

Vergleichende Analyse der Zusammenhänge von Akzeptanz, Auseinandersetzung mit und Nutzung von Ergebnissen von Vergleichsarbeiten und Schulinspektionen

Inga Wagner, Ingmar Hosenfeld, Michael Zimmer-Müller
Zentrum für Empirische Pädagogische Forschung, Landau

Zusammenfassung: Vergleichsarbeiten und Schulinspektionen sollen die Schul- und Unterrichtsentwicklung fördern. Hierzu werden Lehrkräften bei Vergleichsarbeiten Schulleistungsdaten, bei Schulinspektionen hingegen Prozessdaten u. a. aus Unterrichtsbeobachtungen zurückgemeldet. In der aktuellen Studie sollte an einer Stichprobe von niedersächsischen Lehrkräften untersucht werden, ob Akzeptanz und Ergebnisreflexion Unterrichtsveränderungen bei den beiden Verfahren unterschiedlich gut vorhersagen. Weiterhin sollte die Rolle der empfundenen Bestätigung der eigenen Unterrichtsweise für Unterrichtsveränderungen untersucht werden. Die Ergebnisse zeigen, dass Akzeptanz bei Vergleichsarbeiten ein stärkerer Prädiktor für Unterrichtsveränderungen ist als bei Schulinspektionen. Ergebnisauswertung ist bei beiden Verfahren ein schwacher Prädiktor für Unterrichtsveränderungen. Die wahrgenommene Bestätigung durch die Evaluationsergebnisse ist bei Schulinspektionen stärker ausgeprägt als bei Vergleichsarbeiten und übt einen tendenziell positiven Einfluss auf Veränderungen im Unterricht aus.

Schlüsselbegriffe: Vergleichsarbeiten, Schulinspektionen, Unterrichtsentwicklung

Comparative Analysis of the Relations Between Acceptance, Reflection of Results and use of Results of Performance Tests and of School Inspections

Summary: Performance tests and school inspections are supposed to foster school and lesson development at German schools. Test feedback informs teachers about their students' performances, whereas inspection reports depict procedural data of lesson observations. The current study examined at a teacher sample in Lower Saxony (Northern Germany) if acceptance, reflection of results and perceived confirmation by results play different roles for predicting teachers' improvement actions after performance tests and after school inspections. Results show that acceptance is a better predictor for change actions after performance tests than after school inspections. Reflection of results exerts only a weak influence on change actions after both feedback methods. Teachers feel more confirmed by inspection results than by test feedback which has a slightly positive impact on their change actions.

Keywords: Performance tests, school inspections, lesson development

Vergleichsarbeiten (im Folgenden wird auch die Abkürzung *VERA* verwendet) und Schulinspektionen (im Folgenden wird auch die Abkürzung *SI* verwendet) sind externe Evaluationsverfahren, die die Schul- und Unterrichtsentwicklung an deutschen Schulen fördern sollen (Kultusministerkonferenz [KMK], 2016).

VERA wird in jährlichem Turnus in den Bundesländern in den Klassenstufen 3 und 8 durchgeführt. Lehrkräfte erhalten durch *VERA* eine Rückmeldung zum Leistungsstand ihrer Klassen. Dadurch können sie Stärken und Schwächen ihrer Klassen identifizieren und Änderungen für ihren Unterricht ableiten. In Niedersachsen, wo die vorliegende Studie

2015/2016 mit Sekundarschullehrkräften durchgeführt wurde, ist in der Klasse 8 Mathematik verpflichtendes Testfach, sofern die Schule dies nicht anders festlegt. Die Rückmeldungen enthalten Informationen zu den erreichten Kompetenzstufen und den Lösungsquoten der Aufgaben auf Individual- und Klassenebene, teilweise auch im Vergleich zu Landeswerten (s. auch Niedersächsisches Kultusministerium, 2014).

In Deutschland werden in den Bundesländern unterschiedliche Schulinspektionsverfahren eingesetzt, wobei die zugrundeliegenden Konzeptionen oft sehr ähnlich sind (Dedering, 2016; Husfeldt, 2011). In ein paar Bundesländern (z. B. Rheinland-Pfalz) wird die Schulinspektion nicht mehr durchgeführt. In dem niedersächsischen Verfahren von 2012 bis 2017 wurde den Schulen in Inspektionsberichten eine Einschätzung der Prozessqualität sogenannter *Kernaufgaben* zurückgemeldet. Die Kernaufgaben sind mehr auf der schulischen Handlungsebene angesiedelt (z. B. „Kooperationen entwickeln“), sollen aber letztendlich der Unterrichtsentwicklung dienen. Zudem enthielten die Berichte Ergebnisse aus Unterrichtsbeobachtungen, die während des Schulbesuchs von Inspektionsteams bei einer schulabhängig variablen Anzahl von Lehrkräften durchgeführt wurden (genauere Beschreibungen des Verfahrens finden sich bei Niedersächsisches Kultusministerium, 2014, sowie Niedersächsisches Landesinstitut für schulische Qualitätsentwicklung, 2014).

