. Auch wenn die Auswertungsergebnisse keine sicheren Schlüsse auf die exakte Verbreitung von Nebenfolgen zulassen, so ermöglichen sie doch einen ersten explorativen Zugriff auf das Thema Nebenfolgen unter „Low- und No-Stakes“-Bedingungen.

³ Entsprechend den Qualitätsstandards der International Association for the Evaluation of Educational Achievement (IEA) gelten nur Studien mit einer Gesamtteilnehmerrate von 75% als glaubwürdig.

Entsprechend dem explorativen Charakter des Projektes war es Ziel der Fragebogenkonstruktion⁴, möglichst alle gefundenen Nebenfolgen zu operationalisieren. Mit dem Fokus auf die Erfassung der Bandbreite der Nebenfolgen geht notgedrungen einher, dass nicht alle Dimensionen der einzelnen Nebenfolgen erfasst werden konnten. Dies führt zu Einschränkungen hinsichtlich der Validität und Reliabilität der Messung. Für einen ersten Zugriff auf die Frage, inwiefern Nebenfolgen auch unter „Low- und No-Stakes“-Bedingungen auftreten, scheint die zunächst breit angelegte Erfassung von Nebenfolgen aber vielversprechender als die genaue Erhebung einzelner Nebenfolgen. Die verwendeten Fragebogenitems wurden im Projekt entwickelt und mittels eines Reviews durch drei Experten, kognitive Pretests mit Lehrkräften sowie einen quantitativen Pretest an zehn Schulen in Rheinland-Pfalz getestet, bevor sie in der Hauptstudie eingesetzt wurden.

5.1 Ergebnisse

Im Nefo-Projekt wurden 93 Nebenfolgenitems eingesetzt. Da an dieser Stelle nicht alle Antwortverteilungen präsentiert werden können, sind in der folgenden Tabelle (Tabelle 2) die Antwortverteilungen für eine Auswahl von Nebenfolgen dargestellt. Ziel bei der Auswahl der Nebenfolgenitems war es, sowohl Ergebnisse zu Nebenfolgen auf Ebene des Verhaltens als auch auf der Ebene der Einstellungen schulischer Akteure zu präsentieren. In Bezug auf die Nebenfolgen auf Ebene des Verhaltens schulischer Akteure wurden noch einmal solche Nebenfolgenitems, die von der Mehrheit der Lehrkräfte über alle vier Bundesländer hinweg zustimmend beantwortet wurden (anreizkonformes Anpassungsverhalten), und solche, die nur bei einer Minderheit Zustimmung fanden (anreizwidriges Ausweichverhalten), unterschieden. Pro Gruppe wurden jeweils Items mit eher hoher und Items mit eher niedriger Zustimmung ausgewählt, um einen Überblick über die Spanne der Wahrnehmung von Nebenfolgen zu geben.

Um die statistische Signifikanz der Unterschiede zwischen den Bundesländern zumindest ansatzweise einschätzen zu können, wurden die Antwortkategorien „Trifft voll zu“ und „Trifft eher zu“ sowie die Kategorien „Trifft eher nicht zu“ und „Trifft überhaupt nicht zu“ zusammengefasst. Im Anschluss hieran wurde mithilfe logistischer Regressionen in Mehrebenenmodellen überprüft, ob sich die Antwortverteilungen für Berlin, Brandenburg und Rheinland-Pfalz jeweils von der thüringischen Antwortverteilung statistisch signifikant unterscheiden. Thüringen wurde als Referenzland ausgewählt, weil eine überblicksartige Betrachtung des bundeslandspezifischen Antwortverhaltens zu denjenigen Nebenfolgen, die als anreizwidriges Ausweichverhalten verstanden werden können, vermuten ließ, dass sich das Antwortverhalten der thüringischen Lehrkräfte vom Antwortverhalten der Lehrkräfte in den anderen drei Bundesländern systematisch unterscheidet. Zudem unterschei-

4 Die quantitativen Erhebungsinstrumente sind unter <http://www.uni-muenster.de/EW/forschung/> veröffentlicht. Zudem wurden die Datensätze zur Herstellung von Transparenz und intersubjektiver Nachvollziehbarkeit an das Leibniz-Institut für Sozialwissenschaften (GESIS) übermittelt. Sie können dort ab 15.08.2015 auf Antrag abgerufen und für Sekundäranalysen genutzt werden.

det sich das thüringische Steuerungsregime maßgeblich von den Regimen in Berlin, Brandenburg und Rheinland-Pfalz durch die Einbindung neuer Steuerungsinstrumente in den Mechanismus professioneller Selbststeuerung.

Aufgrund nur äußerst geringer Zustimmungsraten konnte kein Signifikanztest für das Item, das die Nebenfolge „Optimierung des Inspektionspools“ operationalisiert, durchgeführt werden.

Tabelle 2: Ergebnisse zu ausgewählten Nebenfolgen (vgl. Text); Angaben in %; in Klammern: Standardfehler; *statistisch signifikante Unterschiede ($p < 0,05$) zwischen den einzelnen Bundesländern und Thüringen						
Nebenfolge	BL	Antwortmöglichkeiten				N
		Trifft voll zu	Trifft eher zu	Trifft eher nicht zu	Trifft überhaupt nicht zu	
I. Nebenfolgen auf Ebene des Verhaltens schulischer Akteure						
Anreizkonformes Anpassungsverhalten						
Coaching Nutzen Sie zur Vorbereitung auf VERA Aufgaben aus alten VERA-Testheften?	BE	25,1 (0,04)	41,5 (0,04)	19,1 (0,03)	14,3 (0,03)	229
	RP	18,5 (0,03)	40,7 (0,03)	17,8 (0,02)	23,0 (0,03)	498
	TH	26,2 (0,03)	42,5 (0,03)	21,2 (0,03)	10,1 (0,02)	341
	BB*	39,7 (0,03)	42,6 (0,03)	9,2 (0,02)	8,4 (0,02)	325
Kurzsichtigkeit Ich habe den Eindruck, dass in den letzten Jahren der Unterricht an der kurzfristigen Erfüllung von definierten Qualitätskriterien orientiert ist.	BE*	27,6 (0,03)	48,0 (0,03)	22,6 (0,02)	1,9 (0,01)	464
	RP*	15,8 (0,01)	54,0 (0,02)	26,4 (0,02)	3,8 (0,01)	923
	TH	11,7 (0,02)	44,1 (0,03)	38,9 (0,03)	5,3 (0,02)	460
	BB*	18,7 (0,02)	52,3 (0,03)	25,0 (0,02)	4,0 (0,01)	532
Unterrichtsassimilation Ich habe den Eindruck, dass in den letzten Jahren die Orientierung an langfristigen Bildungsprozessen der Schüler und Schülerinnen in den Hintergrund getreten ist.	BE*	20,9 (0,02)	46,4 (0,03)	30,3 (0,02)	2,4 (0,01)	466
	RP	13,4 (0,01)	42,5 (0,02)	39,2 (0,02)	4,8 (0,01)	933
	TH	11,9 (0,02)	40,4 (0,03)	43,1 (0,03)	4,6 (0,01)	466
	BB*	17,1 (0,02)	48,8 (0,03)	30,3 (0,03)	3,8 (0,01)	528

Nebenfolge	BL	Antwortmöglichkeiten				N
		Trifft voll zu	Trifft eher zu	Trifft eher nicht zu	Trifft überhaupt nicht zu	
I. Nebenfolgen auf Ebene des Verhaltens schulischer Akteure						
Anreizwidriges Ausweichverhalten						
Window Dressing Während der Schulinspektion zeigen Lehrkräfte unserer Schule andere Unterrichtsmethoden als im normalen Schulalltag.	BE*	13,6 (0,02)	33,8 (0,04)	42,4 (0,04)	10,3 (0,02)	432
	RP*	13,0 (0,01)	39,3 (0,02)	36,8 (0,02)	11,0 (0,02)	803
	TH	5,1 (0,01)	24,5 (0,03)	53,4 (0,03)	17,0 (0,02)	419
	BB*	9,0 (0,02)	26,3 (0,03)	51,0 (0,03)	13,7 (0,02)	452
Betrug Wenn einzelne Schüler und Schülerinnen während eines VERA-Tests nicht weiterkommen, gibt man auf Verständnisfragen von Schülerinnen und Schülern hin schon mal Hilfestellung.	BE*	3,6 (0,01)	17,6 (0,04)	40,6 (0,03)	38,1 (0,05)	216
	RP*	2,1 (0,00)	19,2 (0,03)	37,1 (0,02)	41,7 (0,03)	486
	TH	0,7 (0,00)	10,9 (0,03)	32,6 (0,03)	55,9 (0,04)	318
	BB*	1,8 (0,01)	19,7 (0,03)	38,3 (0,03)	40,2 (0,05)	305
Optimierung des Inspektionspools Es kommt vor, dass einzelnen Lehrkräften nahegelegt wird, an den Tagen der Schulinspektion nicht zur Schule zu kommen.	BE	0,0 (0,0)	2,0 (0,01)	8,9 (0,02)	89,1 (0,02)	392
	RP	0,2 (0,0)	0,3 (0,0)	6,9 (0,01)	92,6 (0,01)	763
	TH	0,0 (0,0)	0,0 (0,0)	4,7 (0,02)	95,3 (0,02)	439
	BB	0,4 (0,0)	0,3 (0,0)	6,3 (0,01)	93,1 (0,01)	444
II. Nebenfolgen auf Ebene der Einstellungen schulischer Akteure						
Psychische Kosten Ich habe den Eindruck, dass in den letzten Jahren der Druck gestiegen ist, gute Testergebnisse erzielen zu müssen.	BE*	30,0 (0,04)	51,3 (0,04)	16,8 (0,02)	1,9 (0,00)	482
	RP*	17,9 (0,01)	55,6 (0,02)	22,2 (0,02)	4,3 (0,01)	960
	TH	13,7 (0,02)	50,9 (0,03)	31,8 (0,02)	3,6 (0,01)	489
	BB*	28,9 (0,03)	48,9 (0,03)	19,8 (0,03)	2,3 (0,01)	547

Nebenfolge	BL	Antwortmöglichkeiten				N
		Trifft voll zu	Trifft eher zu	Trifft eher nicht zu	Trifft überhaupt nicht zu	
Deprofessionalisierung Ich habe den Eindruck, dass die Festlegung von Lernzielen in den letzten Jahren zunehmend zu einer Aufgabe außerschulischer Instanzen geworden ist.	BE*	23,7 (0,02)	54,3 (0,02)	19,3 (0,02)	2,7 (0,01)	468
	RP	21,6 (0,02)	47,3 (0,02)	26,6 (0,02)	4,5 (0,01)	953
	TH	19,9 (0,02)	41,9 (0,03)	32,1 (0,03)	6,1 (0,01)	475
	BB	32,7 (0,02)	46,5 (0,02)	25,5 (0,03)	4,4 (0,01)	537
Abhängigkeit vom Expertenurteil Die Rückmeldung aus VERA gibt mir Orientierung in der Ausübung meines Berufs.	BE*	1,8 (0,01)	15,5 (0,02)	34,5 (2,8)	48,2 (0,03)	233
	RP*	2,3 (0,01)	22,1 (0,02)	37,1 (0,03)	38,5 (0,03)	507
	TH	10,7 (0,03)	43,0 (0,03)	34,1 (0,02)	12,3 (0,02)	348
	BB*	9,2 (0,03)	30,1 (0,03)	37,0 (0,03)	23,7 (0,03)	323
Konkurrenzdenken im Kollegium Die Rückmeldung von Ergebnissen klassenübergreifender Tests führt zu Konkurrenzdenken im Kollegium.	BE	2,2 (0,01)	9,4 (0,02)	42,6 (0,03)	45,8 (0,03)	430
	RP*	2,8 (0,01)	13,7 (0,02)	42,4 (0,02)	41,2 (0,02)	863
	TH	1,2 (0,01)	7,1 (0,02)	42,8 (0,03)	48,9 (0,04)	443
	BB	2,6 (0,01)	10,8 (0,02)	39,4 (0,03)	47,2 (0,04)	491

5.2 Interpretation der Ergebnisse

5.2.1 Nebenfolgen: Verhalten der Akteure

Nimmt man die Wahrnehmung der Lehrkräfte hinsichtlich des Auftretens von Nebenfolgen zum Ausgangspunkt, lässt sich auf Basis der Fragebogenerhebung im Rahmen des Nefo-Projektes festhalten, dass die verschiedenen Nebenfolgen unterschiedlich weit verbreitet sind. Die verschiedenen Nebenfolgen lassen sich zunächst entsprechend ihrem Verbreitungsgrad in zwei Gruppen einteilen: Nebenfolgen, die von der Mehrzahl, d. h. mehr als 50 %, der Lehrkräfte in den vier untersuchten Bundesländern wahrgenommen werden, und solche, die von weniger als 50 % der

Lehrkräfte beobachtet werden. In Anschluss hieran lässt sich auf theoretischer Ebene fragen, was die Nebenfolgen der beiden Gruppen jeweils gemeinsam haben? Es zeigt sich, dass sich die zuerst genannte Gruppe Nebenfolgen als anreizkonformes Anpassungsverhalten beschreiben lässt und die zweite Gruppe als anreizwidriges Ausweichverhalten. In ähnlicher Weise unterscheiden van Thiel und Leeuw (2002) zwischen einem „Unintended Performance Paradox“ und einem „Deliberate Performance Paradox“. Entsprechend müssen für beide Gruppen von Nebenfolgen unterschiedliche Erklärungsansätze gefunden werden.

Anreizkonformes Anpassungsverhalten

Jene Nebenfolgen, die im Rahmen der Nefo-Studie von der Mehrzahl der Lehrkräfte wahrgenommen wurden, korrespondieren mit bildungspolitisch vermittelten Deutungsangeboten und Handlungsoptionen. Insofern lässt sich in Bezug auf diese Nebenfolgen von anreizkonformem Anpassungsverhalten sprechen. Unterschieden werden können solche Deutungsangebote und Handlungsoptionen, die bereits im neuen Steuerungsmodell selbst angelegt sind, und solche, die vonseiten der Bildungspolitik und -administration aktiv als gewünschtes Verhalten nahegelegt werden.

