

“Gravierende Mängel” – Erste deutsche Erfahrungen mit der Benennung von Schulen schwacher Qualität durch die Schulinspektion

Norbert Sommer, Niedersächsisches Landesinstitut für schulische Qualitätsentwicklung, Hildesheim und Universität Osnabrück, Institut für kognitive Mathematik

Abstract

Niedersachsen hat in der ersten Runde systematischer Schulinspektionen zwischen Anfang 2006 und Ende 2012 Schulen benannt, die festgelegte Qualitätsnormen nicht erfüllen. Im Artikel werden Untersuchungsergebnisse zur Wirksamkeit dieser „Offizialisierung“ von Schwächen in Schulen dargestellt. Ausgewertet werden die Qualitätsbewertungen der Erst- und der in den schwach bewerteten Schulen ca. eineinhalb Jahre später durchgeführten Nachinspektionen, sowie Rückmeldungen von Schulleitern¹ im Anschluss an beide Inspektionen.

Einleitung

Schulen, die die in einer Qualitätsevaluation gesetzten Standards verfehlen, stehen in vielen Ländern im Fokus der Schulinspektion. Auch aufgrund finanzieller Erwägungen widmen traditionsreiche Inspektorate wie in England und in den Niederlanden schwachen Schulen in einem „proportionalen“ Vorgehen mehr Aufmerksamkeit (vgl. Matthews/Sammons, 2004, Ehren/Honingh, 2011)). In diesen Schulen scheint ein qualifiziertes Bildungsangebot für Schülerinnen und Schüler nicht gewährleistet und damit das Verfassungsziel gleicher Bildungschancen verletzt. Die Verbesserung dieser Schulen stellt möglicherweise eine besonders effiziente Maßnahme zur Anhebung des Leistungsniveaus des Schulsystems und daraus folgend des individuellen und gesellschaftlichen Nutzens dar (vgl. Psacharopoulos, 2007; The New York Times, 2012).

In Deutschland wird die Idee der datengestützten Schulentwicklung erst im Anschluss an die als zu schwach angesehenen PISA-Ergebnisse aus dem Jahr 2000 bildungspolitisch umgesetzt (vgl. KMK, 2006). Im Unterschied zu angloamerikanischen Staaten mit stärker marktorientierter Bildungspolitik, die league-tables des schulischen Abschneidens publizieren, werden Ergebnisse von Vergleichs- oder Abschlussarbeiten einzelner Schulen oder Regionen in Deutschland nicht veröffentlicht. Ihre Durchführung schlägt sich nur bei besonderen Vorkommnissen in der Presse nieder, z. B. wenn Lehrkräfte sich weigern, ihren Schülern aus sozialen Brennpunkten die Tests zuzumuten (vgl. Vieth-Entus, 2010) oder Fehler in Aufgabenstellungen entdeckt werden (vgl. WAZ, 2011).

In einem durch Ranglisten forcierten Wettbewerb zwischen Schulen wird in Deutschland kein geeignetes Mittel zur Verbesserung der Bildungsqualität gesehen und man fühlt sich in dieser Zurückhaltung durch kritische Analysen aus den USA bestätigt (vgl. Nichols/Berliner, 2007; Wolf/Janssens, 2007; Ravitch, 2010). Maßnahmen als Folge unzureichender Schulqualität wie Schulleiterwechsel, Schulschließung und anschließenden Neugründung (vgl. Quesel & Husfeldt, 2013) wären in Deutschland aufgrund des Beamtenstatus der Lehrkräfte sowie der Interessen der Schulgebäude und nicht lehrendes Personal finanzierenden Kommu-

¹ Im Artikel wird wegen der besseren Lesbarkeit nur die männliche Form verwendet. Selbstverständlich sind Schulleiterinnen, Schülerinnen, Lehrerinnen, Mitarbeiterinnen und Inspektorinnen immer mitgemeint.

nen sowie der Eltern an einer wohnortnahen schulischen Versorgung kaum durchzusetzen. Belastbare Daten darüber, ob die negativen angloamerikanischen Erfahrungen mit dem high-stakes-testing unter den anderen Rahmenbedingungen auf Deutschland übertragbar sind, fehlen.

Mit dem Ziel der Erhöhung von Schulqualität und Schülerleistung war nach der Einführung von Bildungsstandards und zentralen Leistungs- und Abschlusstests ab 2005 die Einrichtung von Schulinspektionsinstitutionen in allen sechzehn deutschen Bundesländern eine weitere bildungspolitische Maßnahme der „Steuerung über Analyse und Feedback“ (Dedering, 2012). In einem Bundesland wurde die Inspektion allerdings bereits wieder abgeschafft, in anderen steht sie auf dem Prüfstand. Immer wieder üben Lehrerverbände massiv Kritik, während die insbesondere die Wirtschaft die externe Evaluation als notwendig ansieht (vgl. Hessischer Philologenverband, 2013; United News Network, 2013).

Für die Schulpolitik sind die sechzehn Bundesländer verantwortlich. Dadurch existieren deutliche Strukturunterschiede in Deutschland nebeneinander (z. B. Anteil der Drei- oder Zweigliedrigkeit von Schulformen unterschiedlichen Anspruchs bzw. von Gesamtschulen, Zeitpunkt (Alter) der ersten externen Differenzierung, Länge der verpflichtenden Schulzeit bis zum mittleren Bildungsabschluss und zum Abitur). Auch die Inspektionsverfahren sind Länderentwicklungen, weisen aber trotz einiger Differenzen im Detail insgesamt eine große Ähnlichkeit auf (vgl. Rürup, 2008).

Ein fundamentaler Unterschied besteht darin, ob die Inspektion besonders schwach bewertete Schulen benennt oder nicht. Vier der 16 Bundesländer haben von Beginn an Schwellen definiert, ab wann die Schulqualität den Ansprüchen nicht mehr genügt. Mit Beginn der zweiten Runde haben einige weitere Länder eine solche Schwelle für schwache Schulen eingeführt. Nur Niedersachsen hat bisher über Ergebnisse und Entwicklungen schwacher Schulen berichtet (vgl. Sommer/Hornig, 2008/09; Sommer/Stöhr/Thomas, 2010; Schwank/Sommer, 2013). „...das bislang teils ignorierte, teils tabuisierte Thema des Versagens von Schule als Organisation“ (Huber, 2013, S. 8) gerät aber auch in Deutschland vermehrt in den Fokus.

Aus den USA, England und den Niederlanden liegen umfassende Erfahrungen mit als schwach bewerteten Schulen vor. Im Unterschied zur bisherigen Praxis in Deutschland kommt Schulleistungsergebnissen in den genannten Staaten ein herausgehobener Stellenwert bei der Schulqualitätsbewertung zu. Zwar liegen inzwischen in allen deutschen Bundesländern nationale Vergleichsarbeits- und landeseinheitliche Abschlussarbeitsergebnisse am Ende der Sekundarstufen vor, die Ergebnisse von Einzelschulen werden aber als vertraulich betrachtet („Die Daten gehören der Schule.“) und in den meisten Bundesländern auch den Schulinspektionen von der Bildungsadministration nicht zur Verfügung gestellt. Bewertet wird dazu in der Evaluation lediglich, ob Schulen Rückmeldungen zu Schülerleistungen schulintern diskutieren und zur Qualitätsentwicklung nutzen. Inspektionsergebnisse weisen die Umsetzung dieser Minimalanforderung nur für ca. 28 % aller Schulen aus (vgl. NLQ, 2012), was den Schluss zulässt, dass in vielen Schulen systematische Bemühungen zur Anhebung von Schülerleistungen nicht im Fokus stehen.

Auch die Inspektionsberichte der Einzelschulen waren im ersten Durchlauf in allen Bundesländern nicht öffentlich. Mit der in Inspektionen anderer Länder an erster Stelle genannten Funktion der professionellen und unabhängigen Information von Eltern über die Qualität von Schulen (vgl. Ofsted, 2014, S. 4) beginnen in Deutschland drei von 15 Bundesländern in ihrer zweiten Inspektionsrunde durch eine

eingeschränkte Veröffentlichung (z. B. Kurzform, Zeitverzögerung) der Inspektionsberichte der einzelnen Schulen (u.a. SenBJW, 2012).

Wo Leistungsergebnisse auf Schulebene ein wesentliches Bewertungskriterium der Schulqualität darstellen, werden Schulen in sozial belasteten Einzugsgebieten überproportional häufig als „failing“ identifiziert. Für einen Zusammenhang zwischen Inspektionsergebnis und sozialer Lage der Schule gibt es in Niedersachsen, wo innerschulische Prozesse und Unterricht im Mittelpunkt der Bewertungen stehen, keinen Hinweis. So sind auch die durch positive Selektion der Schüler leistungsmäßig gut aufgestellten Gymnasien mit vergleichbaren Anteilen unter den als schwach inspizierten Schulen vertreten (vgl. NLQ, 2012). Notwendige Daten für die Einschätzung eines zuverlässigen Sozialindex für Schüler, Schulklassen und Schulen, die eine Adjustierung von Leistungstestergebnissen und auch von Schulinspektionseinschätzungen gemäß wissenschaftlicher Ansprüche (vgl. Bonsen u.a., 2010) ermöglichen würden, liegen im betrachteten Bundesland nicht vor.

Schulische Leistungen sind nur zum Teil von Schule, Schulleitung und Lehrkräften abhängig. Einen erheblichen Anteil der Schulleistungsvarianz wird durch Lernvoraussetzungen der Schüler, ihre Ko-Produktion im Unterricht und die häusliche Unterstützung aufgeklärt (vgl. Hattie, 2009). Daher stehen Schulen, die wegen schwacher Schülerleistungen in den Fokus geraten, vor erheblichen Herausforderungen, die gesetzten Leistungsstandards, z. B. in Form von Schüleranteilen, die bestimmte Abschlussqualifikationen erzielen, zu erreichen. Neben zahlreichen Hinweisen auf ein Scheitern entsprechender innerschulischer Entwicklungsbemühungen und bildungsadministrativer Sanktionen (vgl. Rosenthal, 2004) liegen Studien vor, die Leistungsfortschritte von Schulen belegen, die als failing benannt wurden (vgl. Hussain, 2012; Allen/Burgess, 2012).

Die Rolle von Prozessvariablen für die Qualitätsbewertung von Schulen wird in angloamerikanischen Staaten kritisch gesehen: „The process should be secondary to the primary goal of improving student performance. It seems, however, that the secondary goal of establishing a process of turnaround often takes precedence over the primary goal of improved student achievement.“ (Meyers, 2013, S. 108) Landwehr betont demgegenüber, dass es für moderne komplexe Organisationen durchaus von Interesse ist, ob interne Prozesse „in ihrer Ausführungsqualität gewissen handlungsleitenden Werten entsprechen ...“ (Landwehr, 2011, S. 56)

Schwache Schülerleistungen oder geringe Leistungsfortschritte haben als Outputindikatoren unzureichender Schulqualität eine gewisse Evidenz, die der Definition von „Risikoschülern“ in PISA anhand von Kompetenzstufenmodellen (Deutsches PISA-Konsortium, 2001) vergleichbar ist. Normvorgaben, die überwiegend auf der Bewertung innerschulischer Strukturen und schulinterner Prozesse beruhen, wären mit Funktionsunfähigkeit der Schule, Gefahren für die Lernentwicklung der Schülerinnen und Schüler oder fehlender Kompetenz zur eigenständigen Bearbeitung erkannter Mängel zu begründen (vgl. Quesel u.a., 2013a, S. 12). Da Untersuchungen dazu fehlen, mangelt es Urteilsschwellen der Struktur- und Prozessqualität an empirischer Evidenz.

Ein mit Urteilsschwellen konfundierter Wert, der ebenfalls kritisch zu hinterfragen wäre, ist der in den Ländern der Bundesrepublik und international sehr unterschiedliche Anteil von Schulen, die als schwach benannt werden (vgl. Sommer & Hornig, 2008; Schwank & Sommer, 2013). In Berlin und Niedersachsen treten bei den Anteilen der als schwach identifizierten Schulen des jeweils ersten Inspektionszyklus Jahresschwankungen zwischen 0 und mehr als 10 % auf; in

gleicher Größenordnung bewegen sich in Niedersachsen und den Niederlanden entsprechende Unterschiede zwischen Schulformen (vgl. NLQ, 2012; The Dutch Inspectorate, 2013, S. 23). Der Anteil von failing-schools schwankte in England über sieben betrachtete Jahre zwischen 5,25 und 13,20 % (Allen/Burgess, 2012, Table 1) Internationale Vergleiche der Anteile schwacher Schulen und deren Schwankungen haben neben den Unterschieden der Kriterien periodische Modifikationen der Inspektionsverfahren zu berücksichtigen (vgl. BBC, 2010; Janssens, 2013).

Ziel der folgenden Analysen ist es, auf Grundlage der vorliegenden begrenzten Datenlage aus Inspektionsergebnissen und Rückmeldungen der Schulleiter in schriftlichen Befragungen die Wirksamkeit der „Offizialisierung“ schulischer Schwächen zu untersuchen. „Wirksamkeit“ wird im Sinne Landwehrs verstanden als „Eigenschaft/Fähigkeit von bestimmten Faktoren und Maßnahmen, die beabsichtigte Wirkung auch tatsächlich hervorzurufen.“ (Landwehr, 2011, S. 36) Ein zweiter Untersuchungsaspekt betrifft die kritische Einschätzung des Umfangs von Verbesserungen der als schwach inspizierten Schulen im Verlauf von eineinhalb Jahren bis zur Nachinspektion.

1. Inspektionsintentionen

Dem Inspektionsansatz in Niedersachsen liegt die Annahme zugrunde, dass die Rückmeldung von Qualitätseinschätzungen in Merkmalen, die sich an Erkenntnissen der Schuleffizienzforschung (u.a. Scheerens/Bosker, 2007) orientieren, Schulen einen Impuls für Verbesserungsmaßnahmen liefern. Von den vier „Wirkungsbereichen“ externer Schulevaluation, die Landwehr (2011) benennt, beruht die Wirkungserwartung vor allem auf der „Wissensgewinnung“. Im niedersächsischen Inspektionserlass heißt es: „Die Schulen leiten aus den Ergebnissen der Inspektion Maßnahmen zur Qualitätssicherung und -verbesserung ab und setzen an sie um.“ (MK Niedersachsen, 2006, Abs. 3.4(1)) Die Inspektionsteams sind angehalten nicht zu beraten und die Inspektionsbehörde ist nicht in Folgemaßnahmen von Schulen eingebunden.