Es wird somit deutlich, dass zum Zeitpunkt der vorliegenden Studie deutliche Verfahrensunterschiede bei niedersächsischen Vergleichsarbeiten und Schulinspektionen vorhanden waren, die aufgrund der ähnlichen Konzeptionen zumindest teilweise auf andere Bundesländer übertragbar sind. Folgende Punkte sind dabei hervorzuheben:

1. Der Evaluationsgegenstand von Schulinspektionen sind Unterrichtsprozesse, der von VERA ist die Schülerleistung (Unterrichtsausgangspunkt).

2. Bei Schulinspektionen werden neben dem Unterricht auch andere schulische Bereiche, wie Führungskompetenzen der Schulleitung, bewertet, bei VERA ist dies nicht der Fall.
3. Bei VERA erhält jede Lehrkraft eine individualisierte Rückmeldung zum Leistungsstand ihrer Klasse. Bei Schulinspektionen erhält eine Lehrkraft für die gesamte Schule aggregierte Ergebnisse zu Unterrichtsbeobachtungen.

Aus diesen Unterschieden heraus lässt sich womöglich teilweise erklären, warum die Verfahren von Lehrkräften und Schulleitungen unterschiedlich bewertet werden. Studien zeigten, dass sich das schulische Personal mehr mit Inspektionsals mit VERA-Ergebnissen auseinandersetzt (z. B. Demski, Rosenbusch, van Ackeren, Clausen & Schmidt, 2012). Schulinspektionen werden tendenziell besser akzeptiert, während die Nützlichkeitsbewertungen von VERA häufig niedriger ausfallen (z. B. Groß Ophoff, Koch & Hosenfeld, 2018; Wurster, Richter, Schliesing & Pant, 2013). Es fällt auf, dass Schulinspektionen trotz einer tendenziell höheren Akzeptanz und einer intensiveren Ergebnisrezeption im Vergleich zu VERA häufig ebenso zu marginalen Effekten auf die Unterrichtsentwicklung führen (z. B. Böhm-Kasper & Selders, 2013; Wurster, Richter & Lenski, 2017).

Ziel der aktuellen Studie ist, Wirkmechanismen zu analysieren, die mit erklären könnten, warum Schulinspektionen trotz tendenziell besserer Voraussetzungen als VERA nicht zu mehr Unterrichtsveränderungen führen. Dazu werden Zusammenhangsanalysen von VERA und SI im direkten Vergleich durchgeführt und dargestellt, was in der Forschungspraxis bislang selten erfolgt ist. Als ein möglicher Wirkmechanismus wird zudem die Rolle der von Lehrkräften empfundenen Bestätigung durch VERA- und SI-Ergebnisse für Unterrichtsveränderungen analysiert. Unserem Kenntnisstand nach wurde diese Variable in bisherigen Studien weniger fokussiert betrachtet. Die Ergebnisse der vorliegenden Studie beziehen sich zunächst auf den niedersächsischen Kontext von VERA und Schulinspektionen.

Theoretischer Hintergrund

Theoretische Wirkannahmen

Nach dem Sequenzmodell der Unterrichtsentwicklung (Helmke, 2012), das zunächst für Vergleichsarbeiten formuliert wurde, sind die Rezeption und die Reflexion der Evaluationsergebnisse, die zusammengefasst auch als eine „Auseinandersetzung“ mit den Ergebnissen betrachtet werden können (Groß Ophoff, 2013), Voraussetzungen für die Umsetzung von Maßnahmen nach der externen Evaluation (*Aktion*). Solche Maßnahmen können beispielsweise in der Weiterentwicklung des Unterrichts (Helmke, 2012) oder in Veränderungen des Lehrerverhaltens (Reezigt & Cremers, 2005) bestehen. Akzeptanz ist im Modell von Helmke (2012) ein individueller Bedingungsfaktor, der die Auseinandersetzung mit den Ergebnissen und die Umsetzung von Verbesserungsmaßnahmen positiv beeinflusst. Weitere individuelle Bedingungsfaktoren sind z. B. das *professionelle Wissen* oder die *Selbstwirksamkeit* von Lehrkräften. *Akzeptanz* und *Auseinandersetzung* mit Ergebnissen werden in dem Modell somit als Voraussetzungen für *Anschluss handeln* angesehen, was auch im konzeptionellen Modell der Schulinspektion von Ehren, Altrichter, McNamara und O'Hara (2013) zum Ausdruck kommt. Hier sind Entwicklungsaktivitäten nach der Schulinspektion mit davon abhängig, dass das Feedback wahrgenommen und akzeptiert wird (Altrichter & Kemethofer, 2016). In beiden Modellen spielen auch externe Faktoren, wie das *schulische Evaluationsklima* oder die *Handlungen von Stakeholdern* (z. B. Eltern), eine Rolle für schulische Veränderungen.