Ein Beispiel für Ersteres bildet die Ausrichtung schulischer und unterrichtlicher Arbeit an der kurzfristigen Erfüllung definierter Qualitätskriterien, wobei gleichzeitig langfristige Bildungsprozesse der Schülerinnen und Schüler als entscheidende Zielpunkte schulischen Handelns in den Hintergrund treten. Das Item („Ich habe den Eindruck, dass in den letzten Jahren der Unterricht an der kurzfristigen Erfüllung von definierten Qualitätskriterien erfüllt ist“), das diese Nebenfolge zum Teil operationalisiert⁵, hat in der Nefo-Studie sehr hohe Zustimmung erhalten. In Berlin stimmten insgesamt 75,6 % der Lehrkräfte dem Item entweder „voll“ oder „eher“ zu, in Brandenburg 71,0 %, in Rheinland-Pfalz 69,8 % und in Thüringen 55,8 %. Gleichzeitig nimmt die Mehrheit der Lehrkräfte in allen vier Bundesländern wahr, dass die Orientierung an langfristigen Bildungsprozessen der Schülerinnen und Schüler in den Hintergrund getreten ist. Dies beobachten in Berlin 67,3 %, in Brandenburg 65,9 %, in Rheinland-Pfalz 55,9 % und in Thüringen 52,3 % der Lehrkräfte. Die breite Zustimmung zu den beiden Items lässt sich als Hinweis dafür deuten, dass mit der Einführung neuer Steuerungsinstrumente Schule und Unterricht weniger auf die Erreichung von Outcome, d. h. die langfristigen und eher indirekten Effekte von Unterricht wie beispielsweise gesellschaftliche Teilhabe (vgl. Chapman 2008, S. 259; Schreyer 2012, S. 259), ausgerichtet sind, sondern auf den durch Qualitätsindikatoren operationalisierten Output, d. h. die unmittelbaren Effekte von Lehr-/Lernprozessen wie beispielsweise Schülerleistung, Einstellungen und Fähigkeiten. Dies stellt sich insbesondere in Feldern problematisch dar, in denen ein deutlicher Abstand zwischen Output und Outcome besteht. Je größer dieser Abstand ist, desto größer ist die

5 Wenngleich mit diesem wie auch ähnlichen Items nach einer Einschätzung und nicht nach konkreten Verhaltensweisen gefragt wird, gehen wir davon aus, dass eine Zustimmung zu diesen Items auf ein Verhalten schließen lässt, das die befragten Lehrkräfte beobachten. Insofern dienen auch diese Items als Indikator für die jeweils erfragte Nebenfolge.

Gefahr, dass das Erreichen von Output, anstelle von Outcome, zum entscheidenden Orientierungspunkt wird (vgl. Shavelson & Huang 2003, S. 12). Allein durch die Messung von Output rückt dieser im praktischen Handeln in den Fokus, was in Kapitel 2 im Anschluss an Etzioni (1964) als Grundproblem indikatorengestützter Steuerung bezeichnet wurde und gegenwärtig als „what’s measured is what matters“ (z. B. Bevan & Hood 2006) diskutiert wird.

Neben solchen Deutungsangeboten und Handlungsoptionen, die mit dem Steuerungsmodell selbst transportiert werden, adaptieren Lehrkräfte und Schulleitungen auch solche Verhaltensweisen, die von Bildungspolitik und -administration in gewisser Weise aktiv als gewünschtes Verhalten nahegelegt werden. Hierfür kann Coaching mittels alter Testaufgaben als Beispiel dienen. In allen vier Untersuchungsländern gibt die Mehrzahl der Lehrkräfte an, selbst auf VERA-Tests vorzubereiten, wobei die Anzahl der Lehrkräfte, die dies angeben, in Brandenburg am höchsten ist. So nutzen 82,3 % der Lehrkräfte in Brandenburg zur Vorbereitung auf VERA-Aufgaben aus alten Testheften. Die Anzahl der Lehrkräfte in Thüringen (68,7 %), Berlin (66,6 %) und Rheinland-Pfalz (59,2 %) liegt deutlich darunter. Dass Coaching als bildungspolitisch erwünscht gelten kann, zeigt sich schon daran, dass das Institut für Qualitätsentwicklung im Bildungswesen (IQB) eine öffentlich zugängliche Datenbank mit alten Testaufgaben (vgl. Institut zur Qualitätsentwicklung im Bildungswesen) zur Verfügung stellt und sowohl in Verwaltungsvorschriften (vgl. Senatsverwaltung für Bildung, Jugend und Wissenschaft 2013) und Rundschreiben an die Schulen (vgl. Ministerium für Bildung, Jugend und Sport 2014) als auch auf den Webseiten von Instituten wie dem Institut für Schulqualität (ISQ) auf die Aufgabensammlung des IQB hingewiesen wird (vgl. Institut für Schulqualität der Länder Berlin und Brandenburg e. V.). Aus testtheoretischer Perspektive erscheint Testvorbereitung, die lediglich darauf zielt, Schülerinnen und Schüler mit dem Format der Testaufgaben bekannt zu machen, durchaus wünschenswert, da hierdurch die Validität der Messung erhöht werden kann. Eine Vorbereitung auf Testinhalte kann aber zu einer Verzerrung der Testergebnisse führen (vgl. Koretz 2005). Aus pädagogischer Sicht kann der Einsatz von Testaufgaben im Unterricht deshalb problematisch sein, weil sich Testaufgaben und didaktische Aufgaben in ihrer Funktion unterscheiden. An didaktischen Aufgaben sollen Schülerinnen und Schüler etwas lernen, Testaufgaben sollen zeigen, was gelernt wurde. Dementsprechend sind an die Konzeption von didaktischen Aufgaben andere Anforderungen gestellt als an die Konzeption von Testaufgaben. Eine Verwendung von Testaufgaben als didaktische Aufgaben ist deshalb nicht ohne Weiteres möglich (vgl. Benner 2007; Blömeke 2009).

Anreizwidriges Ausweichverhalten

Von dieser mit Anpassungsverhalten verbundenen Gruppe von verbreiteten Nebenfolgen lässt sich eine Gruppe weniger verbreiteter Nebenfolgen unterscheiden, die mit Formen von Ausweichverhalten gegenüber den Ansprüchen und Anforderungen Neuer Steuerung verbunden ist. Hierzu zählen auf Ebene des Unterrichts Sicherheitsdenken von Lehrkräften als Folge des Drucks neuer Steuerungsregime, aktive

Verstöße gegen Verfahrensvorschriften zu VERA oder der Schulinspektion sowie die Vorbereitung der Schülerinnen und Schüler auf VERA mithilfe spezieller Übungshefte, was über die bildungspolitisch nahegelegte Nutzung alter Testaufgaben hinausgeht. Auf Schulebene finden sich problematische Verhaltensweisen wie z. B. die Optimierung der Schülerpopulation durch die Selektion besonders leistungsstarker und die Ablehnung besonders leistungsschwacher Schülerinnen und Schüler, die Optimierung des Pools der Schülerschaft, die an zentralen Leistungstests teilnehmen z. B. durch den Ausschluss besonders leistungsschwacher Schüler und Schülerinnen sowie den Ausschluss von Lehrkräften aus der Schulinspektion. In diesen Fällen versucht man, dem Druck des Steuerungsregimes gewissermaßen strategisch auszuweichen, indem auf Verhaltensweisen zurückgegriffen wird, die bildungspolitisch weder forciert werden noch gewünscht sein können.

Die Zustimmung zu den jeweiligen Items variiert zwischen bundeslandübergreifend hoher Zustimmung zu dem Item „Während der Schulinspektion zeigen Lehrkräfte unserer Schule andere Unterrichtsmethoden als im normalen Schulalltag“ über Items mit eher geringer Zustimmung, wie z. B. „Wenn einzelne Schülerinnen und Schüler während eines VERA-Tests nicht weiterkommen, gibt man auf Verständnisfragen von Schülern und Schülerinnen hin schon mal Hilfestellung“, bis hin zu dem Item „Es kommt vor, dass einzelnen Lehrkräften nahegelegt wird, an den Tagen der Schulinspektion nicht zur Schule zu kommen“, das bundeslandübergreifend kaum Zustimmung erfahren hat. Das erstgenannte Item erhält die höchste Zustimmung vonseiten der rheinland-pfälzischen Lehrkräfte (52,3 %), gefolgt von Berlin (47,4 %), Brandenburg (35,3 %) und Thüringen (29,6 %). Dem an zweiter Stelle genannten Item stimmten 21,5 % der brandenburgischen, 21,3 % der rheinland-pfälzischen, 21,2 % der Berliner und 11,6 % der thüringischen Lehrkräfte zu. Auch wenn die Zustimmung zum drittgenannten Item äußerst gering ausfällt, so ist es doch bemerkenswert, dass es in Berlin, Rheinland-Pfalz und Brandenburg erwartungswidrig überhaupt Lehrkräfte gibt, die zustimmen, dass Lehrkräften nahegelegt wird, an Tagen der Schulinspektion nicht zur Schule zu kommen.

Über den Block von Nebenfolgen hinweg, die als Ausweichverhalten beschrieben werden können, ist zu beobachten, dass die Zustimmung der thüringischen Lehrkräfte zu den einzelnen Items jeweils am geringsten ausfällt. Lässt man den Schluss von den Selbstauskünften der Lehrkräfte auf den Grad der Verbreitung der Nebenfolgen zu, so sind Verhaltensweisen, die bildungspolitisch weder implizit noch explizit forciert werden, in Thüringen zwar auch, aber in geringerem Maße verbreitet als in den Vergleichsländern. Interpretiert man dies als Effekt bundeslandspezifischer Steuerungsregime, so lässt sich der Schluss ziehen, dass das thüringische Steuerungsregime Ausweichverhalten in geringerem Maße anreizt als die Regime in Berlin, Brandenburg und Rheinland-Pfalz. Im Unterschied zu diesen spielen in Thüringen Druckmomente, sei es durch administrative Kontrolle oder durch marktähnliche Strukturen, wenn überhaupt eine nur marginale Rolle. Stattdessen ist das thüringische Steuerungsmodell überwiegend auf eine professionelle Selbststeuerung angelegt.

In Anschluss hieran lässt sich die These formulieren, dass die bundeslandspezifische Ausgestaltung des Verhältnisses von staatlich-administrativer Steuerung,

marktlich-wettbewerblicher Steuerung und professioneller Selbststeuerung das Auftreten von Ausweichverhalten maßgeblich beeinflusst. Steuerungsmodelle, die primär auf datenbasierte staatlich-administrative Steuerung und Marktmechanismen setzen, reizen Ausweichverhalten schulischer Akteure stärker an als Modelle, die auf die Verarbeitung von Daten in Prozessen professioneller Selbststeuerung setzen.

5.2.2 Nebenfolgen: Einstellungen der Akteure

Neben spezifischen verhaltensbezogenen Nebenfolgen lassen sich auch auf Ebene der Einstellungen der Akteure im Schulsystem Nebenfolgen identifizieren. Anders als auf der Verhaltensebene kann bei den Nebenfolgen auf der Einstellungsebene nicht davon ausgegangen werden, dass es sich um Nebenfolgen handelt, die als Formen des Ausweichverhaltens mit Anforderungen Neuer Steuerung gelten können. Die Einstellungsdimension repräsentiert stärker grundsätzliche Haltungen und Wahrnehmungen der befragten Akteure in Reaktion auf neue Instrumente der Qualitätsentwicklung und rekurriert dementsprechend weniger auf Praktiken als Reaktionsformen auf bestimmte Instrumente. Theoretisch war bereits im Vorfeld der Untersuchung angenommen worden, dass Nebenfolgen auf der Einstellungsebene bereits im Modell der Neuen Steuerung angelegt sind. Es galt jedoch, in der quantitativen Erhebung zu prüfen, inwiefern diese theoretischen Annahmen auch durch die Wahrnehmung schulischer Akteure zu Einstellungsfragen empirisch bestätigt werden.

Ein Beispiel für eine weitverbreitete Nebenfolge auf der Einstellungsebene ist das Auftreten psychischer Kosten im Rahmen der Neuen Steuerung. Zum einen wurde das Belastungserleben der Lehrkräfte im Zuge zentraler Leistungstests erhoben („Sich in zentralen Leistungstests immer wieder einem Vergleich stellen zu müssen bedeutet für Lehrkräfte unseres Kollegiums eine permanente Belastung“) und zum anderen die Einschätzung abgefragt, inwiefern der „Druck in den letzten Jahren gestiegen ist, gute Testergebnisse erzielen zu müssen“. Die Häufigkeitsverteilung lässt deutlich erkennen, dass Vergleiche für die Lehrkräfte bundeslandübergreifend mit psychischen Kosten in Verbindung stehen. Das Erleben von Druck in Reaktion auf die Notwendigkeit, gute Testergebnisse erzielen zu müssen, erhält dabei die höchste Zustimmung. Den Spitzenwert nimmt hier Berlin mit 81,3 % ein, dicht gefolgt von Brandenburg mit 77,8 % und Rheinland-Pfalz mit 73,5 %, während in Thüringen mit 64,6 % deutlich weniger der Befragten dem Item zustimmen.

Die Verteilung der Häufigkeiten lässt zwar nur bedingt Rückschlüsse zu, inwiefern das neue Steuerungsmodell an sich zu psychischen Kosten führt, allerdings scheint sich die im Modell angelegte Verantwortungszuschreibung für schulische Leistungen im Sinne von accountability auf der Einstellungsebene niederzuschlagen. Es kann empirisch gezeigt werden, dass das im Modell über Vergleiche und Feedbacks angelegte Druckmoment Wirksamkeit entfaltet. Dabei scheinen die Ergebnisse passgenau zu den Erwartungen, die aus der Analyse der Steuerungsregime resultieren. In Thüringen, dessen Steuerungsregime weniger durch staatlich-administrativ oder marktlich-wettbewerblich vermittelte Druckmechanismen charakterisiert ist, zeigen schulische

Akteure auch ein geringeres Belastungserleben, während in Berlin die spezifische Ausgestaltung des Steuerungsregimes ein erhöhtes Belastungserleben nach sich zieht.

Ein weiteres Beispiel für eine Nebenfolge, die besondere Verbreitung erfahren hat, ist im Kontext von Deprofessionalisierungseffekten die Delegation bestimmter Aufgaben an professionsfremde Experten oder aber die zunehmende Übernahme professionsfremder Aufgaben. Mit Tacke (2005) kann man dies als Relativierung einer an den Erziehungsaufgaben im Kerngeschäft des Unterrichts festgemachten professionellen Autonomie beschreiben. Betrachtet man die Verteilung der Häufigkeiten im Antwortverhalten der Akteure hinsichtlich des Eindrucks, „dass die Festlegung von Lernzielen in den letzten Jahren zunehmend zu einer Aufgabe außerschulischer Instanzen geworden ist“, kann bundeslandübergreifend festgestellt werden, dass deutlich über 50 % der Befragten Verschiebungen im Aufgabenprofil von Lehrkräften wahrnehmen. Diese werden in der Literatur häufig als Deprofessionalisierung bezeichnet (vgl. ebd.). Besonders in Brandenburg und Berlin ist der Anteil der Lehrkräfte, die in den Kategorien „trifft voll zu“ und „trifft eher zu“ geantwortet haben, mit 79,2 % und 78 % sehr hoch. Rheinland-Pfalz (68,9 %) und Thüringen (61,8 %) liegen etwas darunter. Analog zur Nebenfolge der psychischen Kosten lässt auch das Antwortverhalten an dieser Stelle darauf schließen, dass es sich um eine Nebenfolge handelt, die bereits durch das Modell der Neuen Steuerung nahegelegt und von schulischen Akteuren sehr wohl registriert wird.

Eine weitaus weniger wahrgenommene Nebenfolge auf Einstellungsebene scheint Konkurrenzdenken zwischen Lehrkräften zu sein. Auf das Item „Die Rückmeldung von Ergebnissen klassenübergreifender Tests führt zu Konkurrenzdenken im Kollegium“ reagierten lediglich 16,5 % der Lehrkräfte in Rheinland-Pfalz, 13,4 % der brandenburgischen Lehrkräfte und 11,6 % der Berliner Lehrkräfte mit Zustimmung. Die geringste Zustimmung fand dieses Item in Thüringen. Hier stimmten 8,3 % der Lehrkräfte dem Item zu.