Nur wenn die Qualitätseinschätzung die gesetzten Standards unterschreitet, gibt es in Niedersachsen Vorgaben für die drei anderen Bereiche Landwehrs:

- „Schulentwicklung“: der zitierte Absatz des Inspektionserlasses geht zwar für alle Schulen von Verbesserungsmaßnahmen aus; nur „schwache“ sind durch Zielvereinbarungen mit der Schulaufsicht und die angekündigte Nachinspektion explizit zu Entwicklungen verpflichtet
- „Rechenschaftslegung“: der Inspektionsbericht wird der Schulaufsicht als vorgesetzter Behörde und dem Schulträger, in der Regel der Kommune, zugestellt. Bei einer Bewertung unter Standard ist ein Auswertungsgespräch mit dem für diese Schule zuständigen Dezernenten der Schulaufsicht verpflichtend². Die Schulaufsicht soll Unterstützungsleistungen anbieten und die Umsetzung der Maßnahmenplanung begleiten. Die Rechenschaftslegung gegenüber der Öffentlichkeit ist begrenzt. Inspektionsergebnisse sind nur „schulöffentlich“ (Lehrkräfte, Mitarbeiter, Eltern, Schüler) mit großen Unterschieden zwischen

² Die Bewertung einer Schule durch die Schulinspektion kann vor allem im Falle besonders schwacher Schulqualität durchaus als Bewertung der Arbeit der Schulaufsicht gedeutet werden. Anfängliche Ressentiments der traditionellen Schulaufsicht gegenüber der neu eingerichteten Inspektionsbehörde könnten darin eine Ursache haben.

Schulen in der Art und Weise, wie Berichte der für Eltern und Schüler schulintern zugänglich gemacht werden, sowie hinsichtlich des Interesses der Adressaten.

- „Normendurchsetzung“: die Ebenen des Bildungssystems (Makro, Meso, Mikro, vgl. Fend, 2008) sind lose gekoppelt (Weick, 1976). Durchgriffsmöglichkeiten der Bildungspolitik auf die Einzelschule haben sich als begrenzt erwiesen. Schulen „rekontextualisieren“ (Fend, 2008, S. 174ff) Erlassvorgaben z. T. recht frei, wie negative Inspektionsergebnisse belegen: Lehrpläne sind auch Jahre nach dem Vorliegen der Bildungsstandards nicht überarbeitet, Konferenzen finden nicht regelmäßig statt, Leitbilder und Schulprogramm oder vorgeschriebene Ausschüsse (z. B. zum Arbeitsschutz) fehlen usw. Ähnliche Grenzen wie zwischen Makro- (Schulverwaltung) und Mesoebene (Schule) sind auch zwischen Meso- und Mikroebene (Unterricht) zu vermuten. Schulinterne Beschlüsse, z. B. zur Unterrichtsgestaltung, sind von der Schulleitung bei einer sich verweigernden Lehrkraft kaum durchzusetzen.

In der ersten Runde der Schulinspektion zwischen 2006 und 2012 genügten in Niedersachsen 93 % aller Schulen den Ansprüchen. In diesem Artikel stehen die ca. 7 % allgemein bildenden öffentlichen Schulen im Fokus, die mit dem Urteil „gravierende Mängel“³ und der Aufforderung zur Schulentwicklung konfrontiert wurden.

2. Die Inspektionsbewertung

Je nach Schulgröße besuchen zwei bis fünf Inspektoren eine Schule drei bis fünf Tage lang. 15 Kriterien der Schulqualität werden anhand von vorzulegenden Schuldokumenten, Interviews mit Schulleiter(in), Lehrkräften, Schülern, Eltern und sonstigen Mitarbeitern, Unterrichtseinsichtnahmen und Beobachtungen während des Aufenthalts in der Schule bewertet.

Wenn acht und mehr der 15 Qualitätskriterien oder aus der Teilmenge der vier Unterrichtskriterien drei oder alle vier die gesetzte Norm unterschreiten, beantragt das Inspektionsteam eine Nachinspektion.

Die folgende Abbildung zeigt, dass es deutliche Bewertungsunterschiede zwischen Schulen mit gravierenden Mängeln und solchen gibt, die die Mindeststandards für ausreichende Schulqualität erfüllen, aber auch eine deutlich erkennbare Überlappung der Gruppen.

³ Der Ausdruck „gravierende Mängel“ wird in Niedersachsen für Schulen verwendet, die die Inspektionsvorgaben nicht erfüllten. Ihnen teilt das Inspektionsteam unter Verweis auf das Profil der Qualitätsurteile am Ende der Inspektion mit, dass ein „Antrag auf Nachinspektion“ gestellt wird. Nach ca. 1½ Jahren wird in einer „Nachinspektion“ geprüft, ob die Schule sich in den als defizitär bewerteten Bereichen verbessern konnte. Die Begriffe „Schule mit gravierenden Mängeln“, „Schule mit Antrag auf Nachinspektion“, „Nachinspektionsschule“, „schwache Schule“ und „failing school“ werden im Text synonym für Schulen verwendet, die eine gesetzte Inspektionsnorm nicht erfüllen.

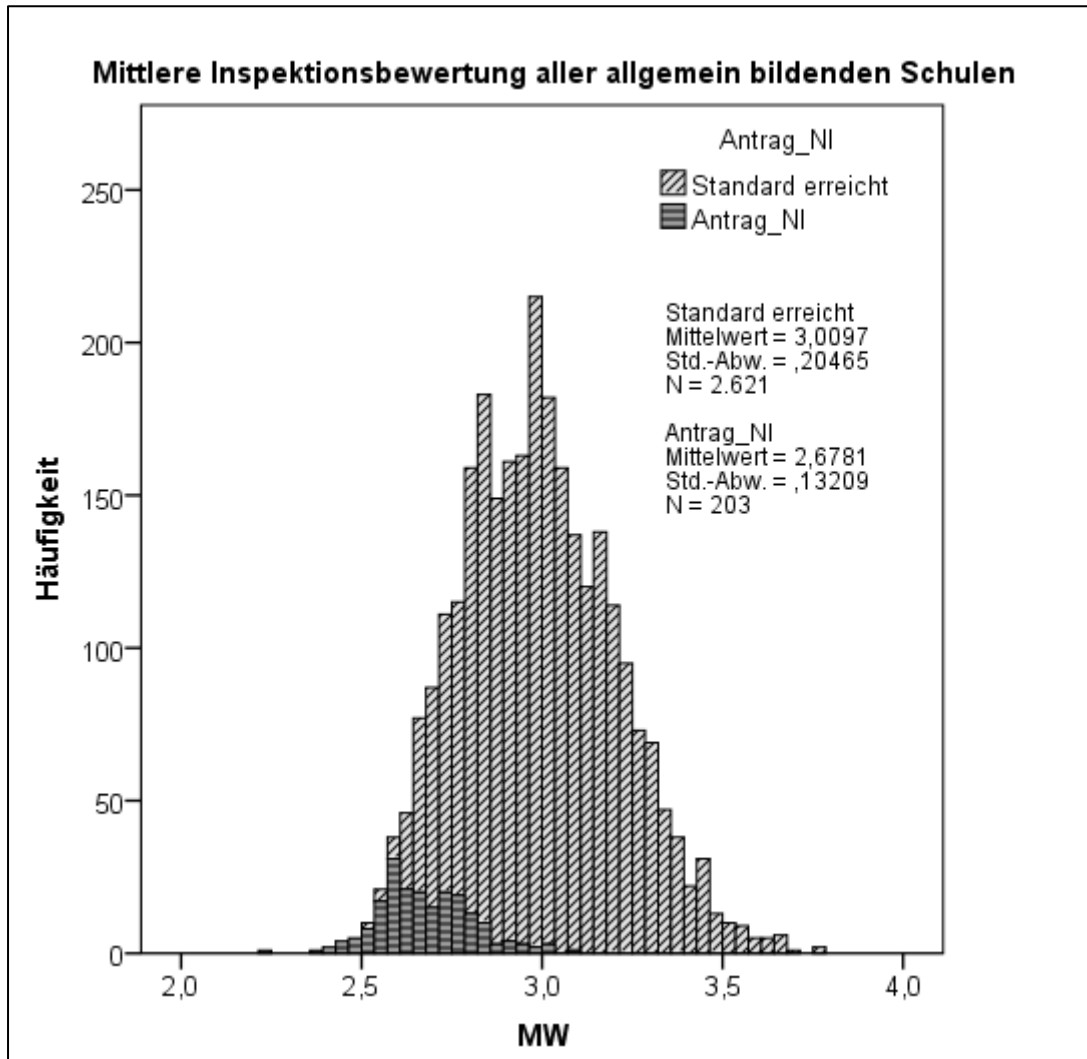


Abb. 1: Teilkriterienmittelwerte in der ersten Runde der Schulinspektion in Niedersachsen

Das Urteil gravierende Mängel basiert auf dem Profil der 15 Qualitätskriterien. Deren vierstufige Bewertung („stark“, „eher stark als schwach“, „eher schwach als stark“, „schwach“ - codiert mit 4, 3, 2, 1) errechnet sich aus jeweils vier bis zehn Teilkriterien, die im niedersächsischen Inspektionsverfahren der ersten Runde von den Inspektionsteams unmittelbar auf einer dreistufigen Skala mit 2 („-“ trifft nicht zu), 3 („+“ trifft zu) und 4 („++“ trifft in besonderem Maße zu) beurteilt werden, nicht rein arithmetisch, sondern nach vorgegebenen „Normen“ (s. u.). Die Überlappung in Abbildung 1 ist so deutlich, weil die Grafik auf den arithmetischen Mitteln der Schulen über alle ca. 95 Teilkriterieneinschätzungen beruht – nicht auf der normbasierten Bewertung.

Die mittlere Bewertung der Schulen, denen die Inspektion eine zumindest ausreichende Qualität bescheinigt, liegt bei ca. $M=3,0$. Sie erfüllen danach in allen Teilkriterien die gesetzten Standards. In der Praxis werden aber in jeder Schule mehr oder weniger Teilkriterien als defizitär eingeschätzt, was rechnerisch durch sehr gute Bewertungen in anderen Merkmalen ausgeglichen wird. Nachinspektionsschulen liegen am linken Rand der Verteilung mit einem signifikant niedrigeren Gruppenmittelwert ($p < 0,01$ ss, Cohens $d > 1,5$).

Für die Überlappung der beiden Gruppen von Schulen sind Normierungsvorgaben verantwortlich. Die Ausprägung eines Qualitätskriteriums wird als angemessen

angesehen, wenn jeweils mehr als 50 % der zugehörigen Teilkriterien positiv bewertet werden, darunter ein als besonders bedeutsam eingeschätztes Teilkriterium, das für eine positive Bewertung des Qualitätskriteriums unbedingt erfüllt sein muss (Ausschluss- oder k.o.-Kriterium). Diese Normierungsvorgaben wirken sich auf die Schulbewertung aus:

- Bei gleichem Mittelwert auf Teilkriterienebene unterscheiden sich zwei Schulen in der Bewertung eines Qualitätskriteriums, dem die Normierungsregeln zugrunde liegen, wenn anstelle des k.o.-Kriteriums ein anderes Teilkriterium nicht erfüllt ist.
- Wenn von zehn Teilkriterien sechs statt fünf hinreichend ausgeprägt sind steigt das arithmetische Mittel nur von $M=2,5$ auf $M=2,6$, das Qualitätskriterienurteil dagegen von 2 (unter Standard) auf 3 (Standard erreicht).
- Reicht die Qualität von drei der vier Unterrichtskriterien nicht aus, liegen „gravierende Mängel“ vor, während Schulen mit bis zu sieben negativen Urteilen in anderen Qualitätsbereichen die Inspektion ohne Konsequenzen durchlaufen. Der Anteil schwacher Schulen, die bei insgesamt weniger als acht negativ eingeschätzten Qualitätskriterien an der Unterrichtsnorm scheitern, beträgt ca. 50 %.

Nur in sehr wenigen Schulen werden mehr als neun der 15 Qualitätskriterien als nicht hinreichend bewertet.

3. Schulische Reaktionen auf das Inspektionsergebnis

Mit der Inspektion wurden erstmals in Deutschland alle Schulen in ihrer Schul- und viele Lehrkräfte (Vorgabe: mindestens 50 % aller Lehrkräfte einer Schule) in ihrer Unterrichtsqualität beurteilt. Zwar oblagen der Schulaufsicht zuvor entsprechende Aufgaben, sie war aber kapazitätsmäßig seit vielen Jahren nicht mehr so aufgestattet, dass Schulen oder Lehrkräfte vom zuständigen Dezernenten regelmäßig außerhalb von Lehramtsprüfungen, Stellenbesetzungsverfahren oder des Vorliegens gravierender Vorfälle (z. B. Elternbeschwerden) besucht wurden. Das von Lortie (1972, vgl. Eder u. a. 2011) für Lehrkräfte beschriebene „Autonomie-Paritäts-Muster“ der Annahme relativer Autonomie im Agieren und prinzipieller Gleichwertigkeit in der Funktionalität beschreibt vermutlich nicht nur die Haltung von Lehrkräften, sondern auch die von Schulen. In diesen Zustand greift die Schulinspektion ein, indem sie Daten über die Schulsituation bereitstellt, durch Hinweis auf Defizite Entwicklungsbedarf aufzeigt, durch ihre daran orientierten Kriterien die Einhaltung gesetzlicher Vorgaben überprüft und Rechenschaft einfordert. Dieser Eingriff von außen konfliktiert mit den Autonomie-Paritäts-Erwartungen von Schulen. Die Bewertung einer Schule unter Standard wird zumindest Schulleitern im regionalen Umfeld nicht verborgen bleiben, was Ursache dysfunktionale Reaktionen der betroffenen Schule auf die Inspektion sein könnte, z. B. Abwertung der Inspektoren oder Ignorieren der Inspektionsergebnisse (vgl. Internetdokument, 2009; Sommer, 2011; Schwank/Sommer, 2013).