Bezogen auf die empirischen Befunde bedeuten die theoretischen Modellvorstellungen, dass die tendenziell bessere Akzeptanz und intensivere Ergebnisrezeption bei Schulinspektionen im Vergleich zu VERA bei gleichen Einflussstärken dieser Faktoren zu eher mehr Unterrichtsentwicklung nach Schulinspektionen als nach VERA führen müssten. Dies scheint jedoch aufgrund der empirischen Befundlage

mehrheitlich nicht der Fall zu sein (Böhm-Kasper & Selders, 2013; Chapman, 2001; Demski et al., 2012). Es ist somit möglich, dass Akzeptanz und Ergebnisrezeption bei Schulinspektionen zwar quantitativ etwas stärker ausgeprägt sind, aber nicht so viel Einflussstärke auf Unterrichtsverbesserungen haben als bei Vergleichsarbeiten. Dadurch würden beide Verfahren letztendlich zu ungefähr ähnlichem Anschluss handeln auf Unterrichtsebene führen. Diese Überlegungen beziehen sich auf eine abstraktere Betrachtungsweise der Zusammenhänge zwischen den drei Variablen.

Dass Schulinspektionen trotz tendenziell höherer Akzeptanz und intensiverer Ergebnisrezeption nicht zu mehr Unterrichtsveränderungen als VERA führen, könnte mehrere praktische Gründe haben. Einer davon könnte in der Ergebnisdarstellung in der Rückmeldung liegen.

Bei VERA erhält eine Lehrkraft eine individualisierte Rückmeldung zum Leistungsstand ihrer Klasse. Schwächere VERA-Ergebnisse können somit von der Lehrkraft eher auf sich selbst oder höchstens auf die vorigen Lehrkräfte der Klassen attribuiert werden. Dadurch fühlt sie sich nach schwächeren VERA-Ergebnissen eher nicht in ihrer Unterrichtsweise bestätigt und kann eher eine subjektive Notwendigkeit zur Veränderung ableiten. Dieses *Gefühl der Notwendigkeit* ist ein weiterer individueller Bedingungsfaktor in dem Modell von Helmke (2012), der Anschluss handeln positiv beeinflusst. Sehr schlechte VERA-Ergebnisse scheinen von Lehrkräften jedoch auch selbstwertschützend externen Ursachen zugeschrieben zu werden (Bonsen, Büchter & Peek, 2006; Mezulis, Abramson, Hyde & Hankin, 2004).

Nach Schulinspektionen werden für die gesamte Schule zusammengefasste Ergebnisse aus Unterrichtsbeobachtungen zurückgemeldet. Diese Darstellung liefert ein vollständigeres Bild über die Unterrichtsqualität an der gesamten Schule. Allerdings ist sie für die einzelne Lehrkraft informationsärmer, da eine gewisse Unsicherheit besteht, ob und inwiefern die Be-

wertung auf sie persönlich zutrifft (s. auch Demski et al., 2012). Aus dieser Unsicherheit heraus bieten aggregierte Rückmeldungen einer Lehrkraft die Möglichkeit, schwächere Ergebnisse, neben möglichen anderen externen Ursachen (Kohler, 2004), auch der schlechteren Unterrichtsqualität anderer Lehrkräfte zuzuschreiben (*self-serving attributional bias*, s. Mezulis et al., 2004). Somit wäre denkbar, dass es Lehrkräften nach schwächeren Inspektionsergebnissen noch mehr als nach VERA möglich ist, sich weiterhin in ihrer Unterrichtsweise bestätigt und keinen Änderungsbedarf zu sehen. Ein höheres Gefühl der Bestätigung und somit eine womöglich geringere Änderungsmotivation könnten bei Schulinspektionen mitbedingen, warum die tendenziell höhere Akzeptanz und Ergebnisreflexion im Vergleich zu VERA nicht zu mehr Unterrichtsveränderungen führen.

Empirischer Forschungsstand

In den Studien des Projektes „Evis“ zeigte sich, dass VERA von Lehrkräften und Schulleitungen leicht nützlicher eingestuft wird als Schulinspektionen. Die Inspektionsergebnisse werden jedoch stärker rezipiert als die VERA-Ergebnisse. Schulleitungen nutzen lediglich Inspektionsergebnisse in stärkerem Maß, Lehrkräfte nutzen weder VERA- noch Inspektionsergebnisse (Demski, 2017). In den Studien des Projektes „Stabil“ zeigte sich, dass Schulleitungen Schulinspektionen als ähnlich aufwändig und nützlich, aber etwas besser in der Diagnosegüte einstufen als VERA. Inspektionsergebnisse werden häufiger in Schulkonferenzen thematisiert als VERA-Ergebnisse. Inspektionsergebnisse werden relativ häufig, aber weniger oft als VERA-Ergebnisse, zur Ableitung von Maßnahmen im Unterrichtsbereich *Lehr- und Lernprozesse* genutzt. Allerdings bleibt die Intensität dieser Aktivitäten unklar (Wurster et al., 2013).

In Studien, die Schulinspektionen für sich genommen fokussierten, zeigte sich, dass Schulinspektionen bei einer größtenteils guten Ak-

zeptanz oft nur geringe bis moderate Effekte auf die Unterrichtsentwicklung haben (Böhm-Kasper & Selders, 2013; Chapman, 2001; Penningcx, Vanhoof, De Maeyer & van Petegem, 2016; Wurster, Feldhoff & Gärtner, 2016). Bei VERA zeigten sich in aktuelleren Studien geringere Nützlichkeitsbewertungen durch Lehrkräfte (Groß Ophoff et al., 2018; Wurster et al., 2016). In der Studie von Wurster, Richter und Lenski (2017) berichten Lehrkräfte eher weniger, nach VERA neue Unterrichtsmethoden eingeführt zu haben.