Hinsichtlich derjenigen Nebenfolgen, die einer weniger flächendeckenden Verbreitung unterliegen, sollen im Folgenden noch exemplarisch besonders interessante Ergebnisse präsentiert werden, die Hinweise darauf geben, dass die Einstellungsebene der Akteure im besonderen Maße von der differentiellen Ausgestaltung des neuen Steuerungsmodells in den Steuerungsregimen der einzelnen Bundesländer tangiert wird. Für die Nebenfolge Abhängigkeit vom Expertenurteil, die mittels dreier Items operationalisiert wurde („Die Rückmeldung aus VERA gibt mir Orientierung in der Ausübung meines Berufs“, „Ich benötige die Ergebnisse aus VERA, um zuverlässige Auskünfte über den Leistungsstand meiner Schüler und Schülerinnen zu bekommen“ und „Wir benötigen die Rückmeldung aus der externen Evaluation, um die Qualität der Arbeit unserer Schule zuverlässig einschätzen zu können“), lassen sich erhebliche Unterschiede zwischen den Bundesländern zeigen. Vor allem für Thüringen kann nachgewiesen werden, dass Rückmeldungen aus standardisierten Instrumenten der Leistungsmessung zur ausschlaggebenden Größe geworden sind, um die eigene Arbeit einschätzen zu können. Dies gilt etwa in Hinblick auf VERA für 53,7 % der Lehrkräfte. Darauf folgt Brandenburg mit 39,3 %, Rheinland-Pfalz mit 24,4 % und Berlin mit 17,5 %. Die Verteilung der Häufigkeiten für die anderen Items dieser Nebenfolge ergibt ein ähnliches Bild. Bringt man diese Ergebnisse in einen Zu-

sammenhang mit der These, dass die Implementierung der Neuen Steuerung einen entscheidenden Einfluss auf die Verbreitung von Nebenfolgen hat, so scheint es, dass in Thüringen die programmatische Etablierung des neuen Steuerungsmodells als Teil der professionellen Selbststeuerung grundsätzlich stärkere Akzeptanz generiert als in den anderen Bundesländern.⁶ Dies kann jedoch dazu führen, dass die Instrumente auch unhinterfragt als Gradmesser der Qualität eigener Arbeit übernommen werden.

6. Fazit

Die Ergebnisse der Nefo-Studie zeigen, dass auch unter „Low- und No-Stakes“-Bedingungen Nebenfolgen auftreten. Die Existenz von stakes kann damit nicht mehr als alleiniges und hinreichendes Erklärungsmodell für die Entstehung bzw. das Ausbleiben von Nebenfolgen dienen. Ebenso wenig sind die Nebenfolgen in jedem Fall zwangsläufige Konsequenz der Etablierung neuer Steuerungsmodelle. Auf Basis der qualitativen Interviewstudie und der quantitativen Fragebogenuntersuchung lassen sich folgende Erklärungsansätze zum Auftreten von Nebenfolgen unter „Low- und No-Stakes“-Bedingungen formulieren. Die Ansätze umfassen habituelle und kontextuelle Erklärungsdimensionen:

Erstens entfalten die modellinhärenten Deutungsangebote und Handlungsoptionen Neuer Steuerung Wirkung in unterschiedlichen Formen eines Anpassungsverhaltens. Das, was gemessen wird, gewinnt an Bedeutung gegenüber dem Nichtgemessenen und nicht notwendigerweise Messbaren. Vorgängige zentrale Orientierungspunkte schulischer Akteure werden verbreitet modellkonform transformiert, sodass schulisches Handeln verstärkt an Output anstelle von Outcome ausgerichtet wird.

Zweitens weisen die Ergebnisse der quantitativen Erhebung darauf hin, dass Lehrkräfte Formen eines Anpassungsverhaltens zeigen, das von der Bildungspolitik und -administration als gewünschtes Verhalten kommuniziert wird. Dies zeigt die weite Verbreitung von Coaching anhand alter Testaufgaben, zu denen bildungspolitisch implizit und explizit aufgefordert wird.

Ausweichverhalten der Lehrkräfte scheint – drittens – vor allem durch die Installation von Druckmomenten innerhalb eines jeweiligen Steuerungsregimes und seiner spezifischen Kombination aus Steuerungsmechanismen systematisch angereizt zu werden. Steuerungsregime, die auf staatlich-administrativen Druck durch Kontrolle oder auf den Druck marktähnlicher Strukturen setzen, scheinen eher Ausweichreaktionen der handelnden Akteure zu erzeugen als Regime, die stärker auf professionelle Selbststeuerung setzen.

Schließlich zeigt die Auswertung der Interviewstudie – viertens –, dass auch die Bedeutungskonstruktionsleistungen im schulischen Feld eine Erklärung für das

6 Die Ergebnisse von Maier (2008) weisen in eine ähnliche Richtung. Er stellt heraus, dass zentrale Vergleichstests in Thüringen weitaus akzeptierter sind als im Vergleichsland Baden-Württemberg. Maier führt die Unterschiede in der Akzeptanz der Tests aber auf Merkmale der Testimplentation und der Ergebnisrückmeldung zurück.

Auftreten von Nebenfolgen liefern. Je nach Reformrezeptionsmuster erlangen die neuen Steuerungsstrukturen eine andere Bedeutung. Es strukturiert auch die Reaktion und den Umgang der Akteure auf und mit den Instrumenten Neuer Steuerung einschließlich der Nebenfolgenproduktion vor. Zusammengenommen eröffnet die Berücksichtigung habitueller und kontextueller Faktoren einen differenzierten Blick auf das Auftreten und die Verbreitung von Nebenfolgen Neuer Steuerung unter Bedingungen von „low-stakes“ und „no-stakes“.

Die vier Erklärungsansätze sind Ergebnis eines ersten explorativen Zugriffs und müssen in weiteren Untersuchungen geprüft und ggf. modifiziert werden. Hierfür könnte die Gültigkeit der Annahmen für andere Bundeslandkontexte überprüft werden. Denkbar wäre beispielsweise, Zwillinge für die im Rahmen der Nefo-Studie untersuchten Bundesländer zu identifizieren, also Bundesländer mit ähnlich ausgeprägten Steuerungsmechanismen. An diesen könnte untersucht werden, ob sich auch eine ähnliche Verteilung und Verbreitung von Nebenfolgen zeigen lässt. Die quantitativen Befunde müssten zusätzlich weiter validiert werden. So könnten ausgewählte Nebenfolgen als latente Konstrukte mittels valider Skalen abgebildet werden. Zu denken wäre hierbei an Phänomene auf der Einstellungsebene wie Vertrauen oder Belastungserleben, die durch mehrere Items operationalisiert werden könnten. So würde auch der komplexen theoretischen Diskussion um diese Phänomene Rechnung getragen werden. Auch eine Wiederholungsstudie in den untersuchten Bundesländern scheint vielversprechend, um das Auftreten von Nebenfolgen tiefer gehend zu untersuchen. Die Etablierung Neuer Steuerung ist bislang keineswegs abgeschlossen. Neue bildungspolitische Regelungen führen u. U. zu einer Rekonfiguration des Zusammenspiels der Steuerungsmechanismen. Ein Wiederholungsdesign würde Rückschlüsse auf den Effekt solcher Entwicklungen ermöglichen.

Aufgrund der methodischen Anlage der Nefo-Studie als Bundesländervergleich wurden Unterschiede in der Verbreitung von Nebenfolgen auf Differenzen in der Ausgestaltung Neuer Steuerung zurückgeführt. Studien mit einem Fokus auf alternative Erklärungsfaktoren, wie z. B. die diskursive Rahmung Neuer Steuerung, die jeweils bundeslandspezifischen Ausgestaltungen einzelner Steuerungsinstrumente, einzelschulische Erklärungsfaktoren⁷, unterschiedliche Arbeitsbedingungen in den Bundesländern etc., könnten weitere Einflussvariablen identifizieren.

Schließlich wurden das Auftreten und die Verbreitung von Nebenfolgen im Nefo-Projekt auf Grundlage von Selbstauskünften von Lehrkräften und Schulleitern abgeschätzt. Der Rekurs auf Selbstauskünfte enthält das Problem eines möglichen self-serving bias, wodurch die der Neuen Steuerung zugeschriebenen Nebenfolgen überzeichnet dargestellt würden. Umgekehrt kann ein social desirability bias

7 Im Rahmen einer ersten Exploration der quantitativen Daten wurden für jedes Item bundeslandspezifische Intraclass Correlation Coefficients (ICC) ermittelt. Die Mehrzahl der ICCs weist Werte kleiner als 0,10 auf. 24 % der ICCs liegen zwischen 0,10 % und 0,20 % auf. Lediglich 11 % der ICCs sind größer als 0,20. Die für die Einzelitems berechneten ICCs variieren über die vier Bundesländer hinweg z. T. erheblich. Diese Ergebnisse lassen vermuten, dass Unterschiede hinsichtlich der Wahrnehmung von Nebenfolgen nur bedingt durch einzelschulspezifische Faktoren erklärt werden können. Allerdings handelt es sich hierbei lediglich um eine Beobachtung im Zuge einer ersten Exploration der Daten. Die Berechnung von ICCs setzt metrische Daten voraus, was im vorliegenden Fall nicht gegeben ist.

auch dazu führen, dass etwa Formen eines anreizwidrigen Ausweichverhaltens in Selbstauskünften seltener berichtet werden, als sie tatsächlich vorkommen. Fakt ist, dass eine auf Selbstauskünften beruhende Studie ergänzt werden müsste durch die Beobachtung schulischer Praxis. So können ethnografische und praxeografische Untersuchungen näher beleuchten, inwieweit die Selbstauskünfte der Lehrerinnen und Lehrer ihr Handeln im Schullalltag widerspiegeln. Hier schließt das Folgeprojekt der Nefo-Studie zu Förder- und Selektionspraktiken als Reaktionsformen auf Instrumente der Output- und Wettbewerbssteuerung im Schulsystem (Self) an.

5. Literaturverzeichnis

- Bellmann, J. (2006). Bildungsforschung und Bildungspolitik im Zeitalter ‚Neuer Steuerung‘. *Zeitschrift für Pädagogik*, 52 (4), S. 487–504.
- Bellmann, J., Dužević, D., Schweizer, S. & Thiel, C. (2016) (paper accepted). Nebenfolgen Neuer Steuerung und die Rekonstruktion ihrer Genese. Differenten Orientierungsmuster schulischer Akteure im Umgang mit neuen Steuerungsinstrumenten. Erscheint In *Zeitschrift für Pädagogik*, 62 (3).
- Bellmann, J. & Weiß, M. (2009). Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. Theoretische Konzeptualisierung und Erklärungsmodelle. *Zeitschrift für Pädagogik*, 55 (2), S. 286–308.
- Bellmann, J. (2014): Warum man eine datengetriebene Steuerung nicht mit einer evidenzbasierten Steuerung verwechseln sollte. Vortrag auf dem ZfE-Forum „Kritik empirischer Bildungsforschung“ in Hamburg am 5. und 6. Dezember 2014. Erscheint In J. Baumert & K.-J. Tillmann, (Hrsg.) (2016): *Der kritische Blick auf die empirische Bildungsforschung*. Sonderheft der ZfE.
- Benner, D. (2007). Unterricht – Wissen – Kompetenz. Zur Differenz zwischen didaktischen Aufgaben und Testaufgaben. In D. Benner (Hrsg.): *Bildungsstandards. Instrumente zur Qualitätssicherung im Bildungswesen. Chancen und Grenzen – Beispiele und Perspektiven*, S. 123–138. Paderborn u.a.: Ferdinand Schöningh.
- Bevan, G. & Hood, C. (2006). What’s measured is what matters: targets and gaming in the English public health care system. *Public Administration*, 84 (3), S. 517–538.
- Blömeke, S. (2004). Empirische Befunde zur Wirksamkeit der Lehrerbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.): *Handbuch Lehrerbildung*, S. 59–91. Bad Heilbrunn: Klinkhardt.
- Blömeke, S. (2009). Allgemeine Didaktik ohne empirische Lernforschung. Perspektiven einer reflexiven Bildungsforschung. In K.-H. Arnold, S. Blömeke, R. Messner & J. Schlömerkemper (Hrsg.): *Allgemeine Didaktik und Lehr-Lernforschung. Kontroversen und Entwicklungsperspektiven einer Wissenschaft vom Unterricht*, S. 13–27. Bad Heilbrunn: Klinkhardt.
- Böttcher, W. (2012). Teaching to the Test. Warnung vor dem falschen Vorbild. *Friedrich Jahresheft*, 30, S. 88–89.

- Campbell, Donald T. (1975): Assessing the Impact of Planned Social Change. In Gene M. Lyons (Hrsg.): *Social Research and Public Policies*, S. 3–45. The Dartmouth/OECD Conference. Hanover, New Hampshire: Dartmouth College
- Chapman, D. W. (2008). Options for improving the management of education systems. In W. K. Cummings & J. H. Williams (Hrsg.): *Policy-Making for Education Reform in Developing Countries: Policy, Options and Strategy*, S. 251–276. Lanham: Rowman & Littlefield Education.
- de Wolf, I. F. & Janssens, F. J. G. (2007): Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education*, 33 (3), S. 379–396.
- Ehren, M. C. M. & Visscher, A. J. (2006): Towards a Theory On the Impact of School Inspections. *British Journal of Educational Studies*, 54 (1), S. 51–72.
- Etzioni, A. (1964). *Modern Organizations*. Englewood-Cliffs: Prentice-Hall.
- Fuchs, H.-W. (2009). Neue Steuerung – neue Schulkultur? *Zeitschrift für Pädagogik*, 55 (3), S. 369–380.
- Frey, B. S. & Jegen, R. (2002): Intrinsische Motivation. Kontraproduktive Wirkung des Motivators „Geld“. In *Verbands-Management*, 28 (3), S. 30–41.
- Holzer, B. (2006). Denn sie wissen nicht, was sie tun? Nebenfolgen als Anlass soziologischer Aufklärung und als Problem gesellschaftlicher Selbstbeschreibung. In S. Bösch, N. Kratzer & S. May (Hrsg.): *Nebenfolgen. Analysen zur Konstruktion und Transformation moderner Gesellschaften*, S. 39–64. Weilerswist: Velbrück Wissenschaft.
- Hopmann, S. T. (2003): On the evaluation of curriculum reforms. *Journal of Curriculum Studies*, 35 (4), S. 459–478.
- Institut für Schulqualität der Länder Berlin und Brandenburg e.V (ISQ). *Vergleichsarbeiten Jahrgangsstufe 8 (VERA)*. Online verfügbar unter <https://www.isq-bb.de/Jahrgangsstufe-8.30.0.html>, zuletzt geprüft am 22.11.2014.
- Institut zur Qualitätsentwicklung im Bildungswesen. *VERA-Aufgabenpool*. Verfügbar unter <http://www.iqb.hu-berlin.de/vera/aufgaben>, zuletzt geprüft am 22.11.2014.
- Jäger, D. (2012). Herausforderung Zentralabitur: Unterrichtsinhalte variieren und an Prüfungsthemen anpassen. In K. Maag Merki (Hrsg.): *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*, S. 179–205. Wiesbaden: VS.
- Jones, G., Jones, B. & Hargrove, T. (2003). *The Unintended Consequences of High-Stakes Testing*. Lanham u. a.: Rowman & Littlefield.
- Koretz, D. M. (2005). Alignment, High Stakes, and the Inflation of Test Scores. In J. L. Herman & E. H. Haertel (Hrsg.), *Uses and Misuses of Data for Educational Accountability and Improvement*, S. 99–118. Malden: Blackwell.
- Kurtz, Thomas (2009): Professionalität aus soziologischer Perspektive. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.): *Lehrprofessionalität*, S. 45–54. Weinheim und Basel: Beltz.
- Levacic, R. & Woods, Ph. A. (2000). Quasi-markets and school performance. Evidence from a study of English secondary schools. In M. Weiß & H. Weishaupt (Hrsg.): *Bildungsökonomie und Neue Steuerung*, S. 53–95. Frankfurt, Main: Lang.