Gemäß der gesetzlichen Vorgabe führt die Inspektionsbehörde in den Schulen Evaluationsbefragungen zum Ablauf der Inspektion durch. Seit Mitte 2007 werden allen schulischen Gruppen mit dem Endbericht über die Inspektion Fragebögen⁴

⁴ zunächst Papier-Versionen, seit 2009 Online-Befragung, moderate Veränderung der Fragebögen bei der Umstellung (Korrekturen, Austausch einiger Items)

zugesandt. In diesem Artikel beziehen wir uns nur auf die Rückmeldung der Schulleiter⁵.

Die Aussagen, die die Teilnehmer in der Regel vierstufig von „stimme zu“ bis „stimme gar nicht zu“ bewerten, beziehen sich auf:

- die schulische Vorbereitung auf die Inspektion
- das Erleben des Inspektionsablaufs
- die Einschätzung des Zutreffens der Qualitätseinschätzung
- im Anschluss an die Inspektion ergriffene bzw. aufgrund des kurzen Abstands zwischen Inspektion bzw. Übersendung des Endberichts und der Befragung in Planung befindliche Maßnahmen
- eine abschließende allgemeine Bewertung der Schulinspektion (in der erfahrenen niedersächsischen Variante).

Nur die Schulleiter sind gesetzlich verpflichtet, an der Befragung teilzunehmen. Die Inspektionsbehörde hat bei fehlenden Rückmeldungen zur Erstinspektionen aufgrund der hohen Inspektionszahlen nicht auf einer Beantwortung insistiert (Normendurchsetzungsdefizit in der Beziehung von Makro- und Mesoebene). Die Rückmeldequote der Schulleitungen beträgt ca. 53 %.

Aufgrund im Inspektionsprofil nicht abgebildeter schulischer Besonderheiten wird jedes Inspektionsergebnis vor Ort als nur eingeschränkt angemessen eingeschätzt werden können. Weil nach dem Autonomie-Paritäts-Muster eine Defizitbenennung schulischer Realität auf Basis eines einheitlichen Rahmens ungewohnt ist, sollten negative Reaktionen auf die Inspektionsurteile in Schulen ausgeprägter ausfallen, die das Urteil „gravierende Mängel“ erhalten haben, als in jenen, die die Mindestvorgaben erfüllen. Letztere nehmen das Ergebnis der Inspektion vermutlich zur Kenntnis, sind aber relativ frei hinsichtlich der Intensität der Auseinandersetzung mit der Rückmeldung und der Umsetzung ableitbarer Entwicklungsmaßnahmen, zumal in Niedersachsen lange Zeit das Ob, Wann und Wie einer zweiten Inspektionsrunde, die fehlendes Engagement hätte benennen können, unklar war.

Gemäß diesen Vorüberlegungen ist zu erwarten, dass

- die Zustimmung zur Inspektion umso höher ausfällt, je besser die Schule in der Inspektion bewertet wurde
- die Rückmeldung über schulische Planungen und ergriffene Maßnahmen um so geringer ist, je besser das Inspektionsurteil ausfällt
- das Urteil „gravierende Mängel“ die Bewertungstendenz verstärkt.

An zwei Items sei dargestellt, dass die Erwartung einer unterschiedlichen Bedeutungszuschreibung des Inspektionsergebnisses in Abhängigkeit von der Bewertung der Schule nicht ungerechtfertigt ist:

- *„Das Inspektionsteam hat sich während der Inspektion freundlich verhalten.“*: Bei einem Mittelwert von $M=3,83$ („volle Zustimmung“ 4,00) bestätigen nahezu alle 1409 rückmeldenden Schulen mit hinreichendem Inspektionsergebnis diese Aussage. Auch die 102 Schulen mit gravierenden Mängeln, die an der Befragung teilgenommen haben, werten überwiegend positiv, der Mittelwert ist aber mit $M=3,40$ signifikant geringer.
- *„Hinsichtlich der weiteren Schulentwicklung ist aufgrund der Inspektionsergebnisse ein detaillierter Plan festgelegt bzw. überarbeitet worden.“*: Die mittlere

⁵ Für eine Gesamtdarstellung der Rückmeldungsergebnisse s. NLQ, 2012 mit den Anhängen III und IV.

Zustimmung in den Schulen ohne Nachinspektionsantrag liegt mit $M=2,75$ nahe am theoretischen Mittelwert $M=2,50$. Aus Schulen mit gravierenden Mängeln wird mit dem Zustimmungswert $M=3,32$ eine signifikant höhere Planungsaktivität zurückgemeldet.

Der Schulleiterfragebogen ist umfangreich und die Auswertungen können daher auf Itemebene nicht vollständig wiedergegeben werden. Entsprechend dem Inhalt der Items sind Skalen gebildet worden, deren Zusammenstellung nach einer explorativen Faktorenanalyse überarbeitet wurde. Die folgende Tabelle stellt fünf Skalen mit befriedigender bis guter Reliabilität dar. Die Spalten beinhalten den Skalennamen, die Itemzahl, ein kennzeichnendes Item der Skala, die Reliabilität und, hier von besonderem Interesse, zwei Korrelationen, in die Fragebogenskalen und die Qualitätsbewertung auf Basis der Normierung eingehen. Die erste Korrelation ist für alle Schulen berechnet worden, die zweite nur für die Teilmenge der Schulen mit hinreichendem und besserem Inspektionsergebnis.

Über die beiden Spalten „Zwillinge“ wird in Abschnitt 4 berichtet.

Tab 1: Skalenbeschreibung und Zusammenhänge von Inspektions- und Nachbefragungsdaten

Skalenbezeichnung	Itemzahl	Typisches Item	Cronbachs α	Korrelationen mit	Zwillinge	Zwillinge
			MW s	Norm-QK Norm-QK nur Schulen ohne NI- Antrag	MW ohne NI- Antrag (N) mit NI-An- trag (N)	MW- Ver- gleich p Signifi- kanz Cohens d
1. Zufriedenheit mit dem Ablauf der Inspektion	12	Die Schule ist insgesamt zufrieden mit der Art und Weise, wie die Schulinspektion durchgeführt wurde.	0,89 3,62 0,42	.34 ss .30 ss	3,45 (66) 3,25 (70)	0,03 s 0,42
2. Bewertung der Berichtsqualität	8	Die Bewertungen/ Erläuterungen im Bericht sind in sich stimmig.	0,87 3,58 0,47	.41 ss .36 ss	3,45 (66) 3,22 (69)	0,02 s 0,46
3. Urteilsvalidität des Inspektionsverfahrens	4	Das Inspektionsverfahren kann die Qualität von Schulen (im Sinne des niedersächsischen Orientierungsrahmens Schulqualität“) gut einschätzen.	0,83 3,12 0,69	.27 ss .23 ss	3,04 (66) 2,80 (69)	0,19 ns 0,41
4. Wirkung der Schulinspektion – Einstellungen	4	Das Kollegium ist bereit, sich mit den durch die Schulinspektion rückgemeldeten Defiziten des Unterrichts auseinander zu setzen.	0,83 2,78 0,79	- .03ns - .02 ns	2,82 (57) 3,06 (62)	0,07 ns 0,35
5. Wirkung der Schulinspektion – Maßnahmen	3	Aufgrund der Ergebnisse der Schulinspektion sind erste Maßnahmen zur Schulentwicklung umgesetzt worden.	0,70 2,84 0,72	-.15 ss -.11 ss	2,80 (61) 3,23 (65)	0,00 ss 0,59

Die Zufriedenheit mit dem Inspektionsablauf (1.) und mit der Berichtsqualität (2.) lässt sich zuverlässig erfassen. Die mittlere Zustimmung ist mit $M = 3,62$ bzw. $M = 3,58$

sehr hoch. Die Korrelationen beider Skalen mit dem Inspektionsergebnisse sind substantiell. Das Urteil der Schulleiter über die Inspektion fällt besser aus, wenn das Inspektionsteam zu einem besseren Qualitätsurteil gelangt ist. Dieser Befund deckt sich mit Ergebnissen anderer Studien (vgl. Wolf/Janssens, 2007, S. 385).

Die mittlere Einschätzung der Urteilsvalidität der Inspektion(-sinstrumente) (3.) ist „eher zustimmend“ positiv, aber deutlich niedriger als die der konkreten Inspektionserfahrung in der Schule. Die Korrelation der Bewertung des Inspektionsverfahrens an sich mit dem Inspektionsurteil ist signifikant, entspricht in der Tendenz der Erwartung, fällt aber gering aus.

Die zweite Hypothese bezieht sich auf Wirkungen der Inspektion. Da, wie oben diskutiert, Leistungs- oder Zufriedenheitsdaten von Schülern nicht zu Verfügung stehen, kann die Wirkung der Schulinspektion nur anhand Struktur- und Prozessveränderungen untersucht werden. Hinsichtlich der unmittelbaren Wirkung der Schulinspektionsergebnisse auf die Arbeit in der Schule sind zwei Skalen gebildet worden:

- Ein Aspekt der „Wirkung auf Einstellungen“ (4.) betrifft die Einschätzung der Schulleitungen zur Aufgeschlossenheit ihrer Kollegien für die anstehende Schulentwicklungsarbeit als Voraussetzung konkreter Maßnahmen. Der Mittelwert der Skala liegt mit $MW=2,78$ knapp oberhalb des theoretischen Mittels, Zeichen einer nicht unerheblichen Skepsis vieler Schulleiter.
- Im zweiten Wirkungsaspekt geht es um den Beginn der konkreten Planung und Umsetzung von Maßnahmen zur Verbesserung aufgezeigter Schwächen (5.). Die Reliabilität ist noch zufriedenstellend. Die Zustimmung liegt im Mittel bei $M=2,86$. Sie ist unwesentlich höher als in Skala 4.

Die Interpretation der geringen Zustimmungswerte in den beiden Wirkungsskalen hat den kurzen Zeitraum von ca. drei bis vier Monate zwischen Inspektion und Abfrage zu berücksichtigen. Die Inspektionsbewertung wird am Ende des Schulbesuchs vorgestellt, der offizielle Bericht geht aber erst nach zwei bis drei Monaten an die Schule. Dann sind offensichtlich selbst die Phasen „Rezeption“ und „Reaktion“ eines umfassenden Inspektionswirkungsprozesses (vgl. Sommer, 2011, S. 152) in vielen Schulen noch nicht abgeschlossen und „Aktionen“ haben vielfach noch nicht begonnen.

Der vermutete negative Zusammenhang zwischen der Schulqualitätsbewertung und der Wirkung in Form von Maßnahmen der Schule - je besser die Schule bewertet wird, umso schwächer fällt die Reformbereitschaft aus - lässt sich nicht bestätigen. Die Vorzeichen sind zwar, wie erwartet, negativ, die Korrelationshöhen sind aber gering und überschreiten die 5 %-Signifikanzschwelle nur für die 5. Skala.

Um die besondere Wirkung des Urteils gravierende Mängel herauszuarbeiten, wurde eine weitere Korrelation nur für die Schulen berechnet, die die Inspektionsstandards erfüllen konnten. Es werden geringere Zusammenhänge zwischen Inspektionsurteilen und Befragungsskalen erwartet, da Schwächen nicht durch das folgenreiche Urteil „gravierende Mängel“ besonders betont werden. Die Erwartung sinkender gemeinsamer Varianz bestätigt sich für die fünf einbezogenen Skalen, die Unterschiede der Korrelationshöhen sind aber nicht signifikant. Einschränkend ist zu berücksichtigen, dass die Streuung der Inspektionsurteile in der Teilgruppe der als hinreichend bewerteten Schule etwas geringer ausfällt als in der Gesamtgruppe aller Schulen ($s=0,28$ gegenüber $s=0,31$), was die Reduktion der Korrelationshöhe

beeinflusst haben könnte. Wir werten das Ergebnis als einen Hinweis darauf, dass die Ankündigung einer Nachinspektion über der mittlere Qualitätsbewertung hinaus einen zusätzlichen Effekt bzw. der Bewertung der erlebten Inspektion ausübt, der deutlicher zu belegen ist.

Die negativeren Reaktionen auf die Inspektion in schwächer bewerteten Schulen kann, muss aber nicht Ausdruck mangelnder Kritikfähigkeit sein. Ein vertiefter Blick auf Inspektionsablauf und -ergebnis ist gerade von den Schulen zu erwarten, in denen Defizite benannt werden. Sie könnten auf tatsächliche oder vermeintliche Inkonsistenzen im Inspektionsverfahren sensibler reagieren. Nicht auszuschließen ist außerdem, dass in einigen der schwach beurteilten Schulen tatsächlich während der Inspektionsdurchführung Probleme, z. B. in der Kommunikation zwischen Inspektionsteam und Schulleitung oder Lehrkräften, aufgetreten sind.

Der Fragebogen zur Nachinspektion enthält Items zu wahrgenommenen Unterschieden zwischen den Teams der Erst- und der Nachinspektion: Die entsprechende Frage nach Differenzen bejahen 68 % der Schulleiter. 76 % der Zustimmungen werden in einem freien Antwortfeld erläutert. Außer in einer dieser Äußerungen wird das Erstinspektionsteam immer kritischer beurteilt als die Inspektoren der Nachinspektion. Die Kritik am Erstinspektionsteam betrifft vereinzelt konkrete Inspektionsaktivitäten: Rahmenbedingungen seien nicht ausreichend berücksichtigt, Bereiche der schulischen Arbeit nicht zur Kenntnis genommen oder gewürdigt worden, die Rückmeldungspräsentation sei demotivierend gewesen. Überwiegend wird in den Rückmeldungen das Auftreten der Inspektoren drastisch kritisiert: „unfreundlich“, „arrogant“, „aggressiv“, „reserviert“, „nicht wertschätzend“, „voreingenommen“, „inkompetent“. Positiv werden demgegenüber die Nachinspektionsteams charakterisiert.

Die Berechtigung der Kritik an den verschiedenen Erstinspektionsteams ist nicht überprüfbar. Die Teams werden den Schulen zufällig zugewiesen und die Nachinspektionen umfassen nur 7 % aller inspezierten Schulen, d.h. alle Teams haben anderen Schulen hinreichende Qualität bescheinigt und von dort keine vergleichbare Kritik erfahren. Der Inspektion ist ein Prozess, in dem die beiden Akteursseiten interagieren (vgl. Lambrecht u.a., 2011). Möglicherweise erhöht sich die Distanz die zwischen Inspektionsteam und Kollegium, wenn sich im Verlauf des Schulbesuchs die schulischen Schwächen zunehmend herauskristallisieren und die Inspektoren besonderen Wert darauf legen (müssen) ihr negatives Urteil gut abzusichern.