In der Studie von Demski (2017) zeigte sich, dass die Nutzung von Lernstandserhebungen durch die wahrgenommene Nützlichkeit und noch mehr durch die Auseinandersetzung mit den Ergebnissen beeinflusst wird. Bei Schulinspektionen hingegen wird die Nutzung der Ergebnisse stärker durch die eingeschätzte Nützlichkeit als durch die Ergebnisreflexion beeinflusst. Wurster et al. (2016) fanden, dass die wahrgenommene Nützlichkeit von VERA der relevanteste Prädiktor für weitere Entwicklungsaktivitäten auf der Schulleitungs- sowie der Lehrkräfteebene ist. Auch die Kommunikation über die VERA-Ergebnisse stellte sich als ein wichtiger Prädiktor heraus.

Kühle und van Ackeren (2012) zeigten ebenfalls, dass die Akzeptanz von Lernstandserhebungen mitentscheidend für eine Verbesserung der Lehr- und Lernprozesse ist. VERA-Ergebnisse werden zudem eher dann zur Planung von Fortbildungen in Fachkonferenzen genutzt, wenn zuvor ein intensiver Austausch darüber stattgefunden hat (Bach, Wurster, Thillmann, Pant & Thiel, 2014). In der Studie von Altrichter und Kemethofer (2015) zeigte sich, dass eine gute Akzeptanz des Inspektionsfeedbacks eher zu Verbesserungen bei schulischen Selbstevaluationen führt. Dieser Befund gilt für Österreich und Schweden, nicht jedoch für England. In der niederländischen Studie von Ehren und Visscher (2008) zeigte sich, dass die Akzeptanz des Inspektionsfeedbacks nicht signifikant mit der Zahl der schulischen Verbesserungsmaßnahmen korreliert ist.

Hypothesen

In der vorliegenden Studie sollen Erklärungsmöglichkeiten analysiert werden, warum die tendenziell höhere Akzeptanz und Ergebnisrezeption bei Schulinspektionen im Vergleich zu VERA in einigen Studien nicht zu mehr Unterrichtsverbesserungen führen. Stringente kausale Schlüsse sind dabei nicht möglich. Wie oben ausgeführt, mag dies, in Anlehnung an das Modell von Helmke (2012) betrachtet, daran liegen, dass Akzeptanz und Auseinandersetzung mit Ergebnissen bei Schulinspektionen schwächere Prädiktoren für Unterrichtsveränderungen sind als bei VERA.

Zur Forschungsfrage 1

Gibt es Unterschiede in den Zusammenhangsstrukturen von Akzeptanz, Ergebnisauswertung und Unterrichtsveränderungen bei VERA und SI?

Es resultieren folgende Hypothesen:

1. Hypothese: Akzeptanz ist bei SI ein schwächerer Prädiktor für Unterrichtsveränderungen als bei VERA.
2. Hypothese: Auseinandersetzung mit Ergebnissen ist bei SI ein schwächerer Prädiktor für Unterrichtsveränderungen als bei VERA.

Ein praktischer Grund, warum die tendenziell besseren Voraussetzungen von Schulinspektionen im Vergleich zu VERA häufig nicht zu mehr Unterrichtsverbesserungen führen, könnte in der Art der Ergebnisdarstellung in den Rückmeldungen liegen. Wie oben ausgeführt, bieten die aggregierten Ergebnisdarstellungen bei Schulinspektionen im Vergleich zu den individualisierten Rückmeldungen bei VERA Lehrkräften eventuell noch mehr die Möglichkeit, auch schwächere Ergebnisse persönlich günstig zu attribuieren. Daher könnten sie sich mehr durch Inspektions- als durch VERA-Ergebnisse in ihrer Unterrichtsweise bestätigt sehen. Dies könnte, durch die eventuell damit einhergehende geringere Änderungsmotivation, eine negative Voraussetzung für Unterrichtsveränderungen nach Schulinspektionen sein.

Zur Forschungsfrage 2

Fühlen Lehrkräfte sich durch die Ergebnisse von VERA und SI unterschiedlich stark in ihrer Unterrichtsweise bestätigt und wirkt sich dies auf Veränderungen im Unterricht aus?

Es resultieren folgende Hypothesen:

3. Hypothese: Lehrkräfte fühlen sich mehr durch Inspektions- als durch VERA-Ergebnisse in ihrer Unterrichtsweise bestätigt.
4. Hypothese: Je mehr sich Lehrkräfte in ihrer Unterrichtsweise nach VERA und Schulinspektionen bestätigt fühlen, desto weniger Veränderungen nehmen sie nach den Verfahren an ihrem Unterricht vor.