- Maier, U. (2008): Vergleichsarbeiten im Vergleich – Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. *Zeitschrift für Erziehungswissenschaft*, 11 (3), S. 453–474.
- Maier, U. (2010). Effekte testbasierter Rechenschaftslegung auf Schule und Unterricht. Ist die internationale Befundlage auf Vergleichsarbeiten im deutschsprachigen Raum übertragbar? *Zeitschrift für Pädagogik*, 56 (1), S. 112–128.
- Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (MBS) (2014). *Rundschreiben Nr.10/2014 vom 15.09.2014*. Online verfügbar unter https://www.isq-bb.de/uploads/media/Rundschreiben_2015_VERA-3_8_2014-09-15.pdf, zuletzt geprüft am 22.11.2014.
- Nichols, S. L. & Berliner, D. C. (2005), *The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing*. Tempe: Arizona State University.
- Nichols, S., Glass, G. & Berliner D. C. (2005). *High-Stakes Testing and Student Achievement: Problems for the No Child Left Behind Act*. Arizona State University. Verfügbar unter <http://files.eric.ed.gov/fulltext/ED531184.pdf>, zuletzt geprüft am 21.02.2014.
- Nichols, S. & Berliner, D. C.. (2007). *Collateral Damage. How High-Stakes Testing Corrupts America's Schools*. Cambridge: Harvard Education Press.
- Nohl, A.-M. (2009). *Interview und dokumentarische Methode. Anleitungen für die Forschungspraxis*. 3. Auflage. Wiesbaden: VS.
- Qesel, C., Husfeldt, V., Landwehr, N. & Steiner, P. (Hrsg.) (2011). *Wirkungen und Wirksamkeit der externen Schulevaluation*. Bern: hep Verl.
- Ridgway, V. F. (1956). Dysfunctional Consequences of Performance Measurements. *Administrative Science Quarterly*, 1 (2), S. 240–247.
- Rürup, M. (2007). *Innovationswege im deutschen Bildungssystem. Die Verbreitung der Idee „Schulautonomie“ im Ländervergleich*. Wiesbaden: VS.
- Schreyer, P. (2012). Output, Outcome and Quality Adjustment in Measuring Health and Education Services. *Review of income and wealth*, 58 (2), S. 257–278.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft (2013). *Verwaltungsvorschrift Schule Nr. 11/2013*. Online verfügbar unter https://www.isq-bb.de/uploads/media/Verwaltungsvorschrift_11_2013_01.pdf, zuletzt geprüft am 22.11.2014.
- Shavelson, R. J. & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change*, 35 (1), S. 10–19.
- Tacke, V. (2005). Schulreform als aktive Deprofessionalisierung. Zur Semantik der Lernenden Organisation im Kontext der Erziehung. In T. Klatetzki & V. Tacke (Hrsg.): *Organisation und Profession*, S. 165–199. Wiesbaden: VS.
- Terhart, E. (2012). *Wirkungsannahmen in Konzepten der Qualitätsverbesserung des Bildungssystems: Hoffen, Bangen, Trauern*. Vortrag am Institut für Bildungsmonitoring und Qualitätsentwicklung, Hamburg am 23.11.2012. Unveröffentlichtes Vortragsmanuskript.
- Thiel, F., Cortina, K. S. & Pant, H. A. (2014). Steuerung im Bildungssystem im internationalen Vergleich. In R. Fatke & J. Oelkers (Hrsg.): *Das Selbstverständnis der Erziehungswissenschaft: Geschichte und Gegenwart*. Zeitschrift für Pädagogik, Beiheft 60, S. 123–138. Weinheim; u.a.: Beltz Juventa.

- Tyack, D. & Cuban, L. (1995). *Tinkering toward utopia: a century of public school reform*. Cambridge u. a.: Harvard University Press.
- van Ackeren, I., Binnewies, C., Clausen, M., Demski, D., Dormann, C., Koch, A. R., Laier, B., Preisendörfer, P., Preuße, D., Rosenbusch, C., Schmidt, U., Stump, M. & Zlatkin-Troitschanskaia, O. (Autorengruppe EviS) (2013). Welche Wissensbestände nutzen Schulen im Kontext von Schulentwicklung? Theoretische Konzepte und erste deskriptive Befunde des EviS-Verbundprojektes im Überblick. *Die Deutsche Schule* 13, Beiheft 12, S. 51–73. Münster: Waxmann.
- van Thiel, S. & Leeuw, F. (2002). The Performance Paradox in the Public Sector. *Public Performance & Management Reviews*, 35 (3), S. 267–281.
- Wacker, A. & Kramer, J. (2012): Vergleichsarbeiten in Baden-Württemberg. *Zeitschrift für Erziehungswissenschaft*, 15 (4), S. 683–706.
- Wacker, A., Maier, U. & Wissinger, J. (Hrsg.) (2012). *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen*. Wiesbaden: VS.

Stefan Brauckmann

Schulleitungshandeln zwischen erweiterten Rechten und Pflichten (SHaRP) unter Berücksichtigung einschlägiger schulrechtlicher Bestimmungen¹

1. Rahmenbedingungen von Schulleitungshandeln zwischen erweiterten Rechten und Pflichten

1.1 Hintergrund

Die aktuellen schulpolitischen Veränderungen in den Bundesländern zielen auf Qualitätssteigerungen durch Stärkung der Selbstständigkeit von Einzelschulen sowie auf jene selbst initiierten und selbst gesteuerten Entwicklungsprozesse ab, die mit externer Standardsetzung und verstärkter Ergebniskontrolle verknüpft werden. In diesem Zusammenhang und aus Gesamtsteuerungsperspektive gewinnt die Rolle von Schulleitung zunehmend an Bedeutung. Als ein zentraler Akteur bei der Umsetzung dieser Konzeptionen rückt die Schulleitung zunehmend ins Blickfeld der Bildungsforschung. An Schulen mit mehr Eigenverantwortung stehen Schulleitungen vor neuen Aufgaben, wie etwa strategischen Führungsaufgaben, da Schulleitungen nicht nur das operative Management im Sinne des administrativen und pädagogischen Alltagsgeschäfts zufällt, sondern sie auch die Ziele jeweiliger Schulentwicklung festzulegen und deren konzeptionelle Umsetzung mit Blick auf die extern herangetragenen Standards zu sichern haben. Somit steht die Einzelschule vor der Aufgabe, ihre Entwicklung eigenständig, zielgerichtet und systematisch zu betreiben. Hierfür benötigt sie eine entsprechende Selbststeuerungsfähigkeit. Den Schulleitungen obliegt es, diese aufzubauen und zu fördern.

Die Mehrbelastung durch neue Aufgaben und erweiterte Anforderungen wird nicht nur von Interessenverbänden der Schulleiterinnen und Schulleiter, sondern auch in der Bildungsforschung als Problem angesehen. So wird insbesondere von der Bildungsforschung infrage gestellt, ob die bisherigen Schulleiterqualifizierungs-

¹ Der hier vorliegende Beitrag beruht im Wesentlichen auf drei (online verfügbaren) Expertisen, die der Bildungsjurist und Ministerialrat a. D. Klaus Hanßen (Hanßen 2011, 2012, 2013c) für das im Rahmen des BMBF Forschungsschwerpunkts „Steuerung im Bildungssystem“ (SteBis) geförderten Forschungsprojekts „Schulleitungshandeln zwischen erweiterten Rechten und Pflichten (SHaRP)“ erstellt hat. Der Autor dankt an dieser Stelle Herrn Hanßen ausdrücklich für dessen kritisches, umsichtiges wie detailreiches Feedback und seine kontinuierliche Diskussionsbereitschaft im Rahmen der Publikationserstellung, die einer Co-Autorenschaft gleichkommt.

maßnahmen noch den erweiterten Aufgaben der Schulleiterinnen und Schulleiter entsprechen.

In diesem Zusammenhang wollte das vom BMBF geförderte Forschungsprojekt „Schulleitungshandeln zwischen erweiterten Rechten und Pflichten“ (SHaRP) zu einer weiteren, empirisch fundierten Klärung des Belastungserlebens von Schulleiterinnen und Schulleitern im Kontext neuer Steuerungsansätze beitragen. Zu diesem Zweck wurden in Grundschulen und Gymnasien von sechs Bundesländern, die laut einem Schulautonomieindex (vgl. Rürup 2007) einen unterschiedlichen Grad an Eigenverantwortlichkeit der Einzelschule ausweisen, quantitative und qualitative Daten (a) zum Schulleitungshandeln selbst, (b) zu Belastungsfaktoren im Schulleitungshandeln und schließlich (c) zu den Wirkungen des Schulleitungshandelns auf die Selbststeuerungsfähigkeit (z. B. Qualitätssicherung und Feedback) der Schule untersucht.

Die Befunde erlaubten es, Belastungsfaktoren und Tätigkeiten von Schulleiterinnen und Schulleitern sowie regionale, institutionelle und auch persönlichkeitsbezogene Unterschiede zu identifizieren.

In dem hier vorliegenden Beitrag soll auf rechtliche Rahmenbedingungen als Einflussgröße von Schulleitungshandeln zwischen erweiterten Rechten und Pflichten aus einer stärker bildungsrechtlichen Perspektive eingegangen werden, zumal Schulleitungen auf die Übernahme solcher neuer Rechte und Pflichten erst seit Kurzem systematisch vorbereitet werden. Hinzu kommt, dass viele Lehrkräfte in den Schulen den neuen Ansätzen der Qualitätsentwicklung und Qualitätssicherung skeptisch gegenüberstehen. Ferner gibt es in den meisten Schulen noch nicht das Personal, das die Schulleitungen bei der Erledigung der erweiterten inhaltlichen und organisatorischen Aufgaben unterstützen könnte (Hanßen 2005). Die Einnahme einer solchen Perspektive scheint umso mehr geboten, als bildungsrechtliche Analysen gegenüber den politischen Konzepten und Anforderungskatalogen den Vorteil des Rechts als objektiven Bezugsrahmen aufweisen.

2. Systemmerkmale als Einflussgröße von Schulleitungshandeln

Blickt man auf die Befundlage der empirischen Schulleitungsforschung, so wird schnell deutlich, dass sich nur wenige Untersuchungen mit der Frage beschäftigen, von welchen systemischen, organisationalen und individuellen Bedingungen Schulleitungshandeln (Bolman & Deal 1992) und Belastungsempfinden von Schulleitern abhängen. Bislang wird in nur wenigen Modellen der Versuch einer integrierenden Gesamtdarstellung von potenziellen Einflussgrößen des Leitungshandelns in Schulen unternommen. Den Systemkontext von der Organisation und die Organisation von den Handelnden empirisch zu trennen bzw. die damit angesprochenen Wechselbeziehungen empirisch adäquat zu modellieren bleibt vorerst ein zentrales Forschungsdesiderat.

Unter systemischen Rahmenbedingungen werden häufig *allgemeingesellschaftliche Herausforderungen* (demografische Entwicklung, Strukturwandel zur Dienstleistungs- und Wissensgesellschaft) und *ökonomische Rahmenbedingungen* (Finanz-

situation der öffentlichen Haushalte, Entgrenzung der nationalen Arbeitsmärkte, zunehmende Internationalisierung von Arbeitskooperationen) gefasst (vgl. Autorengruppe Bildungsberichterstattung 2008, S. 15). Theoretisch-konzeptionell wird dieser Annahme durch die Integration von Systemkontextmerkmalen in Rahmen- (vgl. OECD 2008) wie auch Wirkungsmodellen (vgl. Day et al. 2009) deskriptiv-analytisch in der indikatorengestützten nationalen Bildungsberichterstattung Rechnung getragen (vgl. Autorengruppe Bildungsberichterstattung 2012). Von großer Bedeutung, insbesondere im Zusammenhang mit (internationalen) systemvergleichenden Untersuchungen, sind darüber hinaus die *schulrechtlichen Rahmenbedingungen*, wie etwa der Grad an Autonomie der Einzelschulen (vgl. Wößmann et al. 2007). Da in den vorliegenden Untersuchungen auch auf mögliche schulrechtliche Einflussfaktoren von Schulleitungshandeln eingegangen wird, soll dieser Aspekt im Folgenden ausführlicher dargestellt werden.

Der Zusammenhang von schulrechtlichen Bedingungen und Führungshandeln wird nur selten zum zentralen Untersuchungsgegenstand erhoben. Beispielsweise wird in diesem Zusammenhang erfragt, inwieweit die de jure gewährten erweiterten Handlungsspielräume durch die Schulleitung im Sinne einer praktizierten Gestaltungsautonomie tatsächlich genutzt werden (vgl. Brauckmann 2012). Generell überwiegen allerdings bislang deskriptiv-analytische Untersuchungen, die sich mit den veränderten Anforderungsprofilen von Schulleitungen im Rahmen der bildungspolitischen Stärkung schulischer Selbständigkeiten auseinandersetzen (für eine Zusammenstellung siehe Brauckmann 2012). So können Rechte und Pflichten als subjektive Rechte und Pflichten beschrieben werden, indem die Frage gestellt wird: Was hat eine Schulleiterin oder ein Schulleiter in einer tatbestandlich bestimmten Situation zu tun oder zu lassen? Hierzu zählen auch die Beteiligungs-, Mitwirkungs- und Mitbestimmungsaufgaben in schulinternen und schulexternen Prozessen. Die Tätigkeit von Schulleiterinnen oder Schulleitern wird in ein übergreifendes System staatlicher Planung eingebettet, „von der sie sich in den materiell-äußerlichen Vorbedingungen ebenso wie hinsichtlich des zu vermittelnden Unterrichtsinhaltes in fühlbarer Abhängigkeit befindet“ (Oppermann 1969). Die herrschende juristische Meinung leitet aus dem Grundgesetz die notwendige demokratische Legitimation des Handelns von Schulleiterinnen oder Schulleitern ab und hält deshalb eine autonome Schule nicht für möglich (Niehues/Rux 2014). Allerdings wird zunehmend anerkannt, dass Schulleitungen durch Gesetz Aufgaben übertragen werden dürfen, welche diese mit einem hohen Grad an Selbstständigkeit wahrnehmen dürfen (vgl. Avenarius & Füssel 2010). Dabei handelt es sich vorrangig um pädagogische Aufgaben, aber auch um dienstrechtliche und haushaltsrechtliche Befugnisse. Daneben sind Schulleitungen an gesetzliche und tatsächliche Bedingungen gebunden, welche sie selber nicht beeinflussen können. Zu diesen Rahmenbedingungen zählen der Rechtsstatus der Schule mit den Aufgabenabgrenzungen zur Schulaufsicht und zum Schulträger, die Qualifizierung und Unterstützung der Schulleitungen (Fort- und Weiterbildung, Supervision und Coaching) sowie die Ausstattung der Schule (zeitliche, sachliche und personelle Ressourcen).