Erkennbar sind unterschiedliche Gewichtungen, die Schule und Inspektion einzelnen Qualitätsaspekten zuweisen, Ursache der Unzufriedenheit. Rürup (2008) unterscheidet als zwei für Einschätzungsabweichungen relevante Facetten die „Wahrheits-“ bzw. die „Angemessenheitsorientierung“, die mit den von Allen/Burgess (2012) aufgezeigten Polen von Inspektionsansätzen korrespondieren. Das betrachtete Inspektionsverfahren ist eher wahrheitsorientiert angelegt: es umfasst eine Definition von Schulqualität, ein differenziertes Qualitätskriterienprofil mit Teilkriterien und zugeordneten beobachtbaren Indikatoren sowie eine intensive Schulung der Inspektoren mit der Maßgabe vergleichbarer Umsetzung der Vorgaben. Die Anwendung des Kriterienmodells kann aber durch schulische Rahmenbedingungen erschwert sein (z. B. fehlende Führung durch vakante Schulleiterstelle) oder aus Schulsicht besonders intensiv verfolgte Entwicklungsbemühungen (z. B. Umstellung vom Halbtags- auf den Ganztags-schulbetrieb), die im Kriterienprofil nicht abgebildet sind. Dann vermissen

Schulen die Angemessenheit des Inspektionsrahmens für ihre Situation. Auch Inspektoren stehen vor dem Dilemma, die Erfüllung vorliegender Kriterien „abzuhaken“ oder als professionelle Akteure zu handeln und die beobachteten Besonderheiten zu berücksichtigen. Möglicherweise stellen die Nachinspektionen auch für die Inspektionsteams eine besondere Situation dar, die in geringerem Maße unter dem Anspruch der Vergleichbarkeit aller Inspektionen („Wahrheitsorientierung“) durchgeführt wird. Das wird zu prüfen sein.

4. Zwillinge

Die Bedeutung, die die Inspektionsbewertung „gravierende Mängel“ im Vergleich zu Schulen ohne dieses high-stakes-Urteil hat, lässt sich über einen Vergleich von Schulen beider Gruppen, die ähnliche bewertet werden, genauer erfassen. Dass sich „Zwillinge“ mit vergleichbaren mittleren Qualitätsbewertungen, aber einem Unterschied bzgl. des Urteils über gravierende Mängel bilden lassen sollten, liegt an der oben beschriebenen Normensetzung bei der Zusammenfassung und Gewichtung der Teilkriterienbewertungen.

Zur Ermittlung der Zwillingspaare wird für alle Erstinspektionen über die 15 Qualitätskriterienmittelwerte⁶ eine Faktorenanalyse errechnet. Orientiert am Eigenwertkriterium „>1“ ergeben sich zwei Faktoren, die die vier Unterrichtskriterien von den anderen elf Kriterien trennen. Das Ergebnis erscheint plausibel, weil

- die Beziehung zwischen Merkmalen des Unterrichts und schulischen Rahmenbedingungen überwiegend distal ist
- die Verfahren der Datenerfassung und -analyse während der Inspektion sich erheblich unterscheiden.

Der Unterrichtsbewertung auf Schulebene liegt die Aggregation der von den Mitgliedern des Inspektionsteams einzeln vorgenommenen Bewertungen der Unterrichtseinsichtnahmen (jeweils 20min der in der Regel 45min-Stunden) zugrunde, während die anderen Urteile auf Konsensentscheidungen bei der Triangulation verschiedener Datenquellen (Dokumente, Interviews, Wahrnehmungen während des Schulbesuchs) beruhen.

Hinsichtlich der Analyse der Schulen mit gravierenden Mängeln erscheint die Lösung mit zwei Faktoren außerdem wegen der zwei ineinandergreifenden „Nachinspektions“-Kriterien (s. o.) angemessen.

Ein schwaches Inspektionsergebnis löst Reaktionen aus. Ob diese unterschiedlich sind, wenn das Scheitern eher auf der Unterrichtsbewertung basiert, oder auf der Bewertung der Kriterien, die eher in der Verantwortung des Schulleiters liegen, wäre von Interesse. Dem Aspekt wird hier aber nicht nachgegangen.

Die folgende Grafik zeigt die Position aller Schulen im zweidimensionalen Feld.

⁶ nicht normbasiert; Mittelwert der zugehörigen Teilkriterienurteile

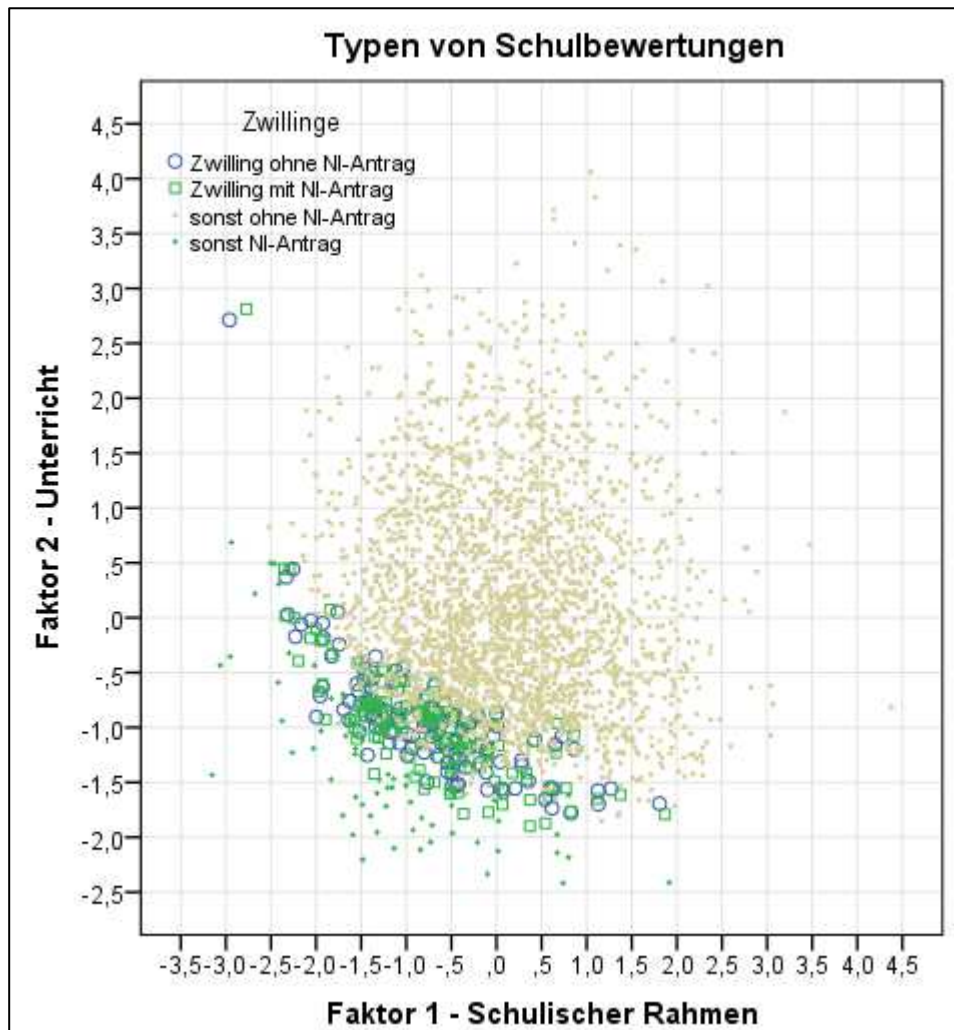


Abb. 2: Position aller inspizierten Schulen in einer zweidimensionalen Darstellung

Zwischen sehr schwachen und deutlich besseren Schulen gibt es einen Überlappungsbereich, aus dem für eine Paarbildung Schulen gleicher mittlerer Bewertung der Teilkriterien, aber mit bzw. ohne das Urteil „gravierende Mängel“ ausgewählt werden können.

Da nur Schulen einbezogen werden sollen, aus denen Rückmeldungen der Schulleiterbefragung vorliegen, kann die Zwillingsbildung nicht zu eng gefasst werden um auf eine für weitere Analysen ausreichend große Fallzahl zu kommen. Daher konnte die Zwillingsbildung neben den beiden Faktorladungen keine weiteren Merkmale in die Auswahl einbeziehen. Gewisse Unterschied in den Faktorladungen (hier euklidische Distanz < 0,35 pragmatisch gewählt) sollten allerdings nicht überschritten werden.

Wenn mehrere Schulen mit zumindest ausreichender Qualität hinsichtlich ihrer Faktorenladungen zu einer Nachinspektionsschule „passten“, wurde bevorzugt die gleiche Schulform ausgewählt. Ansonsten wurde die Schulform für die Zwillingsbildung nicht berücksichtigt. In der Gruppe der Schulen ohne Nachinspektionsantrag ist der Anteil von Schulen des Sekundarbereichs größer, in der anderen der des Primarbereichs. Schulen des Primarbereichs werden insgesamt im Mittel besser bewertet als die des Sekundarbereichs; das gilt auch für die entsprechenden Gruppen mit „gravierenden Mängeln“.

Bei insgesamt 203 Schulen mit gravierenden Mängeln ließen sich insgesamt 130 Zwillingspaare bilden. Damit weisen ca. 65 % der Schulen mit Nachinspektionsantrag auf Teilkriterienebene ein mittleres Qualitätsniveau aus, das in anderen Schulen als hinreichend angesehen wurde.

Ein Vergleich der mittleren Faktorenladungen, die der Zwillingsbildung zugrunde liegen, zeigt keine signifikanten Unterschiede (s. Tabelle 2). Die negativen Mittelwerte der Faktorenladungen drücken aus, dass es sich um schwächer bewertete Schulen handelt.

Tab. 2: Vergleich der Zwillingsgruppen in den der Gruppenbildung zugrunde liegenden Faktoren

Faktoren	Zwillinge	N	MW	s	se	p (t)	s
Schulischer Rahmen	ohne NI-Antrag	130	-,76	,84	,07	0,98	ns
	mit NI-Antrag	130	-,76	,83	,07		
Faktor 2 Unterricht	ohne NI-Antrag	130	-,93	,54	,05	0,61	ns
	mit NI-Antrag	130	-,96	,56	,05		

Die arithmetischen Mittel über die ca. 95 Teilkriterienbewertungen unterscheiden sich nicht signifikant. Die Differenz der Mittelwerte der 15 Qualitätskriterienurteile fällt dagegen erwartungsgemäß signifikant zugunsten der in ihrer Qualität als hinreichend bewerteten Schulen aus den Zwillingsgruppen aus.

Die Anzahl signifikanter Unterschiede in den einzelnen 15 Qualitätskriterien ist mit je sieben Differenzen für die normbasierten Qualitätskriterienurteile und die arithmetisch errechneten Mittelwerte gleich. Allerdings sind die jeweils drei signifikanten Unterschiede, die auf die vier Unterrichtskriterien QK 3 bis QK 6 entfallen, mit größerer Vorsicht zu interpretieren, da die reliable Erfassung der Unterrichtsqualität durch externe Beobachter ein großes Problem darstellt (vgl. Hill, 2012; Newton, 2010; Praetorius, 2014). Während alle sieben Differenzen über die 15 normbasierten Urteile für die Schulen mit Nachinspektionsantrag signifikant geringer ausfallen, weisen drei der sieben signifikanten Unterschiede für die arithmetisch berechneten Mittelwerte der entsprechenden Teilkriterien in die unerwartete Richtung – die Gruppe ausgewählter Schulen mit Nachinspektionsantrag wird sogar besser bewertet als die in der Inspektion erfolgreiche Zwillingsgruppe.

Insgesamt sehen wir die „Zwillingsbildung“ auf der Ebene der beiden Schulgruppen als gelungen an. Die Vergleiche dieser Gruppen in weiteren Merkmalen lässt erwarten, dass Aussagen über Unterschiede im Folgehandeln weniger auf die detaillierte Schulqualitätseinschätzung durch die Inspektion („Wissensgewinnung“) als auf das Urteil „gravierende Mängel“ („Rechenschaftslegung“) zurückzuführen sein wird.

Die beiden letzten Spalten der Tabelle 1 stellen für die fünf Befragungsskalen mit befriedigender Skalenqualität die Ergebnisse von Mittelwertvergleichen der Zwillingsgruppen dar.

Die Zustimmungswerte der Zwillinggruppen weisen in den Skalen der Nachbefragung hohe Übereinstimmung sowohl mit den Erwartungen als auch mit den oben über Korrelationen berichteten Vergleichen für alle inspizierten Schulen mit und ohne Antrag auf Nachinspektion auf. Schulen ohne Antrag auf Nachinspektion weisen auf der proximal an schulischen Arbeitsfeldern angesiedelten Teilkriterienebene bei nahezu gleicher relativ schwacher Schulqualitätsbewertung höhere Zustimmungswerte in drei der fünf Skalen auf als Schulen der Zwillinggruppe, die bei vergleichbarem Qualitätsniveau ein Nachinspektionsurteil erhielten. Nur in den beiden Wirkungsskalen weisen die Zustimmungsdifferenzen in die andere Richtung: Schulen mit Nachinspektionsankündigung berichten eine höhere Bereitschaft und Aktivität. Drei der Vergleiche sind in erwarteter Richtung signifikant. Zwei Mittelwertpaare unterscheiden sich richtungsmäßig in Übereinstimmung mit der Erwartung, die Differenzen sind aber klein und lassen sich statistisch nicht sichern.

Die Ergebnisse stützen die These, dass die aufgrund der festgestellten Mängel in der Schulqualität beantragte Nachinspektion die Bewertung des Inspektionsverfahrens sowie die Bereitschaft zu Folgemaßnahmen in besonderem Maße beeinflusst. Die Zufriedenheit mit dem Inspektionsverfahren und der Berichtsqualität fällt signifikant geringer aus, der Zustimmungswert zu bevorstehenden oder schon ergriffenen Folgemaßnahmen ist höher.

5. Die Veränderung der Schulqualität von Schulen mit „gravierenden Mängeln“

Außer den Befragungsergebnissen zu unmittelbar nach der Erstinspektion beobachtbaren Einstellungen und Planungen liegen inzwischen auch die Ergebnisse der Nachinspektionen für den Großteil der Schule mit gravierenden Mängeln vor. Die folgende Untersuchung beschäftigt sich mit der Veränderung der Qualitätsbewertung schwacher Schulen von der Erst- zur ca. 1½ Jahre später durchgeführten Nachinspektion. In den Veränderungen drückt sich die Inspektionswirkung möglicherweise konkreter aus als in den Fragebogendaten.