Methoden

Design und Ablauf der Studie

Die aktuelle Studie fand im Rahmen des Projekts „Evaluation der Impulswirkung von Schulinspektion und Vergleichsarbeiten auf die Qualitätsentwicklung an Schulen“ (EISVQS) statt, das vom „Zentrum für Empirische Pädagogische Forschung“ der Universität Koblenz-Landau im Auftrag des Niedersächsischen Kultusministeriums durchgeführt wurde. Im Rahmen des Projekts wurde ein Online-Fragebogen konstruiert und implementiert. In dem Fragebogen wurden neben anderen Themenbereichen die Akzeptanz von SI und VERA, die Auseinandersetzung mit ihren Ergebnissen, das persönliche Anschlusshandeln nach den Verfahren sowie die subjektiv wahrgenommene Bestätigung durch die Inspektions- und VERA-Ergebnisse erfragt. Die Bearbeitung des Fragebogens war freiwillig und anonym und konnte jederzeit abgebrochen oder unterbrochen werden. Zur Teilnahme an der Studie wurden Schulleitungen und Lehrkräfte von 100 niedersächsischen Gymnasien, Integrierten Gesamtschulen und Realschulen eingeladen. In der Stichprobe waren ausschließlich Schulen, bei denen zwischen November 2014 und Januar 2016 eine Schulinspektion stattgefunden hatte. Die Befragung des schulischen Personals fand zwischen Juli 2015 und Februar 2016 in verschiedenen Erhebungswellen statt, so dass in den meisten Schulen die Befragung ungefähr 8 Wochen nach der Schulinspektion durchgeführt wurde (Dauer des Befragungszeitraums: 4–32 Wochen nach der SI).

Stichprobe

Da das Hauptkriterium der aktuellen Studie Veränderungen im Unterricht waren, die vor allem Lehrkräfte betreffen, wurden in die Stichprobe 1422 Lehrkräfte aus 29 Gymnasien, 31 Integrierten Gesamtschulen und 28 Realschulen aufgenommen, die den Fragebogen bis zum Ende ausgefüllt hatten. Die Rücklaufquote aus der Gesamtheit der eingeladenen Lehrkräfte lag bei ca. 26%. In die Analysen wurden nur Personen aufgenommen, die im Schuljahr 2014/15 mit einer Klasse an VERA teilgenommen hatten und/oder deren Unterricht während der SI beobachtet worden war und die die Ergebnisse rezipiert hatten. Dies erschien für die Interpretation der Ergebnisse anhand der angenommenen Wirkmechanismen konzeptionell sinnvoll. Dadurch war die potenzielle Stichprobe für VERA-Analysen kleiner als die für SI-Analysen, da an VERA nur Lehrkräfte der achten Klassen, an den Unterrichtsbeobachtungen jedoch Lehrkräfte teilweise aller Klassenstufen teilnahmen. Es wurden zudem Personen ausgeschlossen, die bei den relevanten Variablen fehlende Werte hatten oder die aufgrund von Box Plots als Ausreißerwerte identifiziert wurden. In die VERA-Analysen gingen die Daten von 102 Lehrkräften aus 18 Realschulen, 16 Integrierten Gesamtschulen und 17 Gymnasien ein. Pro Schule waren zwischen 1 und 5 Lehrkräfte in der Stichprobe enthalten ($Md=2$). Bezüglich der SI konnten die Daten von 793 Lehrkräften aus 28 Realschulen, 30 Integrierten Gesamtschulen und 29 Gymnasien ausgewertet werden. Der gültige Rücklauf pro Schule lag zwischen 1 und 37 Personen ($Md=7$).

Erhebungsinstrumente

Messung der Akzeptanz

In den Modellen von Helmke (2012) und von Ehren et al. (2013) wird Akzeptanz als ein Bedingungsfaktor für Anschlusshandeln angesehen, ohne dass das Konstrukt dabei genauer definiert wird. Im „Technology Acceptance Model“ zur Nutzung neuer Informationstechnologien von Davis, Bagozzi und Warshaw (1989) ist die Intention, eine Technologie zu nutzen, von der Einstellung gegenüber der Technologie abhängig. Diese wird wiederum durch die wahrgenommene Nützlichkeit der Technologie mitbestimmt. Definiert man die Akzeptanz von VERA und SI als eine positive Einstellung gegenüber diesen Verfahren, kann sie nach dem Modell von Davis et al. (1989) durch die wahrgenommene Nützlichkeit

dieser Verfahren zumindest teilweise operationalisiert werden. In der vorliegenden Studie wurde die wahrgenommene allgemeine Nützlichkeit von VERA und SI erfasst, die die empfundene Nützlichkeit der Verfahren für Schul- und Unterrichtsentwicklung im Allgemeinen widerspiegelt (Beispielitem: „VERA/SI tragen dazu bei, dass Lehrkräfte sich mehr bemühen.“). Die Skalen wurden zu VERA und zu SI analog konstruiert und beinhalteten sieben Items, von denen zwei Items aus der Einstellungsskala von Ditton und Merz (2000) entnommen wurden. Fünf Items stellten Eigenkonstruktionen dar. Das Antwortformat war vierstufig (*trifft überhaupt nicht zu bis trifft völlig zu*). Konfirmatorische Faktorenanalysen zeigten, bis auf leicht erhöhte RMSEA-Werte, einen guten Modellfit für eine einfaktorielle Struktur der Items sowohl bei VERA als auch bei SI ($RMSEA_{VERA} = .12$, $CFI_{VERA} = .96$, $SRMR_{VERA} = .04$; $RMSEA_{SI} = .09$, $CFI_{SI} = .98$, $SRMR_{SI} = .03$). Itemanalysen wiesen auf gute Reliabilitäten der beiden Skalen hin (Cronbachs $\alpha_{VERA} = .92$; Cronbachs $\alpha_{SI} = .90$). Für die Analysen wurde für jede Person, die mindestens 4 gültige Antworten auf der Skala hatte, ein MEAN-Wert gebildet. Um eine bessere Vergleichbarkeit der MEANs angesichts potenziell unterschiedlicher Itemschwierigkeiten zu gewährleisten, waren die Personenwerte zuvor jeweils um die Itemmittelwerte zentriert worden (außer für die Darstellung der deskriptiven Statistiken).