Je nachdem, was die Schulgesetze der Länder in Sachen Stärkung der Selbstständigkeit vorsehen (vgl. Hanßen 2013b; vbw 2010), unterscheiden sich jedoch die

Kompetenzen und die ihnen zugrunde liegenden Führungsinstrumente der Schulleitungen.

3. Methodisches Vorgehen

Um die Rahmenbedingungen, denen Schulleitungshandeln unterliegt, wissenschaftlich seriös erfassen zu können, sollten die Anforderungsprofile an Schulleitungshandeln über die theorie- und professionsbasierten Perspektiven unter Berücksichtigung einschlägiger schulrechtlicher Bestimmungen beschrieben werden. In den Ländern der Bundesrepublik Deutschland wurde das Ziel einer Stärkung der Selbstständigkeit oder der Eigenverantwortlichkeit in unterschiedlicher Weise umgesetzt. Um unterschiedliche Konzepte vergleichen zu können, wurden exemplarisch Länder ausgewählt, die diesem Ansatz entsprachen. Einer flächendeckenden Erfassung der schulrechtlichen Bestimmungen in den Ländern bedurfte es dafür nicht. Ausgewählt wurden die Länder Berlin, Brandenburg, Hamburg, Hessen und Nordrhein-Westfalen. Die Untersuchung wurde in den Ländern exemplarisch und aus Gründen der Übersichtlichkeit bewusst auf die Schulformen Grundschule und Gymnasium begrenzt. Aufgezeigt werden sollten die Gestaltungsmöglichkeiten für ein schulinternes wie -externes Leitungshandeln bei der Initiierung, Organisation und Moderation qualitätsorientierter Selbststeuerung (z. B. Förderung der Problemlösungsfähigkeit des Kollegiums, Aufbau einer kollegialen Schulkultur, Sensibilisierung des Personals hinsichtlich innovativen und effektiven Unterrichts). Es ging nicht nur um die rechtliche Darstellung von Kompetenzen und Weisungsbefugnissen von Schulleitung, sondern darauf aufbauend auch um die Beschreibung unterschiedlicher Beteiligungs-, Mitwirkungs- sowie Mitbestimmungsformen von Führungskräften in schulinternen wie schulexternen Prozessen der Entscheidungsfindung. Diese Perspektive sollte neue Informationen über inhaltliche und bisher vernachlässigte prozessuale Zusammenhänge eines auf qualitätsorientierte Selbststeuerung bezogenen Schulleitungshandelns bringen. In diesem Zusammenhang wurden auf der Grundlage einer qualitativen Inhaltsanalyse Tätigkeitskategorien erstellt.

Unter Berücksichtigung von „klassischen“ und neuen Aufgabenbereichen (vgl. Avenarius & Füssel 2010) wurden untersucht:

Aufgabenbereich 1 (Unterricht und pädagogische Innovation)

1. Eigener Unterricht,
2. Hospitation, Beratung und Unterstützung der Lehrkräfte,
3. Kontrolle, insbesondere der Einhaltung von Rechts- und Verwaltungsvorschriften,
4. Weiterentwicklung der Unterrichts- und Erziehungsarbeit,
5. Schulprogrammarbeit,
6. Interne Evaluation.

Aufgabenbereich 2 (Beteiligung von Schülerinnen und Schülern sowie ihren Eltern)

1. Information, Beratung und Beteiligung der Schülerinnen und Schüler,
2. Information, Beratung und Beteiligung der Eltern,
3. Schulkonferenz/Schulforum.

Aufgabenbereich 3 (Personalführung und Organisationsentwicklung)

1. Dienstrechtliche Befugnisse generell
2. Personalauswahl,
3. Personalgespräche,
4. Dienstliche Beurteilungen,
5. Beförderungen,
6. Disziplinarmaßnahmen,
7. Planung und Durchführung von Fortbildung,
8. Zusammenarbeit in der Schulleitung,
9. Zusammenarbeit mit den Konferenzen der Lehrkräfte.

Aufgabenbereich 4 (Verwaltungs- und Organisationsaufgaben)

1. Stellen und Personalmittel,
2. Sachmittel,
3. Zuwendungen Dritter,
4. Rechenschaftslegung, Statistiken,
5. Stundenplan und Einsatz der Lehrkräfte.

Aufgabenbereich 5 (Vertretung der Schule nach außen)

1. Außenvertretung generell,
2. Zusammenarbeit mit anderen Schulen,
3. Zusammenarbeit mit dem Schulträger/Schulaufwandsträger,
4. Zusammenarbeit mit den Schulaufsichtsbehörden,
5. Öffnung der Schule, Zusammenarbeit mit anderen Einrichtungen und Organisationen,
6. Teilnahme an Sitzungen von Gremien auf kommunaler Ebene,
7. Öffentlichkeitsarbeit.

Schulleitungshandeln wird aber auch durch einen objektiven Rechtsrahmen geprägt. Deshalb wird neben der Darstellung von Rechten und Pflichten der Schulleiter in schulinternen und schulexternen Prozessen (Aufgabenbereiche) auch auf allgemeine Rahmenbedingungen für Schulleitungshandeln eingegangen. Als objektiver Rechtsrahmen werden hierbei Gesetze und untergesetzliche Festlegungen bezeichnet, die von der Schulleitung einzuhalten sind, ohne dass sie darauf rechtlich Einfluss nehmen kann. Diese Bedingungen schafft der Staat entweder aufgrund seiner Fürsorgepflichten für Schulleiterinnen und Schulleiter oder mit dem Ziel, die Qualität von Erziehung und Unterricht, das Schulklima und die Leistungen der Schülerinnen und Schüler zu sichern oder zu entwickeln.

Wegen der in Deutschland historisch gewachsenen und unterschiedlich ausgeprägten staatlichen Verantwortung für das Schulwesen sind die Ausgangsbedingungen für eine Stärkung der Selbstständigkeit der Schulen und damit auch der Schulleitungen unterschiedlich.

Über den verfassungsrechtlich zu vertretenden Umfang einer Stärkung der Rechte der Einzelschule und damit auch der Schulleitung gibt es unterschiedliche Meinungen. Im Schulgesetzentwurf des Deutschen Juristentages (1981) war vorgeschlagen worden, die Schule lediglich der staatlichen Rechtsaufsicht zu unterstellen. Dem ist kein Land gefolgt. Auch die Forderung einer „pädagogisch tatsächlichen rechtsfähigen Schule“ (Stock 2002) findet mehrheitlich keine Zustimmung.

Ingo Richter (1999) liest Artikel 7 Abs. 1 GG heute wie folgt: „Das gesamte Bildungswesen steht in öffentlicher Verantwortung, die durch eine staatliche Steuerung wahrgenommen wird.“ Er untersucht die Möglichkeiten einer Steuerung durch Recht, Organisation und Verfahren, Geld und Personal. Die Steuerungskraft von Gesetzen hält er für begrenzt. Durch eine Autonomisierung der Bildungsinstitutionen können die Probleme des Bildungswesens seiner Meinung nach nur gelöst werden, wenn es gleichzeitig eine Form zentraler Steuerung gibt. Dieses verlangt eine klare Aufgabenabgrenzung, Verantwortung und Rechenschaftslegung (Blossfeld et al. 2010). Dabei muss der Gesetzgeber die institutionellen Vorkehrungen treffen (Avenarius & Füssel 2010; Müller 2012).

Obschon Gesetze und Verwaltungsvorschriften wichtige Bausteine zur Festlegung der Rechte und Pflichten von Schulleitungen sind, ist Schulleitungshandeln immer auch durch ethische Grundüberzeugungen und politische und verwaltungsmäßige Setzungen geprägt. Solche Überzeugungen und Setzungen kann diese Untersuchung nur ansatzweise einbeziehen. Offen bleiben muss auch die Frage, inwieweit die rechtlichen Bestimmungen in praktisches Handeln umgesetzt werden konnten und damit tatsächlich wirksam wurden. Dies gilt umso mehr, als die „Erprobung“ solcher neuer schulrechtlicher Vorgaben und daraus resultierende Erkenntnisse erst am Anfang eines längerfristigen Prozesses stehen und auch unterschiedliche Implementierungsstadien hinsichtlich der einzelnen Länder zu verzeichnen sind.

4. Ergebnisse der dokumentengestützten Ausgangsanalyse

Einleitend sei darauf hingewiesen, dass die Ergebnisse der umfangreichen Expertise (für eine weiter gehende Betrachtung siehe Hanßen 2011, 2012, 2013a) im vorliegenden Beitrag nur auszugsweise und gerafft wiedergegeben werden können. Als Systematisierungs- wie Differenzierungshilfe werden die bereits weiter oben eingeführten Aufgabenbereiche von Schulleiterinnen und Schulleitern herangezogen.

4.1 Rechte und Pflichten von Schulleiterinnen und Schulleitern

Führungsverantwortung und Beteiligung an schulischen Entwicklungen (Aufgabenbereich 1)

Was die Begrifflichkeit anbelangt, so werden „Schulleitung“ und „Schulleiterin oder Schulleiter“ in den untersuchten Aufgabenbereichen nicht immer eindeutig voneinander abgegrenzt. Weder ist der Begriff „Führungsfunktion“ standardisiert, noch sind die Schulen in vergleichbarer Weise mit Funktionsstellen ausgestattet. Es gibt auch keine einheitlichen Amtsbezeichnungen. Das brandenburgische, das hessische und das nordrhein-westfälische Schulgesetz unterscheiden die Aufgaben der Schulleitung von denen der Schulleiterin oder des Schulleiters.

Berlin, Hamburg und Niedersachsen haben keine gesonderten Regelungen über eine Zusammenarbeit in der Schulleitung. In Hessen wird ausdrücklich festgelegt, dass die Mitglieder der Schulleitung ihre Arbeit insbesondere in regelmäßigen

Dienstbesprechungen koordinieren. In Nordrhein-Westfalen heißt es, dass die Schulleiterin oder der Schulleiter in der Schulleitung mit dem ständigen Vertreter oder der ständigen Vertreterin und gegebenenfalls mit weiteren Personen zusammenarbeitet. Betont wird, dass ihre oder seine Gesamtverantwortung und die abschließende Entscheidungsbefugnis unberührt bleiben. Brandenburg unterscheidet Aufgaben der Schulleitung von Aufgaben, die der Schulleiterin oder dem Schulleiter vorbehalten sind. Soweit dieses nicht der Fall ist, kann die Schulleiterin oder der Schulleiter im Leitungskollegium überstimmt werden.

Erweiterte respektive kollegiale Schulleitungen sehen die Schulgesetze in Berlin, Brandenburg, Niedersachsen und Nordrhein-Westfalen vor. Hamburg hat diese Möglichkeit aus dem Schulgesetz gestrichen. In Brandenburg, Niedersachsen und Nordrhein-Westfalen wird ausdrücklich betont, dass die Führungsverantwortung der Schulleitungen von den übergeordneten staatlichen Stellen zu achten ist. Hamburg konzentriert die Leitungsverantwortung ausschließlich bei der Schulleiterin oder dem Schulleiter. In den übrigen Ländern wird neben der auch dort festgelegten Gesamtverantwortung der Schulleiterin oder des Schulleiters die Kooperation in der Schulleitung betont. Das entspricht Überlegungen, wonach die Funktion der Schulleitung sinnvoller Weise nur gemeinsam verantwortet werden kann.

In den niedersächsischen Schulen ist im Zusammenhang mit der Einführung der Eigenverantwortung der Schulvorstand an die Stelle der Schulkonferenz als zentrales Entscheidungsgremium der Schule getreten. Die Befugnisse des Schulvorstands entsprechen überwiegend denen der Schulkonferenzen in den übrigen Ländern. Der Schulvorstand verfügt jedoch über umfassendere Rechte. Er entscheidet insbesondere über die Inanspruchnahme des erweiterten schulischen Handlungsspielraums der Schule und über die Ausgestaltung der Stundentafel.

Die Führungsverantwortung der Schulleiterin oder des Schulleiters ist also in den Ländern unterschiedlich ausgeprägt. Dieses betrifft sowohl den Aufgabenkreis Unterricht und Erziehung als auch den Aufgabenkreis der äußeren Verwaltung (Avenarius & Füssel 2010).

Eigener Unterricht (Aufgabenbereich 1)

In den Ländern Berlin, Brandenburg, Hessen, Niedersachsen und Nordrhein-Westfalen hat die Schulleiterin oder der Schulleiter grundsätzlich neben Leitungsaufgaben auch eine Unterrichtsverpflichtung. Diese wird je nach Schulform unterschiedlich bestimmt. In Niedersachsen wird darüber hinaus die „Leitungszeit“ gesondert behandelt. Das gilt auch für das Hamburger Arbeitszeitmodell, in dem „funktionsbezogene Aufgaben“ ausgewiesen werden. Dort entscheidet der Schulleiter zudem selbst über den Umfang des von ihr oder ihm zu erteilenden Unterrichts.

Unterrichtsbezogene Führung (Aufgabenbereich 1)

In Hamburg, Niedersachsen und Nordrhein-Westfalen gilt, dass der Schulleiter für die Qualitätsentwicklung und Qualitätssicherung zu sorgen hat. In Berlin, Brandenburg und Hessen ist bestimmt, dass auf die kontinuierliche Verbesserung respektive Weiterentwicklung der Unterrichts- und Erziehungsarbeit hinzuwirken ist. In Nordrhein-Westfalen legt die Dienstordnung ausdrücklich die Pflicht zur Hospitation, Be-

ratung und Unterstützung fest. In den Schulgesetzen aller sechs Länder ist die Schulleiterin oder der Schulleiter für die Erfüllung des Bildungs- und Erziehungsauftrags der Schule verantwortlich. Sie oder er hat sich über das Unterrichtsgeschehen zu informieren und die Lehrkräfte zu beraten.

Evaluation und Rechenschaftslegung (Aufgabenbereich 1)

Daten für die Schulstatistik zu liefern ist traditionell Aufgabe der Schulen und damit des Schulleiters. In allen sechs Ländern wird Rechenschaftslegung im Zusammenhang mit der internen Evaluation erwähnt. In Nordrhein-Westfalen wird sie auch als Aufgabe der Fachkonferenzen genannt und ist hier Teil des Berichtswesens in der selbstständigen Schule und der Fortbildungsplanung. In Hamburg ist die Rechenschaftslegung Teil der zwischen Schulaufsicht und Schulleitungen zu schließenden Ziel- und Leistungsvereinbarungen.

Schulprogrammarbeit und interne Evaluation werden in allen sechs Ländern miteinander verbunden. Die Schulen sind verpflichtet, die besonderen Ziele, Schwerpunkte und Organisationsformen ihrer pädagogischen Arbeit in einem Schulprogramm festzulegen, den Erfolg ihrer Arbeit in regelmäßigen Abständen durch interne Evaluation zu überprüfen und falls erforderlich konkrete Verbesserungsmaßnahmen zu planen und das Schulprogramm regelmäßig fortzuschreiben.

Schüler- und elternbezogene Arbeit (Aufgabenbereich 2)

In allen sechs Bundesländern wird darauf hingewiesen, dass Schülerinnen und Schüler in allen grundsätzlichen und wichtigen Schulangelegenheiten durch die Schulleiterin oder den Schulleiter zu informieren und zu beraten sind. Die Rechte der Eltern auf Information, Beratung und Beteiligung werden in allen sechs Ländern gesetzlich fixiert. Die Rechte des Schulelternbeirats in Hessen sind stärker ausgeprägt als in den übrigen Ländern.