Der Nachinspektion liegt das gleiche Verfahren zugrunde wie der Erstinspektion. Allerdings wird nur der Unterricht immer bewertet. Alle anderen Kriterien werden nur dann erneut in Augenschein genommen, wenn sie vorher nicht der Norm entsprachen. Der Aufwand einer Inspektion ist dadurch für das Inspektionsteam und auch die Schule geringer, die Vergleichbarkeit der Schulqualität zu den Zeitpunkten der Erst- und der Nachinspektion aber eingeschränkt, weil in vorher als hinreichend oder gut bewerteten Bereichen möglicherweise Verbesserungen oder Verschlechterungen eingetreten sind, die nicht erfasst werden.

Abbildung 3 zeigt die Verbesserung der Schulen, die bereits nachinspiziert wurden. Grundlage der Grafik sind die Anteile normgerechter Bewertungen in den Qualitätskriterien. Zwischen den jeweils zwei Bewertungsstufen („stark“ und „eher stark als schwach“ bzw. „eher schwach als stark“ und „schwach“) wird in der Abbildung nicht unterschieden.

Über Schulen, die die Erstinspektion erfolgreich durchlaufen haben, liegen keine Folgedaten vor. Zwar hat die zweite Runde der Schulinspektion in Niedersachsen vor einigen Monaten begonnen. Da das Inspektionsverfahren stark modifiziert wurde, wird ein Vergleich mit den Ergebnissen der Erstinspektion nur beschränkt möglich sein.

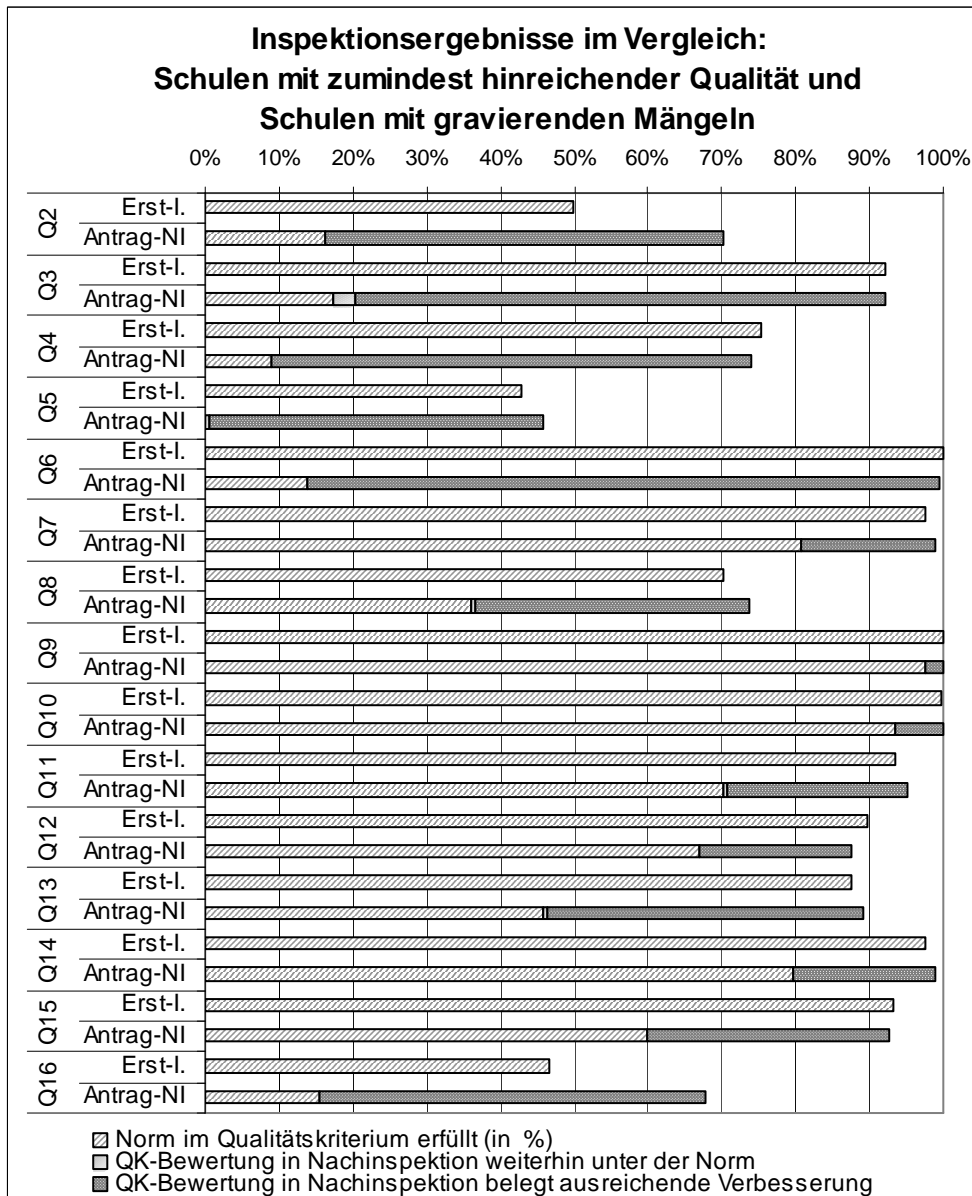


Abb. 3: Vergleich der Qualitätsbewertungen in Erst- und Nachinspektion⁷

⁷

QK1	Ergebnisse und Erfolge (nicht bewertet)	QK2	Curriculum
QK3	Unterrichtsstrukturierung	QK4	Unterrichtsdifferenzierung
QK5	Unterricht-Selbständigkeitsförderung	QK6	Unterrichtsklima
QK7	Leistungsförderung	QK8	Lerndiagnose und -unterstützung
QK9	Schülerberatung	QK10	Schulklima
QK11	Demokr. Mitwirkung von Eltern und SuS	QK12	Kooperationen
QK13	Schulleitungshandeln	QK14	Schulverwaltung
QK15	Personalentwicklung	QK16	Qualitätssicherung und -entwicklung

Für jedes Qualitätskriterium zeigt der obere der beiden Balken an, wie viele Schulen, deren Schulqualität als zumindest hinreichend bewertet wurde, in der Erstinspektion in dem betreffenden Qualitätskriterium den Standard erreichten („+“) bzw. überschritten („++“). Diese Schulen weisen zu 100 % ein gutes Unterrichts- (QK 6) und Schulklima (QK 10) auf; die Schülerberatung z. B. hinsichtlich der weiteren schulischen oder beruflichen Perspektiven (QK 9) erfüllt in allen Schulen den Anspruch. In über 90 % der Schulen verläuft der Unterricht strukturiert (classroom-management), Leistungsbewertungen sind im Kollegium abgestimmt, Schüler und Eltern darüber informiert (QK 7), Eltern und Schüler sind in schulische Entscheidungsprozesse eingebunden (QK 11), die Schulen kooperieren mit Institutionen, aus denen Schüler kommen (z. B. Kindergärten) oder in die sie überwechseln (weiterführende Schulen) sowie außerschulischen Partnern wie z. B. Betrieben (QK 12). Von den Leitungsaufgaben werden die Anforderungen an Schulorganisation (QK 14) und die Personalentwicklung (QK 15) in über 90 % aller in der Erstinspektion insgesamt positiv beurteilter Schulen erfüllt. Deutlich sind aber auch die Schwächen erkennbar. Defizitbereiche, die sich direkt auf das Lernen der Schüler beziehen, sind die schulischen Arbeitspläne (QK 2), die oft Jahre später noch nicht an die ab 2004 eingeführten Bildungsstandards angepasst sind, differenzierende (QK 4) und die Selbständigkeit der Schüler fördernde Unterrichtsformen (QK 5) werden zu wenig beobachtet und auf systematischer Diagnose basierende Förderpläne (QK 8) liegen in 30 % der Schulen nicht vor. Mehr als 50 % der Schulen haben die Aufgabe einer kontinuierlichen Qualitätssicherung und -entwicklung (QK 16) noch nicht für sich umgesetzt. Bedenklich ist auch, dass in ca. 12 % der Schulen mit insgesamt hinreichender Schulqualität die Schulleitung ihrer Führungsaufgabe nicht angemessen gerecht wird, in den 7 % Schulen mit gravierenden Mängeln sind es mehr als die Hälfte.

Der jeweils untere Balken der Schulen mit gravierenden Mängeln, die schon nachinspiziert wurden, zeigt jeweils links den Anteil der Schulen, die das Kriterium bereits in der Erstinspektion erfüllten. Auf den Erstinspektionsergebnissen setzt rechts der Zuwachs an normgerechten Urteilen in der Nachinspektion auf.

Außer im Unterricht (QK 3 - QK 6) wurden die Kriterien, die der Norm bereits in der Erstinspektion entsprachen, in der Nachinspektion nicht erneut überprüft, so dass weder eine Besser- noch eine Schlechterbewertung möglich ist. Der hellgraue Abschnitt in QK 3 betrifft Schulen, die sich in der Nachinspektion von normgerecht auf unter Norm verschlechtert haben.

Abbildung 3 belegt deutliche Verbesserungen der nachinspizierten Schulen, die hinsichtlich der Erreichung der Normen anteilmäßig nicht mehr hinter den schon in der Erstinspektion erfolgreichen Schulen zurückbleiben. Mittelwertvergleiche zeigen allerdings, dass die nachinspizierten Schulen nicht in jedem Kriterium das Niveau der schon in der Erstinspektion erfolgreichen Schulen erreichen. Das ist auf zwei Gründe zurückzuführen:

- Da außer im Unterricht ein Merkmal nur überprüft wird, wenn die Norm unterschritten wurde, können sich in „schweren“ Kriterien mehr Schulen verbessern als in „leichten“. In den Kriterien QK 2 und QK 16 überschreiten die Anteile sogar die der erfolgreichen Erstinspektionen deutlich (ca. + 20 %). In diesen beiden Kriterien greift die „Normdurchsetzung“ am stärksten. Nachweise beruhen weniger auf der Beobachtung der Inspektoren oder Interviewaussagen von Schülern und Eltern, sondern auf Dokumenten, die in schulischen Gremien erstellt werden. Schulen, die durch die angekündigte Nachinspektion zu Entwicklungen gezwun-

gen sind, gestalten z. B. ihre schulinternen Arbeitspläne gemäß den gesetzlichen Anforderungen (QK 2), entwickeln ein Leitbild, werten die Vergleichsarbeits-ergebnisse aus und führen interne Evaluationen durch (QK 16).

- Meistens erreicht der Anteil der in der Nachinspektion normgerecht bewerteten Schulen den der Erstinspektion (Abbildung 3). Mittelwertvergleiche, in die die vier Bewertungsstufen mit 1 bis 4 eingehen, zeigen, dass Schulen sich in der Nachinspektion deutlich verbessern, aber die Ergebnisse der „erfolgreichen“ Schulen nicht erreichen, weil der Anteil besonders positiver Urteile („stark“ – 4) auch in der Nachinspektion deutlich hinter dem entsprechenden Anteil bei den erfolgreichen Erstinspektionen zurückbleibt.
Im QK 5 mit der schwächsten Inspektionsbewertung erreichen 2,9 % der erfolgreichen Schulen das beste Urteil; in der Nachinspektion steigt der Anteil der 4er-Bewertungen in der Gruppe der in Erstinspektion als schwach bewerteten Schulen von 0 % auf 0,6 %; im einem der am besten bewerteten Kriterien QK 10 betragen die Werte in der Erstinspektion 52,2 % bzw. 10,2 % und in der Nachinspektion 10,8 %.
Möglicherweise handelt es sich um ein durch das Verfahren bedingtes Artefakt. Schulen mit gravierenden Mängeln wurden in der Erstinspektion selten in einzelnen Kriterien sehr positiv beurteilt (evtl. Halo-Effekt) und in der Nachinspektion erscheint eine Verbesserung der Bewertung um zwei Stufen nicht gerechtfertigt oder wird möglicherweise bewusst vermieden. Evtl. lässt sich daraus aber auch der Hinweis ablesen, dass es Grenzen der in eineinhalb Jahren möglichen innerschulischen Entwicklung gibt.

Positiv bleibt festzuhalten, dass ehemals schwache Schulen in ihrer Qualität innerhalb von eineinhalb bis zwei Jahren zu denen aufschließen, die in der Erstinspektion eine hinreichende Qualität aufweisen. Die Profile der erfolgreichen und der nachinspizierten Schulen verlaufen allerdings parallel, d.h. schwach bewertete Qualitätsbereiche bleiben auch in Schulen, die eine längere Zeit intensiv auf die Nachinspektion hin gearbeitet haben, defizitär. Wenn man eine 80 % Normerfüllung als Maßstab ansieht, so bleibt die Schulqualität in den Qualitätskriterien 2, 4, 5, 8 und 16 weiter dahinter zurück. Nur Feldstudien könnten darüber Aufschluss geben, warum eine grundlegende Verbesserung vielen Schulen in diesen Kriterien nicht gelingt. Dabei wäre auch zu prüfen, ob die Ansprüche des Inspektionsrahmens angemessen angesetzt sind. Neuere Studien wecken durchaus Zweifel an den von der Inspektion in der ersten Runde gesetzten Standards der Unterrichtsgestaltung und Schulentwicklung (vgl. Hattie, 2009 bzw. Schlee, 2014).

Die Veränderung der Inspektionsergebnisse von der Erst- zur Nachinspektion hat mehrere Ursachen in Erwägung zu ziehen.

- Die Schulaufsicht hat ausgehend vom Inspektionsbericht mit der Schulleitung Vereinbarungen über die Entwicklungsziele und -maßnahmen getroffen. Das dazu eingerichtete Unterstützungssystem ist orientiert am Schwächenprofil durch Schulentwicklungsberatung oder Projekte zur Verbesserung der Unterrichtsqualität in den Schulen tätig geworden (vgl. Sommer, Stöhr, Thomas, 2010). Schulleitung und Kollegium haben u. U. mit Schülern und Eltern in zahlreichen Konferenzen und in Projektarbeit vorhandene Defizite erfolgreich bearbeitet. Die Schulqualität ist tatsächlich besser geworden.
- Das bessere Ergebnis ist auf die Konzentration der schulischen Entwicklungsarbeit auf besonders bewertungsrelevante Indikatoren oder lediglich

erfolgreiches window-dressing zurückzuführen (vgl. Janssens, 2013, S. 72f). Die Schulqualität hat sich nicht substantiell verändert.