Messung der Auseinandersetzung mit den Ergebnissen

Die Messung der Auseinandersetzung mit den VERA-Ergebnissen erfolgte mit Hilfe eines Items, das die Dauer der Beschäftigung mit den VERA-Ergebnissen erfasste. Das Item lautete: „Wie viel Zeit haben Sie persönlich nach der Dateneingabe darauf verwendet, sich mit den Rückmeldungen aus den Vergleichsarbeiten zu beschäftigen (Einsicht in Rückmeldungen im VERA-Portal, diskutieren mit Kolleginnen und Kollegen, in Gremien...)?“ Die Angabe erfolgte in Stunden. Das Item zur SI wurde analog konstruiert. Die Dauer der Beschäftigung mit den Ergebnissen wurde als ein Indikator für die Intensität der Auseinandersetzung mit den Ergebnissen angesehen, da eine längere Rezeptionszeit vermutlich mit vielfältigeren Rezeptionsaktivitäten und mit einer wiederholten Beschäftigung mit den Ergebnissen einhergeht. Zu berücksichtigen ist, dass längere Rezeptionszeiten bei manchen Lehrkräften auch auf Verständnisschwierigkeiten hindeuten können.

Messung des persönlichen Anschlusshandelns im Unterricht

Die Messung des persönlichen Anschlusshandelns nach VERA erfolgte ebenfalls mit Hilfe eines Items, das mittels einer vierstufigen Ratingskala beantwortet werden konnte. Das Item lautete: „An meiner Unterrichtsweise hat sich durch die Vergleichsarbeiten nichts geändert.“ Das Item zur SI wurde analog konstruiert. Für die Analysen wurden die Items umgepolt. Änderungen in der Unterrichtsweise deuten auf ein höheres persönliches Anschlusshandeln im Unterricht hin. Veränderungen im Unterricht können in den meisten Fällen als ein positiver Effekt von VERA und SI gewertet werden, da auch bei guten Evaluationsergebnissen häufig noch ein Änderungspotenzial für den Unterricht gegeben ist. Die Studie von Bellmann, Schweizer und Thiel (2016) in vier deutschen Bundesländern zeigte, dass eindeutig negative Veränderungen im Lehrerverhalten, wie beispielsweise Betrug, zumindest nach Angabe von Lehrkräften, weniger häufig vorkommen.

Messung der subjektiv wahrgenommenen Bestätigung

Die Messung der subjektiv wahrgenommenen Bestätigung der eigenen Unterrichtsweise durch die VERA-Ergebnisse erfolgte mittels eines Items, das lautete: „Ich sehe mich durch die Ergebnisse der Vergleichsarbeiten in meiner Art zu unterrichten bestätigt.“ Das Item konnte auf einer vierstufigen Ratingskala beantwortet werden. Das Item zu SI wurde analog konstruiert.

Auswertungsmethoden

Die Regressionskoeffizienten von Akzeptanz, Auseinandersetzung und Bestätigung auf Anschlusshandeln im Unterricht für VERA und für SI wurden mit bivariaten und multiplen linearen Regressionsmodel-

len berechnet. In die multiplen Modelle gingen die Prädiktoren Akzeptanz, Auseinandersetzung und Bestätigung sowie *Schulform* als Kontrollvariable simultan ein. Um die Nützlichkeiten der einzelnen Prädiktoren zu bestimmen, wurden zusätzlich sequenzielle multiple Regressionen berechnet. Die Prädiktoren gingen in aufsteigender Stärke in die Modelle ein.

Da die Daten eine hierarchische Struktur aufweisen, wurden bei SI für alle Schulen mit min. 5 teilnehmenden Lehrkräften zusätzlich Mehrebenenanalysen mit der Software „HLM“ durchgeführt. Die multiplen Modelle waren analog zu denen der linearen Regressionen aufgebaut, wobei die Kontrollvariable *Schulform* auf der zweiten Ebene lag. Mit Hilfe von Random Intercepts-Modellen, in die die Level 1-Prädiktoren um die Gruppenmittelwerte zentriert eingingen, konnten somit die Regressionskoeffizienten unter Einbezug der schulischen Ebene nochmals geschätzt werden. Als zusätzliche Effektstärken werden sog. „Pseudo-R²“-Werte (Eid, Gollwitzer & Schmitt, 2010, S. 717) berichtet. Diese Werte geben an, inwiefern sich die Level 1-Residualvarianz des Modells ohne den betreffenden Prädiktor durch die Hinzunahme des Prädiktors anteilig verringert. Von 51 Schulen in der VERA-Stichprobe konnten bei 41 Schulen nur die Daten von jeweils 1 oder 2 Personen ausgewertet werden. Daher wurde bei VERA auf die Durchführung von Mehrebenenanalysen verzichtet. (Eine ausführlichere Beschreibung der statistischen Verfahren findet sich z. B. bei Eid et al., 2010.)