Personalführung/-entwicklung (Aufgabenbereich 3)

Entscheidungen in Personalangelegenheiten blieben lange Zeit höheren Entscheidungsebenen vorbehalten und werden nur in kleinen Schritten der Schulleiterin oder dem Schulleiter übertragen (vgl. Avenarius & Füssel 2010; Terhart 2010). In allen sechs Ländern gibt es neben zentralen Einstellungsverfahren der Schulbehörden auch schulbezogene Verfahren. Die Länder unterscheiden sich aber erheblich hinsichtlich des Umfangs. Während z. B. in Brandenburg nur eine begrenzte Zahl von Schulen die Entscheidung selbst treffen kann und in Hessen 2011 lediglich von einem Drittel „schulscharf“ eingestellter Lehrkräfte die Rede ist, liegt die Entscheidung in Berlin, Hamburg, Niedersachsen und Nordrhein-Westfalen ganz überwiegend bei den Schulen selbst.

Was die Personalentwicklung angeht, so sind die Lehrkräfte der genannten Länder zur Fortbildung verpflichtet. Die Schulleiterin oder der Schulleiter hat auf die Fortbildung der Lehrkräfte hinzuwirken und auf die Einhaltung dieser Pflicht zu achten. Er kann die Lehrkräfte aus eigenem Recht verpflichten, an notwendigen Fortbildungsmaßnahmen für die Entwicklung der Qualität und Organisation der Schule teilzunehmen.

Die Schulleiterin oder der Schulleiter ist in allen sechs Ländern weisungsbe-rechtigt gegenüber den Lehrkräften ihrer Schule. Dieses Recht ist allerdings in den Ländern unterschiedlich ausgeprägt. So kann die Schulleitung in Hessen den Unter-richt der Lehrkräfte zwar jederzeit besuchen, in die Unterrichts- und Erziehungsar-beit darf sie aber nur bei einem Verstoß gegen Rechts- und Verwaltungsvorschrif-ten, allgemein anerkannte pädagogische Grundsätze oder Bewertungsmaßstäbe, verbindliche pädagogische Grundsätze des Schulprogramm und Konferenzbe-schlüsse eingreifen und die Weisung erteilen, diese Vorgaben zu beachten, darf aber nicht die Entscheidung der Lehrkraft ersetzen. In Berlin und Brandenburg ist sie zu Eingriffen in die Unterrichts- und Erziehungsarbeit der Lehrkräfte nur bei Verstoß gegen Rechts- und Verwaltungsvorschriften, Weisungen der Schulbehör-den, Beschlüssen der schulischen Gremien oder bei Mängeln in der Qualität der pädagogischen Arbeit verpflichtet. Die umfassenden Aufsichtsbefugnisse werden mit dem Demokratieprinzip begründet, wonach die Letztentscheidung eines dem Parlament verantwortlichen Verwaltungsträgers gesichert sein muss (BVerfGE 93, 37 [70 ff.]; Niehues & Rux 2013). Unstrittig ist jedoch, dass es Grenzen einer behördli-chen Reglementierung von der Natur des Unterrichtens her gibt (so bereits Oppermann 1969). Es hat sich inzwischen auch die Einsicht durchgesetzt, dass die einzel-ne Schule ihre Erziehungsaufgabe nicht hinreichend wahrnehmen kann, wenn sie ausschließlich zentral gesteuert wird (Avenarius 2010). Bestrebungen, die Fachauf-sicht durch eine bloße Rechtsaufsicht zu ersetzen, waren erfolglos. Die dienstlichen Beurteilungen galten lange Zeit als Domäne der Dienstbehörden selbst (Avenarius & Füssel 2010). Sie werden inzwischen jedoch in allen sechs Ländern überwiegend von der Schulleiterin oder dem Schulleiter erstellt. Sie obliegen in der Mehrzahl der untersuchten Länder den Schulbehörden dann, wenn es um die Besetzung von Funktionsstellen, insbesondere wenn es um die Stellvertretung der Schulleitung geht. Zuständig sind die Schulbehörden auch dann, wenn ein Wechsel des Dienst-herrn ansteht. Sie können sich die dienstliche Beurteilung vorbehalten, wenn erhebliche Zweifel an der Eignung von zu beurteilenden Lehrkräften bestehen.

Die dienstrechtlichen Befugnisse liegen in fünf der sechs hier vorgestellten Län-der überwiegend bei übergeordneten staatlichen Behörden. Allerdings sehen die dortigen Schulgesetze jeweils eine Übertragung dienstrechtlicher Befugnisse vor. Die Einzelheiten werden in untergesetzlichen Vorschriften geregelt. Dabei werden in einigen Ländern Befugnisse unterschieden, die allen Schulleiterinnen und Schullei-tern übertragen wurden, und solche, die nur für bestimmte Schulen gelten. Lediglich in Hamburg ist die Schulleiterin oder der Schulleiter in der Regel weitgehend Dienst-vorgesetzte oder Dienstvorgesetzter.

Disziplinarmaßnahmen sind grundsätzlich Sache der Dienstbehörde (Claussen et al. 2003). In Nordrhein-Westfalen hat die Schulleiterin oder der Schulleiter zwar je-den Verdacht eines Dienstvergehens der Schulaufsichtsbehörde zu melden, sie oder er hat aber keine eigenen Disziplinarbefugnisse. Auch in Berlin und Niedersachsen ist die Bearbeitung von Dienstaufsichtsbeschwerden und Disziplinarmaßnahmen in Grundschulen und Gymnasien Sache der übergeordneten staatlichen Behörden. In Hessen sind mündliche oder schriftliche missbilligende Äußerungen (Zurechtwei-

sungen, Ermahnungen, Rügen und dergleichen), die nicht ausdrücklich als Verweis bezeichnet werden, Aufgabe der Schulleiterin oder des Schulleiters. Dies gilt auch in Brandenburg für einen begrenzten Kreis von Schulen. In Hamburg kann die Schulleiterin oder der Schulleiter einen Verweis als Disziplinarmaßnahme aussprechen. Die Ahndung von Dienstvergehen angestellter Lehrkräften allerdings wird zunehmend der Schulleiterin oder dem Schulleiter übertragen.

Mittelbewirtschaftung (Aufgabenbereich 4)

Ein wichtiger Teil der Stärkung der Selbstständigkeit der Schule ist eine erweiterte finanzielle Eigenständigkeit. Eine Budgetierung von Personal- und Sachmitteln schafft die Möglichkeit zur Profilierung und zur Umsetzung von Maßnahmen, die zur Entwicklung und Förderung eines vielfältigen Schullebens beitragen (Hanßen 2005).

In Brandenburg, Hessen und Nordrhein-Westfalen kann das Land den Schulen Personalmittel zur eigenverantwortlichen Bewirtschaftung zuweisen. In Brandenburg wurde dieses für alle Schulen als Instrument zur Unterrichtsabsicherung festgelegt, weiter gehende Rechte erhielt nur ein begrenzter Kreis von Schulen. Die Zumessung von Lehrkräftestunden erfolgt in Berlin als idealtypische Bemessungsgrundlage der Unterrichtsversorgung, die in der Verantwortung der einzelnen Schule organisatorisch umgesetzt wird. Schulen in Hamburg und Niedersachsen bewirtschaften Budgets im Rahmen der ihnen nach den geltenden Bedarfsgrundlagen zugewiesenen Personalmittel. In Hamburg erhalten Schulen die Möglichkeit, die ihnen zugewiesenen Personalmittel im Rahmen der bestehenden haushalts- und personalrechtlichen Grenzen flexibel einzusetzen.

Die Bewirtschaftung der Sachmittel ist in den vier Flächenländern vergleichbar geregelt. Sie richtet sich nach den für die Schulträger geltenden haushalts- und kasensrechtlichen Bestimmungen, wobei Brandenburg, Hessen und Niedersachsen die Schulträger ausdrücklich dazu verpflichten, den Schulen Mittel der laufenden Verwaltung und Unterhaltung zur Verfügung zu stellen. In Berlin erhalten die Schulen diese Mittel im Rahmen ihrer sächlichen Verantwortung. In Hamburg werden Sachmittel im Selbstbewirtschaftungsfonds als Budget für die Schulen bewirtschaftet. Dabei sind die Mittel deckungsfähig und können unter bestimmten Bedingungen punktuell auch für Personalmittel verwendet werden.

Dezentrale Ressourcenverantwortung erschöpft sich nicht in der Verwaltung staatlich oder kommunal zugewiesener Mittel, sondern umfasst die aktive Beteiligung der Schulen bei der Mittelbeschaffung. In allen sechs Ländern dürfen die Schulen zur Erfüllung ihrer Aufgaben Zuwendungen von Dritten entgegennehmen (Avenarius & Füssel 2010). Diese müssen mit dem Bildungs- und Erziehungsauftrag der Schule vereinbar sein. Die Entscheidung trifft die Schulleiterin oder der Schulleiter. Die Schulkonferenz und der Schulträger sind nach jeweiligem Landesrecht zu beteiligen. Berlin und Brandenburg legen fest, dass Einnahmen- oder Ausgabenminderungen, die eine Schule durch eigenes Handeln erzielt, der Schule in voller Höhe verbleiben sollen.

Verwaltungs-/Organisationsaufgaben (Aufgabenbereich 4)

In allen sechs Ländern ist die Schulleiterin oder der Schulleiter für einen geordneten Schulbetrieb und damit für die Stundenpläne, den Einsatz der Lehrkräfte und die Vertretungspläne verantwortlich.

Vertretung der Schule nach außen (Aufgabenbereich 5)

In allen sechs Ländern ist die Schulleiterin oder der Schulleiter wie bisher verpflichtet, die Schule nach außen zu vertreten. Brandenburg bindet die Außenvertretung ausdrücklich an Beschlüsse der Schulleitung und der schulischen Gremien. In Hessen wird festgelegt, dass die Außenvertretung im Einvernehmen mit dem Schulträger zu erfolgen hat, wenn dessen Angelegenheiten berührt werden. Auch die Zusammenarbeit mit anderen Schulen ist in allen sechs Ländern Pflicht der Schulleiterin oder des Schulleiters. Niedersachsen regelt die Zusammenarbeit mit anderen Schulen für die Schulformen durch einen gesonderten Erlass. In Nordrhein-Westfalen wird die Zusammenarbeit mit Schulen in freier Trägerschaft ausdrücklich einbezogen. In Hamburg ist diese Zusammenarbeit in die Arbeit regionaler Bildungskonferenzen institutionell eingebunden. Aus der Pflicht, die Schule nach außen zu vertreten, folgt in allen sechs Ländern für die Schulleiterin oder den Schulleiter die Aufgabe der Öffentlichkeitsarbeit. Die jeweiligen Pressegesetze verlangen, Vertreterinnen und Vertretern der Presse Auskunft zu erteilen. In Hessen ist sie oder er dabei im Rahmen der jeweiligen Zuständigkeiten an die Beschlüsse der Schulkonferenz und der Gesamtkonferenz sowie an das Einvernehmen mit dem Schulträger gebunden. In Brandenburg und Nordrhein-Westfalen sind bei Angelegenheiten von besonderer Tragweite Abstimmungen mit der Schulaufsichtsbehörde oder dem Schulträger erforderlich.

4.2 Allgemeine Rahmenbedingungen für das Schulleitungshandeln

Auswahl, Qualifizierung und Fortbildung der Schulleiterinnen und Schulleiter

Die Bestellung der Schulleiterin oder des Schulleiters ist Sache der Schulaufsichtsbehörde (Avenarius & Füssel 2010) und darf wegen ihrer politischen Tragweite nicht generell der Regierungsverantwortung entzogen werden. Sie folgt verfassungsrechtlich verankerten beamtenrechtlichen Grundsätzen (Schnellenbach 2010).

Allgemein anerkannt ist aber, dass die Schulleiterin oder der Schulleiter für ihre oder seine Aufgabe gesondert qualifiziert werden muss. Weiter gehende Professionalitätsanforderungen sind dadurch gekennzeichnet, dass sie sich auf die Reflexion über Unterricht, über Bildung und Erziehung als Voraussetzung für die Bewertung der Arbeit anderer Lehrkräfte erstrecken und damit zugleich die Grundlage für die Beratung innerhalb der Schule bilden (Füssel 2009). In Berlin darf das Amt einer Schulleiterin oder eines Schulleiters nur übertragen werden, wenn die ausgewählte Lehrkraft an einer Qualifizierungsmaßnahme für künftige Schulleiterinnen und Schulleiter erfolgreich teilgenommen hat. Entsprechend wird in Hamburg zur Schulleiterin oder zum Schulleiter nur bestellt, wer Kenntnisse und Fähigkeiten erworben hat, die über die Ausbildung für das Lehramt hinausgehen und die für die Leitung einer Schule erforderlich sind. Entsprechende gesetzliche Regelungen fehlen

in den übrigen vier Ländern. Aber auch in Nordrhein-Westfalen ist die Bestellung zur Schulleiterin oder zum Schulleiter grundsätzlich von einer erfolgreich abgeschlossenen Teilnahme an einer Qualifizierungsmaßnahme abhängig. In Niedersachsen müssen Lehrkräfte, die sich erfolgreich um eine Schulleitungsfunktion beworben haben, an einer Erstqualifizierung für ihren neuen Tätigkeitsbereich teilnehmen. Die Qualifizierung der Schulleiterin oder des Schulleiters obliegt in Hessen der Führungsakademie im Landesschulamt, in Brandenburg dem Landesinstitut für Schule und Medien.

Die Aufgaben der Schulleiterin oder des Schulleiters sind im Zusammenhang mit neuen Steuerungsvorstellungen erweitert worden, ihr oder sein Status hat sich grundlegend verändert. Die klassische Lehrkräfteausbildung bildet für die mit der Leitung einer Schule verbundenen Aufgaben nicht hinreichend aus. Alle sechs Länder qualifizieren Lehrkräfte sowie Schulleiterinnen und Schulleiter für die neuen Aufgaben und haben entsprechende Bestimmungen und Angebote geschaffen. Dies geschieht jedoch in unterschiedlicher Weise. In Brandenburg kann eine Zusatzqualifikation in Schulmanagement erworben werden. Hessen, Niedersachsen und Nordrhein-Westfalen haben gestufte Verfahren zur Qualifizierung für die Übernahme von Funktionstellen in der Schule eingeführt. Die Teilnahme an entsprechenden Qualifizierungsangeboten ist in Berlin und Hamburg verpflichtend.

Einen gesetzlich abgesicherten Anspruch auf Supervision und Coaching gibt es in keinem der sechs Länder. Supervision und Coaching werden aber in allen sechs Ländern im Rahmen von Qualifizierungs- und Unterstützungsangeboten für Schulleiterinnen und Schulleiter angeboten.