- Die Beurteilungen von Schulen mit gravierenden Mängeln liegen am unteren Ende der Skala. Bei einer zweiten Messung werden ursprüngliche Extremwerte einer Regression zur Mitte unterliegen. Die bessere Bewertung beruht auf einem rein statistischen Effekt.

Als weitere Ursache besserer Nachinspektionsbewertungen, der in der Literatur zur Verbesserung schwacher Schulen kaum benannt wird, kommt auch ein „milderes“ Urteil der Nachinspektionsteams (Zusammensetzung grundsätzlich abweichend von der der Erstinspektion) in Frage. Ein neuer Schulleiter im Amt wird möglicherweise einen „Bonus“ erhalten; ein begonnener Entwicklungsprozess, dessen Stand in einer Erstinspektion nicht gewürdigt würde; erscheint in der Nachinspektion mit Blick auf die kurze Zeit zwischen den Inspektionen als angemessen. Eine Tendenz zum „im Zweifel für den Angeklagten“ (auf den das Fehlen von „++“-Bewertungen hinweisen könnte) ließe sich auch darauf zurückführen, dass es in der ersten Runde der externen Evaluation nur eine vage Regelung für die weitere Vorgehensweise bei Schulen gab, deren Qualität auch in der Nachinspektion als nicht hinreichend bewertet wurde und dass die Gestaltung einer zweiten Inspektionsrunden lange offen war, so dass unklar war, ob und wann eine erneut negativ beurteilte Schule in einer weiteren Inspektion würde zeigen können, dass sie ihre Defizite beseitigt hat.

Analysen der 20 Items der Unterrichtsbeobachtung liefern belastbare Hinweise darauf, dass die Inspektionsteams Schulen in den Nachinspektionen wohlwollend bewerten. Sie sind bzgl. der Frage nach Unterschieden im Inspektionsverhalten besonders aussagekräftig, weil

- Unterricht in allen Nachinspektionen erneut in Augenschein genommen wird, so dass aus allen Schulen eine Neubewertung vorliegt
- in der überwiegenden Mehrzahl der Schulen mit gravierenden Mängeln die Qualitätskriterien 3, 4 und 5 des Unterrichts als nicht hinreichend bewertet wurden und damit die zweite gesetzte Nachinspektionsnorm unterschritten wird
- viele Schulen mit Hilfe externer Unterstützung Entwicklungsprojekte zur Verbesserung der Unterrichtsqualität verfolgen; diese sind allerdings langwierig, u.a. weil eingeschliffene Routinen und stabile subjektive Theorien der Lehrkräfte betroffen sind (vgl. Groeben, Wahl, Schlee, Scheele, 1984); in 65 % aller Rückmeldungen ist angegeben, dass zum Zeitpunkt der Nachinspektion die Schulentwicklungsaktivitäten, darunter häufig die Verbesserung der Unterrichtsqualität, noch nicht abgeschlossen sind; internationale Studien deuten zudem auf eine eingeschränkte Bereitschaft von Lehrkräften hin, den eigenen Unterricht aufgrund von Inspektionsergebnissen zu verändern (vgl. Husfeldt, 2011, S. 16f).

Die Urteile in Unterrichtseinsichtnahmen unterliegen einem hohen Beurteilungsfehler (vgl. Hill, 2012; Newton, 2010; Praetorius, 2014); dieser sollte zufällig sein und auch in negativer Richtung ausschlagen; in den Nachinspektionen erfüllen aber nur noch 3 % der Schulen die Unterrichtsnorm wiederum nicht.

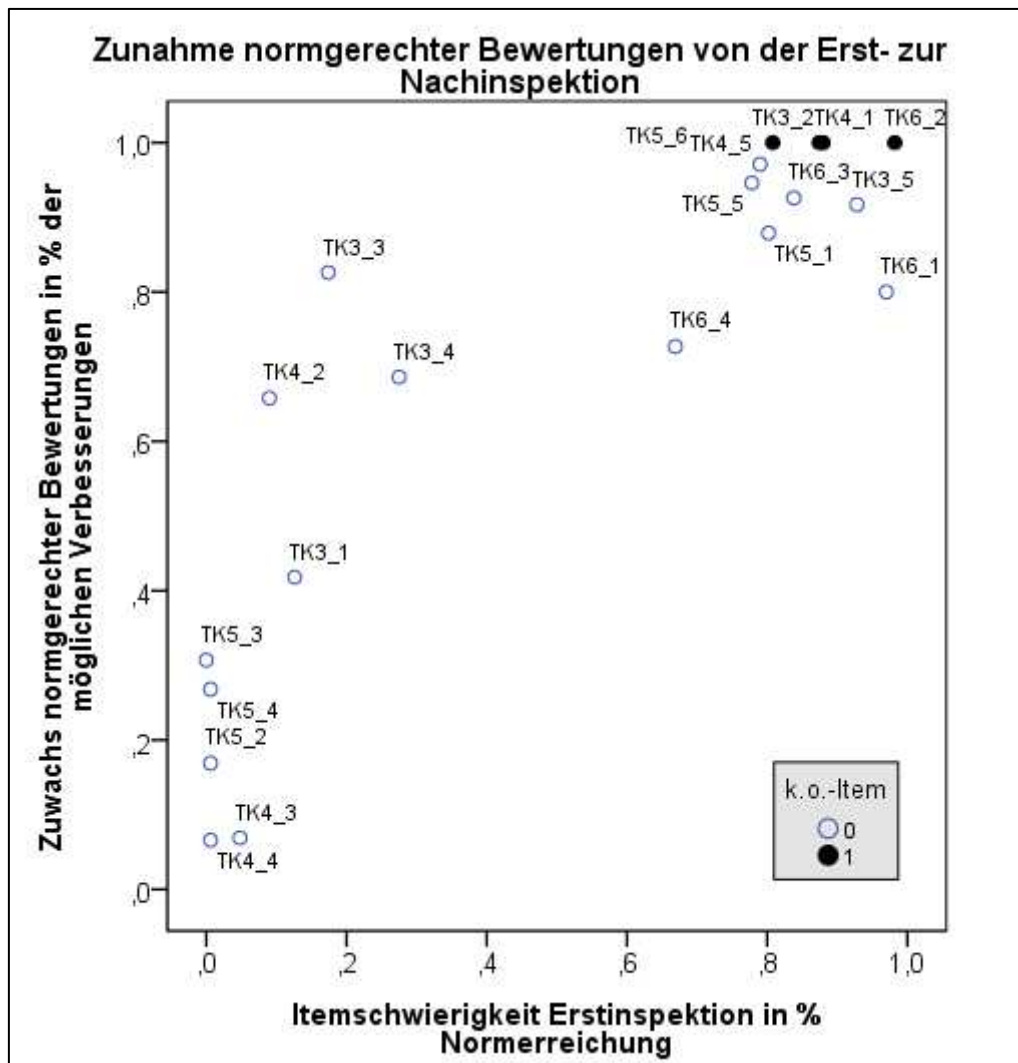


Abb. 4: Die Verbesserung der Schulen mit gravierenden Mängeln im Unterricht

Abbildung 4 zeigt auf der Abszisse für die 20 Teilkriterien des Unterrichtsbeobachtungsbogens den Anteil der Schulen mit Nachinspektionsantrag und bereits durchlaufener zweiter Inspektion, die bereits in der Erstinspektion den gesetzten Anspruch zumindest hinreichend erfüllten, und auf der Ordinate den Anteil an ehemals schwach bewerteten Schulen, die in der Nachinspektion besser bewertet wurden.

Unten links befinden sich Items, die in dieser Auswahl schwacher Schulen in der Erstinspektion nur in wenigen Fällen positiv bewertet wurden (z. B. TK3_1 Zielklarheit, TK4_3 Anforderungsspektrum, TK4_4 Differenzierung, TK5_2 Förderung von Selbständigkeit, TK5_3 Kooperation und TK5_4 Medienkompetenz) und in denen der Zuwachs an normgerechten Bewertungen (Ordinate) in der Nachinspektion unterdurchschnittlich ausfällt, was auf begrenzte Erfolge der Entwicklungsarbeit hindeutet.

Die rechts oben platzierten Unterrichtsmerkmale wurden bereits in der Erstinspektion überwiegend gut umgesetzt, z. B. alle Kriterien des Unterrichtsklimas TK6_..., und in der Nachinspektion kommt ein hoher Anteil der wenigen vorher als unzureichend bewerteter Schulen zu einem hinreichenden Ergebnis.

In diesem Spektrum gibt es Indizien für eine positive Verzerrung des Inspektionsurteils mit der Intention, dass eine Schule in der Nachinspektion die Norm für Unterricht möglichst nicht wieder unterschreiten sollte.

- Die sogenannten k.o.-Kriterien, die unbedingt erfüllt werden müssen um eine hinreichende Bewertung des gesamten Qualitätskriteriums zu erhalten, wurden bereits in der Erstinspektion besonders gut bewertet. Das ist damit zu erklären, dass als Ausschlusskriterien Merkmale gewählt wurden, die als Basis von Standardunterricht gelten und deren Umsetzung daher von allen Lehrkräften zu erwarten ist. Bemerkenswert ist dennoch, dass in der Nachinspektion alle vier k.o.-Kriterien TK3_2, TK4_1, TK5_6 und TK6_2, und nur diese, in allen darin vorher schwach bewerteten Schulen jetzt als erfüllt angesehen werden.
- Der besonders hohe Zuwachs der vorher schwach bewerteten Kriterien TK3_3, TK3_4 und TK4_2 passt in das Bild wohlwollender Inspektionsbewertung. Eine Schule erfüllt die Unterrichtsnorm, wenn drei der vier Qualitätskriterien normgerecht bewertet werden. Aus der Abbildung 4 lässt sich die Rangreihe der Erfüllung des Anspruchs in den vier Qualitätskriterien erschließen: QK 6 Unterrichtsklima (überwiegend erfüllt) > QK 3 Unterrichtsstrukturierung > QK 4 Differenzierung > QK 5 Aktives selbständiges Lernen (sehr schwach ausgeprägt). QK 6 ist leicht zu erfüllen und daher nicht entscheidend, QK 5 ist anspruchsvoll, aber das Unterschreiten der Norm in diesem Kriterium allein hat keine Konsequenzen; die Grafik zeigt, dass der Zuwachs in den drei Teilkriterien TK5_2, TK5_3 und TK5_4 gering ist; es hat sich an der Unterrichtsqualität im Bereich der Selbstständigkeitsförderung wenig geändert. Anders verhält es sich in den Teilkriterien der Qualitätskriterien 3 und 4, die für das Urteil gravierende Mängel besonders bedeutsam sind. Von jeweils 5 Teilkriterien müssen drei hinreichend erfüllt sein. In QK 4 muss neben den schon in der Erstinspektion als leicht erscheinenden Items TK4_1 und TK 4_5 ein weiteres Item die Norm jetzt erfüllen. Das ist mit TK4_2 gegeben, wo von den zuvor 90 % Schulen mit unzureichenden Bewertung 70 % in der Nachinspektion die Standards hinreichend erfüllen. Vergleichbar ist das Bewertungsverhalten in QK 3. Die Anforderungen in den TK3_2 und TK3_5 sind leicht erfüllbar, die Zuwächse in den für die Erfüllung der Nachinspektionsnorm besonders wichtigen Items TK3_3 und TK3_4 sind unter den schweren Items überdurchschnittlich.

Mit einem Unterrichtsbeobachtungsbogen wird das ganzheitliche Bild von Unterricht analytisch gegliedert. Dass nach einer Phase der Unterrichtsentwicklung einige Merkmale, und dabei genau die für das Inspektionsurteil relevanten, sich stark positiv verändern, andere wesentliche Qualitätsmerkmale aber kaum, ist angesichts des Handlungszusammenhangs der Merkmale in einer Unterrichtssequenz und in Unterrichtsentwicklungsmaßnahmen wenig plausibel.

Mit den vorliegenden Daten kann die Annahme einer Senkung der Ansprüche in der Nachinspektion gegenüber der Erstinspektion nicht belegt werden. Sich auf das Ernten von low-hanging-fruits bzw. im Sinne der „indikatorengestützten Praxisgestaltung“ (Landwehr, 2011, S. 62) auf das Erreichen der für die Normvorgaben relevanten Teilkriterien zu konzentrieren, mag als unerwünschter Nebeneffekt von externer Evaluation gelten (vgl. Wolf/Janssens, 2007, S. 382), wäre aber strategisch klug. Wie beim teaching-to-the-test kann es sich um eine sinnvolle und notwendige Orientierung an den durch Vorgaben charakterisierten Anforderungen handeln, aber auch eine unerwünschte Einengung von Interventionen und Bewältigungsstrategien auf die begrenzte Komplexität von Prüfungssituation.

Hinweise auf ein gezieltes Unterlaufen der Inspektionsansprüche durch Konzentration auf oberflächliche Aspekte von Schulqualität, Schmeicheln der Inspektoren oder Verunsichern durch in Frage Stellen der Qualität des Inspektionsverfahrens sind nicht auszuschließen (vgl. The Telegraph, 2012), aber in der Schulleiterbefragung nicht erkennbar. Wenn Inspektionsverfahren in Richtung eines kommunikativen Austauschs von schuleigener und externer Wahrnehmung der aktuellen Schulsituation und konsensorientierte Beratung hin verändert werden, wächst die Gefahr des gaming (vgl. Bruijn u.a., 2007). In entsprechenden Evaluationsverfahren sollte dem Aspekt daher Aufmerksamkeit gewidmet werden.

Aufgrund mehrerer möglicher Ursachen ist der Umfang tatsächlicher Veränderungen zwischen Erst- und Nachinspektion und deren Nachhaltigkeit schwer einzuschätzen. Es bleibt eine gewisse Skepsis. Die Tendenz, positive Veränderungen zu überhöhen, ist nicht auf Verfahren beschränkt, die schulische Prozesse und Strukturen in Augenschein nehmen. Auch Inspektionen, die sich stark auf Schülerleistungsdaten stützen, weisen möglicherweise die Tendenz auf, ihre eigene Bedeutsamkeit durch Hinweise auf vermeintliche Verbesserungen von Schulen zu belegen, wie Quesel anhand von Entwicklungen in England aufzeigt (vgl. Quesel, 2011).