Ergebnisse

Zur besseren Interpretation der Regressionskoeffizienten werden zunächst die deskriptiven Statistiken der relevanten Variablen für die Lehrkräfte aufgelistet, die die Fragen zu mindestens einem der Verfahren beantwortet haben (s. Tabelle 1):

Tab. 1: Deskriptive Statistiken von Akzeptanz, Auseinandersetzung, Anschlusshandeln und wahrgenommener Bestätigung bei VERA und SI.

	VERA (n = 102)		SI (n = 793)		VERA vs. SI
	M	SD	M	SD	d
Akzeptanz ^a	2.15	0.64	2.53	0.59	.62
Auseinandersetzung (in h)	2.15	1.27	3.64	2.95	.66
Anschlusshandeln ^a	2.15	0.78	1.89	0.84	.32
Bestätigung ^a	2.59	0.69	2.74	0.85	.19

Anmerkungen: ^a: Antwortkategorien: 1 (trifft gar nicht zu) – 4 (trifft völlig zu).

Tabelle 1 zeigt, dass SI im Vergleich zu VERA mehr akzeptiert und ihre Ergebnisse mehr reflektiert werden. Die Unterschiede weisen mittlere Effektstärken auf. SI führen zu leicht weniger Unterrichtsveränderungen als VERA. Die Effektstärke ist gering. Lehrkräfte fühlen sich durch die Ergebnisse der SI stärker in ihrer Unterrichtsweise bestätigt als durch VERA-Ergebnisse. Der praktische Effekt ist gering. Dieses hypothesenrelevante Ergebnis wird durch den direkten Vergleich der Lehrkräfte, die die Fragen zu beiden Verfahren beantwortet haben ($n=67$), bestätigt ($M_{\text{VERA}}=2.61, SD_{\text{VERA}}=0.63; M_{\text{SI}}=2.81, SD_{\text{SI}}=0.82; t(66)=-2.20, p(1\text{-seitig})=.02, d=.27$). Hypothese 3 wird somit bestätigt.

Tabelle 2 zeigt die Kennwerte für die Zusammenhänge zwischen Akzeptanz, Auseinandersetzung und Bestätigung und dem persönlichen Anschlusshandeln im Unterricht bei VERA und bei SI aus bivariaten und multiplen Regressionen.

Die Tabelle zeigt, dass in der bivariaten Betrachtung sowohl die B- als auch die β -Koeffizienten der Akzeptanz bei VERA 1.5- bis 2-mal so hoch sind wie bei SI. In der multiplen Betrachtung sind die Unterschiede etwas geringer ausgeprägt. Akzeptanz klärt im bivariaten Modell bei VERA ca. 14% der Varianz im Anschlusshandeln auf, bei SI sind es nur 4%. Die Ergebnisse bestätigen somit die Annahme der Hypothese 1, dass Akzeptanz bei SI ein schwächerer Prädiktor für Unterrichtsveränderungen ist als bei VERA.

Auseinandersetzung klärt im bivariaten Modell bei VERA 0.4% der Varianz des Anschlusshandelns auf, bei SI sind es 1.6%. Die Regressionskoeffizienten zeigen, dass Auseinandersetzung bei SI ein schwach positiver Prädiktor für Unterrichtsveränderungen zu sein scheint und bei VERA keinen signifikanten Einfluss auf das Anschlusshandeln hat. Hypothese 2, dass Auseinandersetzung ein schwächerer positiver Prädiktor für Unterrichtsveränderungen bei SI ist als bei VERA, kann nicht bestätigt werden.

Die gefühlte Bestätigung durch die Evaluationsergebnisse ist bei VERA ein positiver Prädiktor für Unterrichtsveränderungen. Im bivariaten Modell ist der Zusammenhang signifikant. Bei SI hat die empfundene Bestätigung kaum einen Einfluss auf das Anschlusshandeln. Im bivariaten Modell klärt die empfundene Bestätigung bei VERA ca. 6% der Varianz im Anschlusshandeln auf, bei SI sind es 0.1%. Bestätigung ist bei VERA ein stärkerer und tendenziell positiver Prädiktor für Unterrichtsveränderungen als bei SI. Hypothese 4, dass Bestätigung generell ein negativer Prädiktor für Anschlusshandeln ist, kann nicht bestätigt werden. In den simultanen multiplen Regressionen liegt der adjustierte Determinationskoeffizient aller Prädiktoren bei VERA bei .141, bei SI liegt er bei .060.