Aufgabenabgrenzung zwischen Schule und Schulverwaltung im Hinblick auf Qualitätsentwicklung

Die Aufgabenverteilung zwischen Schule und Schulaufsichtsbehörden hat sich verändert. Selbstständigkeit der Schule und Qualitätssicherung stehen manchmal in einem schwer auflösbaren Spannungsverhältnis zueinander (Avenarius & Füssel 2010). Mit der Stärkung der Eigenverantwortung von Schulen haben die Länder von der Detailsteuerung Abstand genommen. Auf der anderen Seite haben sie aber durch Maßnahmen der Qualitätssicherung den Druck auf die Schule erhöht. Beides beeinflusst die Aufgabenstellung der Schulleiterin oder des Schulleiters. Während die Schulaufsicht in Nordrhein-Westfalen dreistufig aufgebaut ist, werden die Aufgaben in Brandenburg auf zwei Ebenen wahrgenommen. Hessen und Niedersachsen haben Landes-schulbehörden eingerichtet, die als nachgeordnete Behörden der für Schule zuständigen Ministerien schulaufsichtliche Aufgaben bündeln. Berlin verteilt die Aufgaben auf Hauptverwaltung und Bezirke, während in Hamburg die Behörde für Schule und Berufsbildung umfassend zuständig ist. In Brandenburg übt das für Schule zuständige Ministerium die Fach- und Dienstaufsicht unmittelbar gegenüber dem Landesschulamt aus und nur mittelbar gegenüber den Schulen. Entsprechend führt das zuständige Ministerium in Nordrhein-Westfalen die Aufsicht über die nachgeordneten Schulaufsichtsbehörden, entscheidet aber lediglich über Angelegenheiten von grundsätzlicher Bedeutung und sichert die landeseinheitlichen Grundlagen für die pädagogische und organisatorische Arbeit der Schulen und für ein leistungsfähiges Schulwesen.

Die besondere Bedeutung der Zusammenarbeit zwischen Schule und Schulaufsichtsbehörden wird demnach in allen sechs Ländern hervorgehoben und insbesondere auf die Qualitätsentwicklung bezogen. Dabei stehen in Hamburg und Nordrhein-Westfalen Stellung und Aufgabe der Schulaufsichtsbehörden im Vordergrund, während in Berlin, Brandenburg, Hessen und Niedersachsen die Qualitätsentwicklung vorrangig als Aufgabe der Schule angesehen wird. In Brandenburg, Niedersachsen und Nordrhein-Westfalen wird ausdrücklich betont, dass die Verantwortung der Schulleitungen für Erziehung und Unterricht von den übergeordneten staatlichen Stellen zu achten ist. In Nordrhein-Westfalen wird auf die Führungsverantwortung der Schulleitung hingewiesen.

Unterstützungssysteme für aufgabenbezogenes Schulleitungshandeln

Die Kosten für das Verwaltungspersonal in der Schule trägt in Hamburg das Land, in Berlin sowie in den vier Flächenländern grundsätzlich der Schulträger. In Brandenburg und Niedersachsen ist dieses im Schulgesetz ausdrücklich geregelt. In Hessen wird überlegt, Landesmittel für die Beschäftigung von zusätzlichem Verwaltungspersonal zur Entlastung der Schulleiterin oder des Schulleiters zur Verfügung zu stellen. In Nordrhein-Westfalen wurde eine Schulverwaltungsassistenz erprobt. Der hier vorliegende Abschlussbericht kommt zum Ergebnis, dass eine Übernahme in den Regelbetrieb aus fachlicher Sicht uneingeschränkt empfehlenswert ist.

5. Fazit zu Rechten und Pflichten der Schulleiterin und des Schulleiters aus schulrechtlicher Perspektive

Die Entwicklung der Rechte und Pflichten der Schulleiterin bzw. des Schulleiters und die Fortschreibung der allgemeinen Rahmenbedingungen des Schulleitungshandelns sind Teil tief greifender Änderungen des Schulverfassungsrechts. Der Weg von der „verwalteten“ Schule (Becker 1954) hin zur „selbstständigen“ oder „eigenverantwortlichen“ Schule ist weit. Eine „autonome“ Schule kann es wegen der Festlegung in Art. 7 Abs. 1 des Grundgesetzes verfassungsrechtlich nicht geben, die Schule ist in den deutschen Ländern aber längst „geschäftsfähig“ geworden, auch wenn die Eigenverantwortung dem föderativen Prinzip entsprechend unterschiedlich ausgeprägt ist.

Das Bundesverfassungsgericht hat in den 1960er-Jahren das „besondere Gewaltverhältnis“ als unvereinbar mit dem Grundgesetz angesehen (BVerfGE 33, 1) und den Lehrpersonen ebenso wie den Schülerinnen und Schülern sowie Eltern als Nutzer Rechte und Pflichten im Rahmen des Schulverhältnisses eingeräumt. Das hat die Stellung der Schulleiterin oder des Schulleiters als „verlängerter Arm“ des Staates verändert. Individuelle Rechte wurden durch kollektive Rechte ergänzt. Neben Entscheidungen der Leitungsperson traten Entscheidungen schulischer Gremien. Die Schule selbst wurde damit zu einem Akteur im gesellschaftlichen Rahmen. Begriffe wie Öffnung der Schule verdeutlichen diese Entwicklung. Diese Schule bedurfte nach den Regeln der „Neuen Steuerung“ einer starken Leitung, die sich nicht nur um Unterricht und Erziehung, sondern auch um Personalentwicklung, Personalführung sowie Personal- und Sachmittel zu kümmern hat.

Die unter dem Gesichtspunkt der Partizipation in Schulen entwickelten Mitwirkungsstrukturen haben nach Abkehr von traditionellen Steuerungsformen und Hinwendung zur Kontextsteuerung eine neue Bedeutung erlangt. Viele Entscheidungen der selbstständigen Schulen bedürfen der Beteiligung oder Zustimmung schulischer Gremien. Hierbei ist die Frage zu stellen, ob sich die Balance zwischen Schulleiterin und Schulleiter, Schulleitung, Steuerungsgruppe und Gremien der Schule effizient gestaltet oder ob mit z. T. „komplizierteren“ Leitungs- und Gremienstrukturen auch in längeren Entscheidungsprozessen überhaupt ein sicheres Ergebnis erzielt werden kann. Bezüglich des Verhältnisses von Zeitaufwand und Entscheidungssicherheit ist somit weiterer Klärungsbedarf angezeigt.

Gegenwärtig ist nicht absehbar, inwieweit sich die in den Länderbestimmungen angelegte Balance zwischen der Stärkung schulischer Eigenständigkeit und der qualitätsbezogenen Rechenschaftspflicht (einschließlich jeweils entsprechender Unterstützungssysteme) in der Praxis als tragfähig erweist. Es verstärkt sich der Eindruck, dass Rechenschaftspflichten stärker der schulischen Qualitätssicherung und -entwicklung dienen, als dies bei den „neuen“ erweiterten Handlungsspielräumen der Fall zu sein scheint. Die (künftige) Funktion von mehr schulischer Eigenständigkeit wird in den rechtlichen Regelungen nicht mehrheitlich qualitätsbezogen interpretiert und signalisiert. Hiermit ist der Zielkonflikt zwischen Selbstständigkeit und Qualitätskontrolle angesprochen. Das ist allerdings keine Frage der (juristischen) Verhältnismäßigkeit, sondern eine politische Entscheidung, die, wenn sie ausschließlich zugunsten der Qualitätskontrolle ausginge, mögliche Erfolge einer stärkeren Selbstständigkeit in der Tat wieder einschränkt.

Daher ist ein weiterer Klärungsbedarf notwendig, wenn es um die Verhältnismäßigkeit der Rechte und Pflichten von Schulleiterinnen und Schulleitern geht.

Die Aufgaben des Dienstvorgesetzten werden in den deutschen Ländern in unterschiedlicher Weise auf die Schulleiterin bzw. den Schulleiter übertragen. Lediglich in Hamburg kann bislang davon gesprochen werden, dass der Schulleiter die Funktion des Dienstvorgesetzten übernommen hat. So kann die Schulleiterin oder der Schulleiter auch nur in eingeschränkter Weise personalrechtliche Befugnisse der Schule bei Einstellungen und anderen dienstrechtlichen Entscheidungen wahrnehmen. Auch wenn Einstellungen vereinzelt auf der Basis schulbezogener Ausschreibungen vorgenommen werden, liegt das Letztentscheidungsrecht zumeist bei den Schulaufsichtsbehörden. Zugleich ist die Schulleiterin oder der Schulleiter bezüglich der Qualitätsentwicklung ihrer oder seiner Schule auch weiterhin nicht nur angehalten, mit den Schulaufsichtsbehörden Rücksprache zu halten, sondern auch deren Placet abzuwarten. Dieser personelle Spielraum wird aber benötigt, wenn beispielsweise das Schulprogramm in personeller Hinsicht umgesetzt werden soll. Dies lässt eine verstärkte gleichzeitige Berücksichtigung schulinterner wie -externer Gruppeninteressen notwendig erscheinen. Hier wird deutlich, dass eine schulische Qualitätssicherung und -entwicklung ohne die Kommunikation und Be(ob)achtung sozialer Ebenen und Prozesse durch den „Moderator Schulleitung“ nur bedingt realisierbar erscheint.

Eine Entlastung der Schulleiterin bzw. des Schulleiters von unterrichtlichen Aufgaben ist insbesondere in Hamburg und Niedersachsen zu erkennen. Im Hinblick auf die Wahrnehmung von Managementaufgaben bleibt jedoch die Frage offen, ob

die weiter zuständigen staatlichen und kommunalen Stellen als Serviceeinrichtungen organisiert werden können. So besteht beispielsweise die Möglichkeit und Notwendigkeit, dass bei der Einstellung von Lehrkräften die Entscheidung zwar in der Schule getroffen wird, die in diesem Zusammenhang notwendigen Verwaltungstätigkeiten (von der Ausschreibung bis zur Klärung der Besoldung und evtl. Schutzrechte bei Behinderungen u. Ä.) von der unteren Schulbehörde (Schulaufsicht) erledigt werden. Bei der Schulverwaltung (Schulträger) kann entsprechend die Entscheidung über Baumaßnahmen und Investitionen (bis zu einer bestimmten Höhe der nötigen Ausgaben) in der Schule erfolgen, die Abwicklung des Vertragswerks aber in der Behörde. Auch stellt sich die Frage, ob diese bereit sind, den damit einhergehenden Rollenwechsel beispielsweise in Form gemeinsamer Auswertungen der einzelschulischen Rechenschaftslegungen zu vollziehen.

Eine Abstimmung oder aber Verzahnung externer wie interner Evaluationsverfahren, und sei es nur hinsichtlich ihres zeitlichen Turnus, ist schulrechtlich vorgesehen, in der Praxis aber noch wenig entwickelt. Dies gilt in besonderer Weise für die Einführung neuer Steuerungsinstrumente wie Zielvereinbarung, Mitarbeitergespräch sowie Entwicklung und Fortschreibung des Schulprogramms.

Gegenwärtig scheint offen zu sein, ob und in welchem Ausmaß es mit den aufgrund der gestärkten Eigenverantwortung einhergehenden zusätzlichen Rechenschafts- und Evaluationspflichten gelingt, insbesondere den Schulleitungen den Eindruck zu vermitteln, dass die juristische und die pädagogische Perspektive auf das Mehr an Eigenverantwortlichkeit im Schulalltag in Einklang stehen können und müssen (vgl. Rux 2008). Dies wirft die Frage auf, inwieweit eine eigenständige Profilierung der Schulen realisiert werden kann, die nicht ausschließlich auf kurzfristige Ergebnisse hin orientiert ist, sondern auf das mittel- und langfristige Erreichen strategischer und konkret benannter Ziele.

6. Ausblick

Hinsichtlich des zeitlichen Aufwandes für bestimmte Tätigkeitsbereiche schulleiterischen Handelns konnten durch die in der quantitativen Feldphase (Untersuchungsphase 2) durchgeführten Analysen Einflussfaktoren auf der System- und Organisationsebene verzeichnet werden. Individuelle Einflussfaktoren hingegen konnten hingegen nicht festgestellt werden. Es schienen somit vor allem die Rahmenbedingungen des Schulleitungshandelns zu sein, die in den Tätigkeitsprofilen der Schulleiterinnen und Schulleiter zum Ausdruck kommen. Statistisch bedeutsame Unterschiede zeigten sich hinsichtlich der Zeitanteile, die für Verwaltungs- und Organisationsaufgaben aufgewendet wurden. Demgegenüber wirkten sich auf das individuelle Belastungserleben ausschließlich Organisations- und Individualmerkmale aus. Ein statistisch bedeutsamer Einfluss auf Systemebene, d. h. zwischen den Bundesländern, war in Bezug auf die Beanspruchungsmaße nicht festzustellen. Dies galt auch für länderspezifische Vergleiche innerhalb der Schulformen.

Weitere erste empirische Befunde aus SHaRP weisen ferner darauf hin, dass sich in den Tätigkeitsfeldern Organisationsführung und -entwicklung, Personalführung

und -entwicklung sowie unterrichtsbezogene Führungsarbeit der zeitliche Aufwand und das Beanspruchungsempfinden von Schulleiterinnen und Schulleitern deutlich voneinander unterscheiden. Zugleich sind es gerade diese drei entwicklungsbezogenen Tätigkeitsfelder, denen die Schulleiterinnen und Schulleiter aus SHaRP die höchste Handlungspriorität einräumen. Es zeigt sich, dass nur äußerst vereinzelt Zusammenhänge zwischen dem, was Schulleiterinnen und Schulleiter tun und dabei empfinden, einerseits und ihrem personen- wie sachbezogenen Führungsverhalten andererseits existieren. Ferner ist es der Grad einzelschulischer Problemlagen, seien es strukturelle Defizite (z. B. bedarfsgerechte Unterrichtsversorgung) oder aber soziale Komponenten (mangelnde Lernförderung durch die Eltern), der die Ausbildung entwicklungsbezogenen Schulleitungshandelns deutlich einengen kann. Diese Befunde in ihrer Gesamtheit machen deutlich, dass man noch stärker als bisher nach den system-, organisations- und personenbezogenen Bedingungen für die Ausübung, vor allem aber für die Nichtausübung jener führungs- wie entwicklungsbezogenen Tätigkeiten fragen muss, denen im Rahmen „Neuer Steuerungsansätze“ eine besondere Bedeutung bei der Sicherung und Steigerung einzelschulischer Qualität zugeschrieben wird. So zielt die Ausübung dieser Tätigkeiten darauf ab, sowohl in der Organisation Schule insgesamt als auch bei den einzelnen Mitarbeitern bislang ungenutzte Ressourcen, Fähigkeiten, Begabungen und Handlungsmöglichkeiten zu erkennen, aufzuspüren und dabei gleichzeitig Freiräume, Anreize und günstige Entwicklungsbedingungen zu schaffen.