6. Befragungen zur Nachinspektion

Wie nach der Erstinspektion werden Schulleitung und schulische Gruppen auch im Anschluss an die Nachinspektion zum Inspektionsablauf befragt. Ab 2009 wurde dazu ein modifizierter Schulleiterfragebogen eingesetzt, in dem Items zur Bewertung des Evaluationsablaufs ergänzt wurden um solche zu

- Reaktionen schulischer Gruppen (insbesondere Kollegium und Eltern) auf das schwache Inspektionsergebnis
- eingeleitete Entwicklungsmaßnahmen
- deren Unterstützung durch die Schulaufsicht
- der Belastung durch die Entwicklungsarbeit.

Die Rücklaufquote konnte auf ca. 80 % erhöht werden, indem die Schulleiter eine gewisse Zeit nach Zustellung der Zugangsinformationen telefonisch auf die Bedeutung der Befragung hingewiesen wurden. Es liegen 75 Datensätze vor.

In eineinhalb bis zwei Jahren können sich schulische Rahmenbedingungen erheblich verändern (vgl. Houtveen u.a., 2007). In 25 % der vorliegenden Rückmeldungen wird angegeben, dass sich die personelle Situation der Schule gegenüber dem Zeitpunkt der Erstinspektion verbessert hat, in jeweils 38 % schätzt der Schulleiter die Lehrmittelausstattung und den Zustand von Schulgebäude und -gelände als besser ein. Etwa 22 % der Schulen haben, dem Trend der demographischen Entwicklung folgend, mehr als 10 % ihrer Schüler eingebüßt, in einem Drittel der Schulen hat die Schulleitung gewechselt.

Die auf den Inspektionsablauf bezogenen Skalen weisen eine gute Reliabilität auf. Die Zustimmung der Schulleitung zum Verfahren liegt mit $MW = 3,76$ noch über den für alle Erstinspektionen ermittelten Werten, die Urteilsvalidität des Inspektionsverfahrens wird ebenfalls positiv, aber niedriger als die Verfahrensumsetzung in der Schule eingeschätzt und trifft nahezu auf den Wert, der in der Befragung zur Erstinspektion ermittelt wurde.

Die Schulleitung schätzt die Bereitschaft des Kollegiums sich mit den (guten) Nachinspektionsergebnissen auseinanderzusetzen sehr positiv ein; etwas geringer war die Zustimmung zu den ähnlichen Items in der Befragung zur Erstinspektion. Beide Werte zeugen von einer „partizipatorischen Problemlösekultur“ im Kollegium

als Voraussetzung der Wirksamkeit des Inspektionswissens (Landwehr, 2011, S. 46f).

Oben wurde darauf hingewiesen, dass in Deutschland Leistungs- und Inspektionsdaten nicht veröffentlicht werden, weil die Unterstützung einer Wettbewerbssituation zwischen Schulen als dysfunktional angesehen wird. Das scheint zurzeit auch öffentlich nicht strittig. Eltern und kommunale Schulträger, scheinen auf die Schule keinen Druck auszuüben, wie geringe Zustimmung der Schulleiter auf entsprechende Items zur Kritik von außen (MW = 2,01 – „stimme eher nicht zu“) belegt.

Eineinhalb bis zwei Jahre nach der Erstinspektion mit dem Urteil gravierende Mängel wird die damalige Bewertung zwar mehrheitlich als zutreffend eingeschätzt, der Mittelwert knapp unter MW=3,00 und eine relativ hohe Streuung weisen aber darauf hin, dass ein erheblicher Teil der Schulleiter auch im Rückblick das negative Erstinspektionsurteil als nicht gerechtfertigt ansieht. Auffällig ist, dass die Zustimmung zur überwiegend negativen Bewertung des Unterrichts bei den Schulleitern höher ist als zur Bewertung der Schule insgesamt, in der der ihnen selbst zuzurechnende Einfluss stärker ist. Den Aussagen über mögliche schulinterne Ursachen, u. a. Versäumnisse innerschulischer Entwicklungsarbeit, insbesondere der Schulleitung (MW=2,71) wird tendenziell eher zugestimmt, die zum mangelnden Engagement des Kollegiums werden eher abgelehnt (MW =2,24).

Hinsichtlich der geleisteten Entwicklungsarbeit beurteilen die Schulleiter die Bereitschaft des Kollegiums moderat positiv (MW = 2,98). Der Wert ist angesichts der durch die Aktivitäten verursachten zusätzlichen Belastung (MW = 3,44) als hoch einzuschätzen. Die Unterstützung, die der Schule von der für deren Bereitstellung verantwortlichen Schulaufsicht gewährt wird, ist hoch und wird in ihrem Effekt positiv bewertet, sie trägt in der Schule aber z.B. durch zusätzliche Termine zur empfundenen Belastung bei.

Weitere Skalen beziehen sich auf Veränderungen in „weichen“ schulischen Strukturen zwischen den beiden Inspektionen. Sie werden alle als moderat bis sehr positiv eingeschätzt – die kollegiale Zusammenarbeit hat zugenommen, Schulleiter haben in höherem Maße Führungsverantwortung übernommen. Dazu ist zu ergänzen, dass Schulleiter in Deutschland an den meisten Schulen eine hohe Unterrichtsverpflichtung haben, bis zur Schulgesetzänderung 2007 nur an Gymnasien, also in weniger als 10 % aller allgemein bildenden Schulen Dienstvorgesetzte waren und sich dementsprechend überwiegend als „primus inter pares“ fühlten und z. T. Probleme mit dem vorzunehmenden Rollenwechsel vom Kollegen zur Führungsperson hatten.

Die in der Schule insgesamt erreichten Veränderungen werden als sehr positiv eingeschätzt.

Zwar wird dem Nachinspektionsergebnis noch ein Erkenntniswert für die weitere schulische Entwicklung zugeschrieben, die Aussagen zu konkreten weiteren Entwicklungsmaßnahmen erfahren aber demgegenüber nur noch moderate Zustimmung. Aus detaillierten Items zu den Bereichen, in denen die Arbeit fortgesetzt wird, sticht die Unterrichtsentwicklung hervor, in der begonnene Projekte noch nicht abgeschlossen sind.

Interessante Ergebnisse weist ein Vergleich des Antwortverhaltens von Schulleitern auf, die bereits während der Erstinspektion die Schule führten, mit denen, die

zwischenzeitlich die Leitung neu übernommen haben. Genau ein Drittel der Rückmeldungen stammen von „neuen“ Schulleitern.

Vergleicht man, wie verbliebene und neue Schulleiter die Enttäuschung des Kollegiums über das Urteil „gravierende Mängel“ und die daraus folgende Belastung der schulischen Arbeit, eine relativ gut bestätigte nicht intendierte Wirkung (Wolf/Janssens, 2007), wahrnehmen, so lassen sich keine signifikanten Unterschiede feststellen. In beiden Gruppen melden 35 % zurück, das empfundene Beeinträchtigungen der schulischen Arbeit sich über mehr als ein halbes Jahr erstreckten (s. Abbildung 5). „... eine negative Diagnose (kann also) selbst eine lähmende Wirkung haben“ (Quesel u.a., 2013a, S. 14), die sich in „Entmutigung, Verringerung der Identifikation, Stigmatisierung der Schule, negative Auswirkungen auf die Personalführung und die Konzentration auf weniger relevante Aspekte der Schulentwicklung“ äußern kann (Husfeldt, 2011, S. 24f).

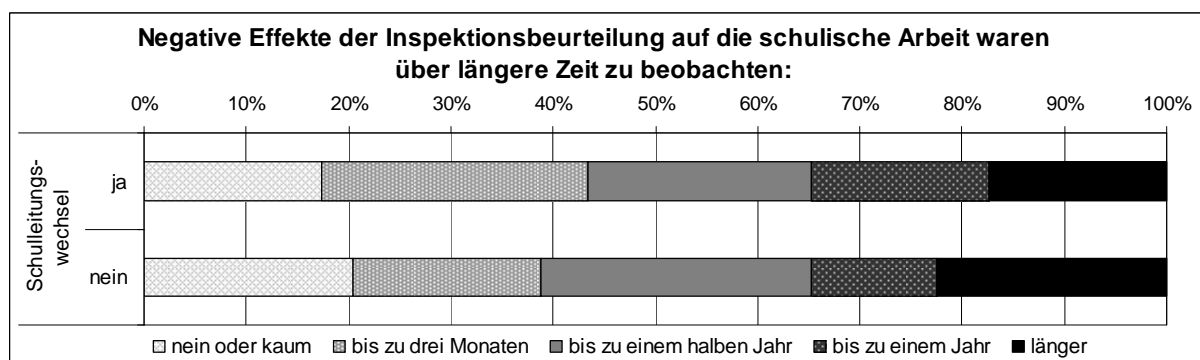


Abb. 5: Belastung der schulischen Arbeit durch das Inspektionsurteil „gravierende Mängel“

Bei anderen Aussagen fallen die Einschätzungen der beiden Schulleitergruppen dagegen z. T. signifikant unterschiedlich aus. Schulleiter, die zum Zeitpunkt der Erstinspektion bereits die Verantwortung für die Schule trugen, schätzen das Erstinspektionsurteil deutlich häufiger als falsch ein als neu eingesetzte Schulleiter. Sie sehen auch wesentlich seltener Schulleitungsversäumnisse als Ursache des schwachen Abschneidens in der Erstinspektion an, wie Abbildung 6 zeigt.

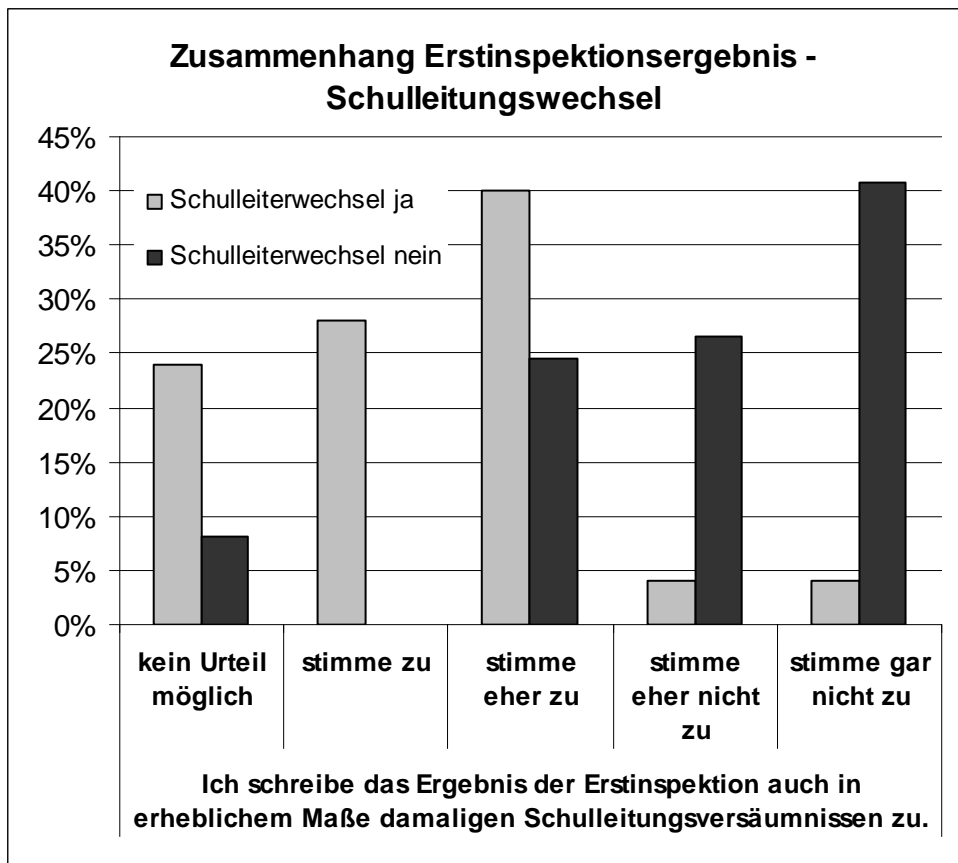


Abb. 6: Einschätzung des Erstinspektionsurteils durch die Schulleiter

Dieser Unterschied lässt sich anhand verschiedener Skalen nachweisen. Schulleiter, die schon in der Erstinspektion verantwortlich waren, stimmen

- der Bewertung des Erstinspektionsurteils
- der Urteilsvalidität des Inspektionsverfahrens
- den vorgegebenen innerschulischen Ursachen für das schwache Ergebnis
- der Veränderung ihres Führungsverhalten und
- dem Nutzen der Nachinspektion aus Sicht des Schulleiters

in signifikant geringerem Maße zu als neu eingesetzte Schulleiter.

Über die Ursachen können nur Vermutungen angestellt werden. Darin ausschließlich eine Verdrängung der eigenen Verantwortung durch die länger im Amt befindlichen Schulleiter zu sehen, wäre zu einseitig. Der Blick neu eingesetzter Schulleiter auf die Schule und das Inspektionsergebnis ist vermutlich unvoreingenommener, der Informationswert des Inspektionsberichts aber für sie auch höher als für diejenigen, die die Schule lange schon geführt haben.

Fazit

Die Untersuchung belegt positive Zusammenhänge zwischen dem Inspektionsurteil über die Schulqualität und der Bewertung des Inspektionsablaufs durch Schulleiter und negative zwischen dem Urteil und geplanten Entwicklungsmaßnahmen. An Zwillinggruppen mit nahezu gleicher Schulqualitätsbewertung zeigt sich der besondere Einfluss der Offizialisierung von Schwächen durch das Urteil „gravierende Mängel“. Die Möglichkeit der Zwillingbildung - vergleichbare Detailbewertung schulischer Qualität, aber unterschiedliches Globalurteil bzgl. des Erreichens gesetzter Standards – macht auf die Ambivalenz zugrundeliegender

Aggregierungsvorgaben bei der Zusammenfassung von Teil- zu Qualitätskriterienurteilen aufmerksam.