Literatur

- Autorengruppe Bildungsberichtserstattung (2008). *Bildung in Deutschland 2008. Ein indikatorengestützter Bericht mit einer Analyse zu Übergängen im Anschluss an den Sekundarbereich I*. Bielefeld: Bertelsmann.
- Autorengruppe Bildungsberichtserstattung (2012). *Bildung in Deutschland 2012. Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Bielefeld: Bertelsmann.
- Avenarius, H. & Füssel, H.-P. (2010). *Schulrecht. Ein Handbuch für Praxis, Rechtsprechung und Wissenschaft* (8., neu bearb. Aufl.). Köln u. a.
- Becker, H. (1954). Die verwaltete Schule. Gefahren und Möglichkeiten. *Recht der Jugend und des Bildungswesens*, 41 (1993) 2, S. 130–147.
- Blossfeld, H.-P., Bos, W., Daniel, H.-D., Hannover, B., Lenzen, D., Prenzel, M. & Wößmann, L. (2010). *Bildungsautonomie. Zwischen Regulierung und Eigenverantwortung. Jahresgutachten 2010*. Wiesbaden.
- Bolman, L. G. & Deal, T. E. (1992). Leadership and Management Effectiveness: a Multi-Frame, Multi-Sector Analysis. *Human Resource Management*, 30, S. 509–534.
- Brauckmann (2014). Ergebnisbericht im Rahmen des BMBF-Forschungsschwerpunkts „Steuerung im Bildungssystem“ (SteBis) geförderten Forschungsprojekts „Schulleitungshandeln zwischen erweiterten Rechten und Pflichten (SHaRP)“. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung-

- Brauckmann, S. (2012). Schulleitungshandeln zwischen deconcentration, devolution und delegation (3D). Empirische Annäherungen aus internationaler Perspektive. *Empirische Pädagogik* 26 (1), S. 76–100.
- Claussen, H. R., Benneke, F. & Schwandt, E.-A. (2003). *Das nichtförmliche Disziplinarverfahren (Verfahren des Dienstvorgesetzten): Leitfaden mit einer Übersicht über das materielle Disziplinarrecht und einer Mustersammlung* (4. Aufl.).
- Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., Brown, E., Ahtaridou, E. & Kington, A. (2009). *The Impact of School Leadership on Pupil Outcomes* (Research Report RR108). London: DCSF.
- Füssel, H.-P. (2009). *Schulaufsicht und Schulentwicklung – die neue Rolle der Schulleitung in einer selbständigen Schule aus juristischer Sicht*. Online unter: https://www.bllv.de/fileadmin/Dateien/Land-PDF/BLLV/Fuessel.Schulaufsicht_und_Schulentwicklung-1.pdf
- Hanßen, K. (2005). Mit Instrumenten der Verwaltungsreform zu mehr Qualität in Schule. In Buchen, H., Horster, L. & Rolff, H.-G. (Hrsg.), *Schule auf dem Weg zur Selbstständigkeit. Als Schulleitung mehr Verantwortung tragen und bewusst gestalten*, S. 90–108. Stuttgart: RAABE.
- Hanßen, K. (2011). *Rechtliche Regelungen zu Tätigkeitsfeldern von Schulleiterinnen und Schulleitern bei erweiterter Eigenverantwortung von Schulen. Eine Untersuchung der Rechtslage in den Ländern Bayern, Hessen und Nordrhein-Westfalen*. Online unter: http://www.dipf.de/de/forschung/projekte/pdf/steufi/Endfassung_BY_HE_NW.PDF
- Hanßen, K. (2012). *Rechtliche Regelungen zu Tätigkeitsfeldern von Schulleiterinnen und Schulleitern bei erweiterter Eigenverantwortung von Schulen. Eine Untersuchung der Rechtslage in den Ländern Brandenburg und Hamburg (Ergänzung zu der Untersuchung der Rechtslage in den Ländern Bayern, Hessen und Nordrhein-Westfalen)*. Online unter: http://www.dipf.de/de/forschung/projekte/pdf/steufi/Endfassung_BB_HH.PDF
- Hanßen, K. (2013a). *Rechtliche Regelungen zu Tätigkeitsfeldern von Schulleiterinnen und Schulleitern bei erweiterter Eigenverantwortung von Schule. Eine Untersuchung der Rechtslage in den Ländern Berlin und Niedersachsen (Ergänzung zu der Untersuchung der Rechtslage in den Ländern Bayern, Hessen und Nordrhein-Westfalen sowie in den Ländern Brandenburg und Hamburg)*.
- Hanßen, K. (2013b). *Bildungsrecht unter dem Aspekt rechtlich selbstständiger Schule im Spannungsverhältnis zwischen staatlicher und regionaler Verantwortung. Zeitschrift für Bildungsverwaltung*, 29 (2013) 2, S. 33–43.
- Müller, J. (2012). Die eigenverantwortliche Schule. In Sauerland, F. & Uhl, S. (Hrsg.), *Selbstständige Schule: Hintergrundwissen und Empfehlungen für die eigenverantwortliche Schule und die Lehrerbildung*, S. 1–14.
- Oppermann, T. (1969). *Kulturverwaltungsrecht. Bildung – Wissenschaft – Kunst*. Tübingen: Mohr Siebeck.
- Organisation for Economic Co-operation and Development OECD (2008). *Improving School Leadership Policy and Practice*. Paris: OECD.
- Richter, I. (1999). *Die sieben Todsünden der Bildungspolitik*. München, Wien: Hanser.

- Rürup, M. (2007). *Innovationswege im deutschen Bildungssystem. Die Verbreitung der Idee „Schulautonomie“ im Ländervergleich*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rux, J. & Niehues, N. (2013). *Schulrecht* (5., vollständig neu bearbeitete Aufl.). München.
- Schnellenbach, H. (2013). *Beamtenrecht in der Praxis* (8. Aufl.). München: C.H. Beck.
- Stock, M. (2002). Auf dem mühsamen Weg zur „Selbstständigen Schule“ – ein Modellversuch in Nordrhein-Westfalen im Zeichen der PISA-Debatte, *RdJB* 4, S. 468–494.
- Terhart, E. (2010). Personalauswahl, Personaleinsatz und Personalentwicklung an Schulen. In Altrichter, H. & Maag Merki, K. (Hrsg.), *Neue Steuerung im Schulsystem*, S. 255–275). Wiesbaden: Springer Verlag.
- vbw – Vereinigung der Bayerischen Wirtschaft e.V. (2010). *Bildungsautonomie: Zwischen Regulierung und Eigenverantwortung – die Bundesländer im Vergleich. Expertenrating der Schul- und Hochschulgesetze der Länder zum Jahresgutachten 2010*. München: VS Verlag.
- Wößmann, L., Lüdemann, E., Schütz, G. & West, M. R. (2007). School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003, *Education Working Paper, No. 13 EDU/WKP 8*.

Autorinnen und Autoren

Anabel Bach, wissenschaftliche Mitarbeiterin am Arbeitsbereich Schulpädagogik/Schulentwicklungsforschung der Freien Universität Berlin

Johannes Bellmann, Prof. Dr., Professur für Allgemeine Erziehungswissenschaft an der Westfälischen Wilhelms-Universität Münster

Carmen Binnewies, Prof. Dr., Professur für Arbeitspsychologie an der Westfälischen Wilhelms-Universität Münster

Oliver Böhm-Kasper, Prof. Dr., Professur für Forschungsmethoden und Jugendforschung an der Universität Bielefeld

Stefan Brauckmann, Prof. Dr., Professur für Qualitätsentwicklung und Qualitätssicherung im Bildungsbereich am Institut für Unterrichts- und Schulentwicklung der Alpen-Adria-Universität Klagenfurt

Thomas Brüsemeister, Prof. Dr., Professur für Soziologie mit Schwerpunkt Sozialisation und Bildung an der Justus-Liebig-Universität Gießen

Marten Clausen, Prof. Dr., Professur für Unterrichtsforschung an der Universität Duisburg-Essen

Kathrin Dederling, Prof. Dr., Professur für Bildungsinstitutionen und Schulentwicklung an der Universität Erfurt

Denise Demski, wissenschaftliche Mitarbeiterin am Institut für Pädagogik der Universität Duisburg-Essen

Fabian Dietrich, Dr., wissenschaftlicher Mitarbeiter am Institut für Erziehungswissenschaft der Leibniz Universität Hannover

Christian Dormann, Prof. Dr., Professur für Wirtschaftspädagogik an der Johannes Gutenberg-Universität Mainz

Christina Funke, wissenschaftliche Mitarbeiterin in der Arbeitsgruppe Bildungsforschung am Institut für Pädagogik der Universität Duisburg-Essen

Laura Fölker, wissenschaftliche Mitarbeiterin an der Fakultät für Bildungswissenschaften, AG Migrations- und Ungleichheitsforschung, der Universität Duisburg-Essen

Tanja Graf, wissenschaftliche Mitarbeiterin am Arbeitsbereich Weiterbildung und Bildungsmanagement der Freien Universität Berlin

Lisa Gromala, wissenschaftliche Mitarbeiterin am Institut für Soziologie der Justus-Liebig-Universität Gießen

Martin Heinrich, Prof. Dr., Professur für Schulentwicklung und Schulforschung sowie Leiter der Wissenschaftlichen Einrichtung Oberstufen-Kolleg an der Universität Bielefeld

- Thorsten Hertel**, wissenschaftlicher Mitarbeiter an der Fakultät für Bildungswissenschaften, AG Migrations- und Ungleichheitsforschung, der Universität Duisburg-Essen
- Anna Rosa Koch**, Dr., wissenschaftliche Mitarbeiterin am Institut für Arbeitspsychologie an der Westfälischen Wilhelms-Universität Münster
- Harm Kuper**, Prof. Dr., Professur für Weiterbildung und Bildungsmanagement an der Freien Universität Berlin sowie Dekan des Fachbereichs Erziehungswissenschaft und Psychologie der Freien Universität Berlin
- Bastian Laier**, wissenschaftlicher Mitarbeiter am Institut für Soziologie der Johannes Gutenberg-Universität Mainz
- Maike Lambrecht**, wissenschaftliche Mitarbeiterin an der Universität Bielefeld, Fakultät für Erziehungswissenschaft, AG 4 „Schulentwicklung und Schulforschung“
- Uwe Maier**, Prof. Dr., Professur für Erziehungswissenschaft und empirische Schulforschung an der Pädagogischen Hochschule Schwäbisch Gmünd
- Olga Mater**, wissenschaftliche Mitarbeiterin am Lehrstuhl für Wirtschaftspädagogik der Johannes Gutenberg-Universität Mainz
- Barbara Muslic**, wissenschaftliche Mitarbeiterin am Arbeitsbereich Weiterbildung und Bildungsmanagement der Freien Universität Berlin
- Hans Anand Pant**, Prof. Dr., Professur für Erziehungswissenschaftliche Methodenlehre an der Humboldt-Universität zu Berlin sowie Geschäftsführer der „Deutschen Schulakademie“
- Nicole Pfaff**, Prof. Dr., Professur für Ethnische Differenzierung und Heterogenität an der Universität Duisburg-Essen
- Marc Piopiunik**, Dr., wissenschaftlicher Mitarbeiter am ifo Institut, ifo Zentrum für Bildungsökonomik, München
- Peter Preisendörfer**, Prof. Dr., Professur für Soziologie, Arbeitsbereich Organisation von Arbeit und Betrieb, an der Johannes Gutenberg-Universität Mainz
- Bianca E. Preuß**, Dr., wissenschaftliche Mitarbeiterin am Institut für Soziologie der Justus-Liebig-Universität Gießen
- Daja Preuße**, Dr., wissenschaftliche Assistentin am Lehrstuhl für Wirtschaftspädagogik der Johannes Gutenberg-Universität Mainz
- Kathrin Racherbäumer**, Dr., wissenschaftliche Mitarbeiterin in der Arbeitsgruppe Bildungsforschung am Institut für Pädagogik der Universität Duisburg-Essen
- Carolin Ramsteck**, wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaft und empirische Schulforschung der Pädagogischen Hochschule Schwäbisch Gmünd
- Mathias Rürup**, Dr., wissenschaftlicher Mitarbeiter am Institut für Bildungsforschung der School of Education, Arbeitsbereich Empirische Bildungsforschung, der Bergischen Universität Wuppertal
- Anna Schliesing**, wissenschaftliche Mitarbeiterin am Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin

- Uwe Schmidt**, Dr., Leiter des Zentrums für Qualitätssicherung und -entwicklung (ZQ) und der Geschäftsstelle des Hochschulevaluierungsverbundes an der Johannes Gutenberg-Universität Mainz
- Sebastian Schweizer**, wissenschaftlicher Mitarbeiter am Institut für Erziehungswissenschaft an der Westfälischen Wilhelms-Universität Münster
- Guido Schwerdt**, Prof. Dr., Professur für Public Economics an der Universität Konstanz
- Odette Selders**, Dr., wissenschaftliche Mitarbeiterin am Institut für Forschungsmethoden und Jugendforschung der Universität Bielefeld
- Martin Stump**, wissenschaftlicher Mitarbeiter am Lehrstuhl für Wirtschaftspädagogik der Johannes Gutenberg-Universität Mainz
- Jasmin Tarkian**, wissenschaftliche Mitarbeiterin am Arbeitsbereich Schulpädagogik/Schulentwicklungsforschung der Freien Universität Berlin
- Corrie Thiel**, wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaft an der Westfälischen Wilhelms-Universität Münster
- Felicitas Thiel**, Prof. Dr., Professur für Schulpädagogik/Schulentwicklungsforschung an der Freien Universität Berlin
- Katja Thillmann**, Dr., wissenschaftliche Mitarbeiterin am Arbeitsbereich Schulpädagogik/Schulentwicklungsforschung der Freien Universität Berlin
- Isabell van Ackeren**, Prof. Dr., Professur für Bildungssystem- und Schulentwicklungsforschung an der Universität Duisburg-Essen sowie Prorektorin für Studium und Lehre an der Universität Duisburg-Essen
- Jochen Wissinger**, Prof. Dr., Professur für Erziehungswissenschaft mit dem Schwerpunkt Schulpädagogik an der Justus-Liebig-Universität Gießen
- Ludger Wößmann**, Prof. Dr., Professur für Volkswirtschaftslehre, insb. Bildungsökonomik, an der Ludwig-Maximilians-Universität München, Leiter des ifo Zentrums für Bildungsökonomik, München
- Sebastian Wurster**, wissenschaftlicher Mitarbeiter am Institut für Erziehungswissenschaft, AG Schulforschung/Schulpädagogik, der Johannes Gutenberg-Universität Mainz
- Lena Maria Zimmer**, wissenschaftliche Mitarbeiterin am Zentrum für Qualitätssicherung und -entwicklung (ZQ) der Johannes Gutenberg-Universität Mainz
- Olga Zlatkin-Troitschanskaia**, Prof. Dr., Professur für Wirtschaftspädagogik an der Johannes Gutenberg-Universität Mainz

Impressum

Herausgeber

Bundesministerium für Bildung und
Forschung (BMBF)
Referat Bildungsforschung
11055 Berlin

Bestellungen

schriftlich an
Publikationsversand der Bundesregierung
Postfach 48 10 09
18132 Rostock
E-Mail: publikationen@bundesregierung.de
Tel.: 030 18 272 272 1
Fax: 030 18 10 272 272 1

Stand

April 2016

Druck

M&E Druckhaus

Gestaltung

W. Bertelsmann, Bielefeld

Text

siehe Autorenverzeichnis

Diese Druckschrift wird im Rahmen der Öffentlichkeitsarbeit vom Bundesministerium für Bildung und Forschung unentgeltlich abgegeben. Sie ist nicht zum gewerblichen Vertrieb bestimmt. Sie darf weder von Parteien noch von Wahlwerberinnen/Wahlwerbern oder Wahlhelferinnen/Wahlhelfern während eines Wahlkampfes zum Zweck der Wahlwerbung verwendet werden. Dies gilt für Bundestags-, Landtags- und Kommunalwahlen sowie für Wahlen zum Europäischen Parlament. Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen und an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Unabhängig davon, wann, auf welchem Weg und in welcher Anzahl diese Schrift der Empfängerin/dem Empfänger zugegangen ist, darf sie auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl nicht in einer Weise verwendet werden, die als Parteinahme der Bundesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte.