Als schwach benannte Schulen, die sich aufgrund der Inspektionskriterien, die den Output nicht bewerten, auf die Entwicklung innerschulischer Strukturen und Prozesse konzentrieren können, erzielen in der Nachinspektion ca. 1½ Jahre später deutlich bessere Inspektionsbewertungen. Nur in 3 % dieser Schulen ist die Schulqualität auch dann noch nicht hinreichend. Nicht auszuschließen ist aber, dass in den Nachinspektionen Kriterien nicht mit gleicher Strenge angewendet werden, indem z. B. bereits Entwicklungsbemühungen und -fortschritte gewürdigt werden, auch wenn das Ziel noch nicht erreicht ist. Der Vergleich der detaillierten Unterrichtsbewertungen zeigt, dass sich die Qualität in Kriterien, die für das Erreichen der Inspektionsstandards entscheidend sind, stärker verbessert als in anderen Merkmalen.

Aus der weitgehenden Übereinstimmung der Inspektionsprofile von Schulen, die die Erstinspektion erfolgreich durchliefen, und denen, die aufgrund gravierender Mängel eineinhalb Jahre Zeit für die Aufarbeitung der Defizite hatten, lässt sich schlussfolgern, dass ausgeprägte Schwächen, die sich im Profil der Qualitätskriterien zeigen, auch durch besondere Entwicklungsbemühungen nicht ausgeglichen werden können. Möglicherweise sind aber auch einige Inspektionsansprüche unrealistisch.

Zwischen Erst- und Nachinspektion neu ernannte Schulleiter, die durch das Erstinspektionsurteil „gravierende Mängel“ nicht vorbelastete sind, beurteilen die schwache Erstinspektionsbewertung signifikant häufiger als zutreffend als die andere Gruppe. Ob diese Haltungsunterschiede den Schulentwicklungsprozess zwischen den Inspektionen beeinflusst haben, lässt sich nicht feststellen.

Für eine Wertung, ob die Ergebnisse die Einführung bzw. Beibehaltung einer besonderen und mit Sanktionen verbundenen Benennung unzureichender Schulqualität rechtfertigen, reichen die Befunde nicht aus. Einzelfallstudien zu schulischen Reaktionen, u. a. zu negativen Effekten auf die Motivation des Kollegiums oder zur Substanz von Entwicklungsmaßnahmen wären sinnvoll.

Mittelfristig wird auch in Deutschland die Schülerleistung als wesentliches Qualitätskriterium das bisher prozessbezogene Inspektionsprofil ergänzen müssen, um mit dem durch die erweiterte Selbständigkeit von Schulen vorgenommenen Wechsel von einer Input- zur einer Outputsteuerung zu korrespondieren. Dann stellt sich die Frage der Bewertung einer Schule als „failing“ möglicherweise neu und anders.

Literatur

Allen, R.; Burgess, S. (2012): How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England. CMPO Working Paper, 12/287, Bristol.

<http://www.bris.ac.uk/cmppo/publications/papers/2012/wp287.pdf>

BBC (2010): More schools are failing Ofsted checks. BBC News. 10 March 2010.

http://news.bbc.co.uk/2/hi/uk_news/education/8559402.stm

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. BMBF (Hrsg.): Reihe Bildungsforschung Band 31. Berlin

- Bruijn, H. D.; Heuvelhof, E. T.; Koopmans, M. (2007): Law Enforcement: The game between inspectors and inspectees. Boca Raton, Fla: Universal Publishers
- Dedering, K. (2012): Steuerung und Schulentwicklung. Bestandsaufnahme und Theorieperspektive. Wiesbaden: Springer VS.
- Deutsches PISA-Konsortium (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Eder, F.; Dämon, K.; Hörl, G. (2011): Das „Autonomie-Paritäts-Muster“: Vorberuflich erlerntes Stereotyp, Bewältigungsstrategie oder Ergebnis der beruflichen Sozialisation? *Zeitschrift für Bildungsforschung*, 1(3), S.199-217
DOI 10.1007/s35834-011-0021-1
- Ehren, M.C.M; Honingh, M.E . (2011): Risk-based school inspections in the Netherlands: A critical reflection on intended effects and causal mechanisms. In: *Studies in Educational Evaluation* 37 (4), S. 239–248.
- Fend, H. (2008²): Neue Theorie der Schule: Einführung in das Verstehen von Bildungssystemen. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Groeben, N.; Wahl, D.; Schlee, J.; Scheele, B. (Hrsg.) (1988): Das Forschungsprogramm Subjektive Theorien. Eine Einführung in die Psychologie des reflexiven Subjekts. Tübingen: Francke.
- Hattie, J.A.C. (2009): Visible Learning. A synthesis of over 800 meta-analyses relating to achievement. London & New York: Routledge.
- Hessischer Philologenverband (2013): Stellungnahme der im dbb Hessen organisierten Lehrerverbände zur Schulinspektion.
<http://www.hphv.de/hphv/posts/stellungnahme-der-im-dbb-hessen-organisierten-lehrerverbaende-zur-schulinspektion-499.php>
- Hill, H. C.; Charalambous, C. Y.; Kraft, M. A. (2012): When Rater Reliability Is Not Enough: Teacher Observation Systems and a Case for the Generalizability Study. In: *Educational Researcher* 41 (2), S. 56–64.
- Houtveen, T.; van de Grift, W.; Kuijpers, J.; Boot, M.; Groote, F.; Kooijmane, H. (2007). Improving Underperforming Schools. In: *Journal of Education for Students Placed at Risk*, 12 (4), 361–381.
- Huber, R. (2013): Vorwort. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.), Bern 2013, S. 7–8.
- Husfeldt, V. (2011): Externe Schulevaluation - Ein länderübergreifender Blick auf Forschung und Modelle. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.), Bern 2011, S. 13–34.
- Hussain, I. (2012): Subjective Performance Evaluation in the Public Sector: Evidence from School Inspections. Centre for the Economics of Education London School of Economics. London (CEE DP 135).
<http://cee.lse.ac.uk/ceedps/ceedp135.pdf>.
- Internetdokument (2009): Die Schulinspektion – Kommentare.**
http://cfg.wtal.de/03/insp/CFG_QA2009_Kommentare.pdf
- Janssens, F. J. G. (2013): Sehr schwache Primarschulen im Fokus der niederländischen Schulinspektion; In: Quesel, C.; Husfeldt, V.; Landwehr, N.; Steiner, P. (Hrsg.): Bern 2013, S. 61-76

- KMK (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Bonn.
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_01_01-Gesamtstrategie-Endf.pdf
- Lambrecht, M.; Kotthoff, H.-G.; Maag Merki, K. (2011): Rekontextualisierungen: Schulinspektionsprozesse als indirekte Kommunikation zwischen Schulleitung und Kollegium. In: S. Müller, M. Pietsch und W. Bos (Hrsg.): Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht. Münster u.a: Waxmann, S. 165–191.
- Landwehr, N. (2011): Thesen zur Wirkung und Wirksamkeit der externen Schulevaluation. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.): Bern 2011, S. 35–70.
- Lortie, D. C. (1972): Team Teaching: Versuch der Beschreibung einer zukünftigen Schule. In: Dechert, H.W. (Hrsg.): *Team teaching* in der *Schule*. München: Piper. S. 37-76.
- Meyers, C. V. (2013): School Turnaround as National Policy in the United States: Considerations from Three Studies Conducted in the Midwest. In: *Journal of International and Comparative Education*, 2013, Volume 2, Issue 2 2 (2), S. 98–111.
- MK Niedersachsen (2006): Schulinspektion in Niedersachsen. RdErl. d. MK v. 07.04.2006 – 25-80260/2 - VORIS 22410.
<http://www.nibis.de/nibis3/uploads/1nschi/files/materialien/erlass-nschi.pdf>
- Matthews, P.; Sammons, P. (2004): Improvement through inspection. An evaluation of the impact of Ofsted's work. [London]: OFSTEAD; Institute of Education.
- Müller, S.; Pietsch, M.; Bos, W. (Hrsg.) (2011): Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht. Münster u.a: Waxmann.
- Newton, X. A. (2010): Developing indicators of classroom practice to evaluate the impact of district mathematics reform initiative: A generalizability analysis. In: *Studies in Educational Evaluation* 36 (2010) 1–13, S. 1–13.
- Nichols, S. L.; Berliner, D. C. (2007): Collateral damage. How high-stakes testing corrupts America's schools. Cambridge, Mass: Harvard Education Press.
- NLQ (2012): Abschlussbericht Schulinspektion (2006-2012). Hildesheim.
http://www.mk.niedersachsen.de/download/83314/Hier_klicken_zum_Abschlussbericht_Schulinspektion_2006_-_2012.pdf
- Ofsted (2014): The framework for school inspection. Nr. 120100.
<http://www.ofsted.gov.uk/sites/default/files/documents/inspection--forms-and-guides/t/The%20framework%20for%20school%20inspection.doc>
- Praetorius, A.-K. (2014): Messung von Unterrichtsqualität durch Ratings. Münster: Waxmann.
- Psacharopoulos, G. (2007): The Costs of School Failure - A Feasibility Study. Analytical Report for the European Commission. EENEE Analytical Report No. 2. Hg. v. European Expert Network on Economics of Education (EENEE).

http://www.eenee.de/portal/page/portal/EENEEContent/_IMPORT_TELECE NTRUM/DOCS/EENEE_AR2.pdf

- Quesel, C.; V. Husfeldt, V.; N. Landwehr N.; Steiner, P. (Hrsg.) (2011): Wirkungen und Wirksamkeit der externen Schulevaluation. Bern: HEP.
- Quesel, C. (2011): Die beschränkte institutionelle Rationalität von Evaluationen. Fallstudien zum Wandel der Schulinspektion in England. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.): Bern 2011, S. 163–198.
- Quesel, C.; Husfeldt, V.; Landwehr, N. und Steiner, P. (Hrsg.) (2013): Failing schools. Herausforderungen für die Schulentwicklung. Bern: hep.
- Quesel, C.; Husfeldt, V.; Landwehr, N.; Steiner, P. (2013a): Einleitung. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.): Bern 2013, S. 9–23.
- Quesel, C.; Husfeldt, V. (2013): »Failing Schools« im Kontext marktorientierter Bildungspolitik - Eine Analyse zur Reformdynamik in England und den USA. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.): Bern 2013, S. 23–59.
- Ravitch, D. (2010): The death and life of the great American school system. How testing and choice are undermining education. New York: Basic Books.
- Rosenthal, L. (2004) Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review*, 23 (2) 143-151.
- Rürup, M. (2008). Typen der Schulinspektion in den deutschen Bundesländern. *Die Deutsche Schule* 100(4), S. 467–477
- SenBJW (2012): Die zweite Runde der Schulinspektion in Berlin. Handbuch ab Schuljahr 2011/12. Berlin.
- http://www.berlin.de/imperia/md/content/sen-bildung/schulqualitaet/schulinspektion/handbuch_schulinspektion.pdf?start&ts=1332512342&file=handbuch_schulinspektion.pdf#page=1&zoom=auto,0,286
- Scheerens, J.; Bosker, R. (2007): The Foundations of Educational Effectiveness. Oxford (Elsevier)
- Schlee, J. (2014): Schulentwicklung gescheitert! Die falschen Versprechen der Bildungsreformer. Stuttgart: Kohlhammer.
- Schwank, E.; Sommer, N. (2013): Krisendiagnose und Veränderungsdruck - Schulentwicklungen zwischen Erst- und Nachinspektionen in Niedersachsen. In: C. Quesel, V. Husfeldt, N. Landwehr und P. Steiner (Hrsg.): Bern: hep, S. 94–121.
- Sommer, N. (2011): Wie beurteilen schulische Gruppen die erlebte Schulinspektion? Ergebnisse einer Befragung. In: S. Müller, M. Pietsch und W. Bos (Hrsg.): Münster u.a: Waxmann, S. 137–164.
- Sommer, N; Stöhr, C.; Thomas, D. (2010): Schulen mit „gravierenden Mängeln“ - Situation in Niedersachsen und Einsatzmöglichkeiten der Schulentwicklungsberatung. In: Böttcher, W.; Dicke, J. W.; Hogrebe, N. (Hrsg.): *Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit*. Münster: Waxmann, S. 209 – 228

- Sommer, N.; Hornig, G. (2008/2009): Die Diskussion um »failing schools«. Erste Erfahrungen mit Nachinspektionen. In: *Schulverwaltung Niedersachsen*, Teil I, 19, Heft 12/2008, S. 322-325, Teil II, 20, Heft 2/2009, S. 36-39
- Sommer, N. (2013): Schulleitungshandeln im Lichte der Ergebnisse der Schulinspektion. Vortrag auf der 78. AEPF, Dortmund, S. 159
<http://www.aepf2013.de/assets/Abstraktband.pdf>
- The Dutch Inspectorate of Education (2013): The State of Education in the Netherlands. Highlights of the 2011/2012 Education Report. Utrecht.
http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2013/the-state-of-education-in-the-netherlands-2011-2012.pdf#page=1&zoom=auto,0,403
- The New York Times (2012): Panel Says Schools' Failings Could Threaten Economy and National Security.
http://www.nytimes.com/2012/03/20/education/panel-says-schools-failings-could-threaten-economy-and-national-security.html?_r=1&
- The Telegraph (2012): Schools 'bribing pupils' to cheat Ofsted inspections.
<http://www.telegraph.co.uk/education/educationnews/8995377/Schools-bribing-pupils-to-cheat-Ofsted-inspections.html>
- United News Network GmbH (2013): Arbeitgeber weisen Kritik großer Lehrerverbände als unzutreffend und polemisch zurück.
<http://www.pressebox.de/inaktiv/vereinigung-der-hessischen-unternehmerverbaende-e-v-vhu/Arbeitgeber-weisen-Kritik-grosser-Lehrerverbaende-als-unzutreffend-und-polemisch-zurueck/boxid/638548>
- Vieth-Entus, S. (2010): Lehrer wagen keinen Boykottaufruf. Der Tagesspiegel vom 21.4.2010.
<http://www.tagesspiegel.de/berlin/schule/vergleichstests-lehrer-wagen-keinen-boykottaufruf/1805500.html>
- WAZ (2011): Wieder Fehler beim Zentralabitur aufgetaucht.
<http://www.derwesten.de/politik/wieder-fehler-beim-zentralabitur-aufgetaucht-id4607202.html>
- Weick, K. (1975): Educational organizations as loosely coupled systems. In: *Administrative Science Quarterly*, 21 (1976), 1-9 (part).
- Wolf, I. de; Janssens, F. J. G. (2007): Effects and side effects of inspections and accountability in education; an overview of empirical studies. In: *Oxford Review of Education* 33 (3), S. 379–396.