Die gefühlte Bestätigung durch die Evaluationsergebnisse ist bei VERA ein positiver Prädiktor für Unterrichtsveränderungen. Im bivariaten Modell ist der Zusammenhang signifikant. Bei SI hat die empfundene Bestätigung kaum einen Einfluss auf das Anschlusshandeln. Im bivariaten Modell klärt die empfundene Bestätigung bei VERA ca. 6% der Varianz im Anschlusshandeln auf, bei SI sind es 0.1%. Bestätigung ist bei VERA ein stärkerer und tendenziell positiver Prädiktor für Unterrichtsveränderungen als bei SI. Hypothese 4, dass Bestätigung generell ein negativer Prädiktor für Anschlusshandeln ist, kann nicht bestätigt werden. In den simultanen multiplen Regressionen liegt der adjustierte Determinationskoeffizient aller Prädiktoren bei VERA bei .141, bei SI liegt er bei .060.

Tab. 2: Regressionskoeffizienten aus linearen Regressionen von Akzeptanz, Auseinandersetzung und Bestätigung auf das persönliche Anschlusshandeln im Unterricht bei VERA und bei SI.

	Bivariate Regression				Multiple Regression			
	B	SE	β	R^2_{adj}	B	SE	β	ΔR^2_{adj}
Akzeptanz								
VERA	.48***	.11	.39	.144	.46**	.14	.38	–
SI	.30***	.05	.21	.042	.32***	.06	.22	–
Auseinandersetzung								
VERA	-.04	.06	-.06	.004 [‡]	-.08	.06	-.12	.008
SI	.04***	.01	.13	.016 [‡]	.04***	.01	.12	.015
Bestätigung								
VERA	.29**	.11	.26	.059	.12	.12	.11	.004
SI	.05	.04	.05	.001	-.05	.04	-.05	.001

Anmerkungen: *, $p < .05$, **, $p < .01$, ***, $p < .001$; ‡: nicht adjustierte Determinationskoeffizienten.

Tab. 3: Regressionskoeffizienten aus Mehrebenenanalysen von Akzeptanz, Auseinandersetzung und Bestätigung auf das persönliche Anschlusshandeln im Unterricht bei SI.

	Bivariate Mehrebenenanalyse				Multiple Mehrebenenanalyse			
	B	SE	β	Pseudo- R^2	B	SE	β	Pseudo- R^2
Akzeptanz	.30***	.06	.24	.037	.34***	.06	.25	.037
Auseinandersetzung	.04**	.01	.13	.013	.04**	.01	.13	.012
Bestätigung	.02	.04	.04	-.001	-.08*	.04	-.06	.003

Anmerkungen: *: $p < .05$, **: $p < .01$, ***: $p < .001$.

Für die Daten zu Schulinspektionen wurden zusätzlich Mehrebenenanalysen durchgeführt. Die Intraklassenkorrelation liegt bei .02. Tabelle 3 zeigt, dass Mehrebenenanalysen zu ähnlichen Ergebnissen führen wie die linearen Regressionen und somit die oben gemachten Aussagen tendenziell bestätigen. Im simultanen multiplen Modell klären alle Prädiktoren 5.5 % der Level 1- und 12.5 % der Level 2-Residualvarianz des Nullmodells auf.

Als Zusatzbefund zeigt sich, dass Gymnasiallehrkräfte VERA mindestens marginal signifikant weniger gut akzeptieren als Lehrkräfte anderer Schulformen ($M_{\text{GYM}} = 1.85$, $SD_{\text{GYM}} = 0.59$; $M_{\text{RS}} = 2.18$, $SD_{\text{RS}} = 0.59$; $M_{\text{IGS}} = 2.39$, $SD_{\text{IGS}} = 0.62$; $F(2, 99) = 7.42$, $p < .001$, part. $\text{Eta}^2 = .13$). Bei SI werden die Akzeptanzunterschiede zwischen Gymnasiallehrkräften und Lehrkräften anderer Schulformen signifikant, allerdings ist die globale Effektstärke geringer als bei VERA ($M_{\text{GYM}} = 2.35$, $SD_{\text{GYM}} = 0.61$; $M_{\text{RS}} = 2.66$, $SD_{\text{RS}} = 0.57$; $M_{\text{IGS}} = 2.61$, $SD_{\text{IGS}} = 0.53$; $F(2, 790) = 22.31$, $p < .001$, part. $\text{Eta}^2 = .05$). Zudem fühlen sich Gymnasiallehrkräfte durch VERA-Ergebnisse tendenziell weniger stark in ihrer Unterrichtsweise bestätigt als andere Lehrkräfte ($M_{\text{GYM}} = 2.44$, $SD_{\text{GYM}} = 0.71$; $M_{\text{RS}} = 2.79$, $SD_{\text{RS}} = 0.57$; $M_{\text{IGS}} = 2.57$, $SD_{\text{IGS}} = .75$; $F(2, 99) = 1.94$, $p = .15$, part. $\text{Eta}^2 = .04$). Bei SI werden die Unterschiede zwischen Gymnasiallehrkräften und den Lehrkräften anderer Schulformen signifikant. Die globale Effektstärke ist jedoch klein und wiederum geringer als bei VERA ($M_{\text{GYM}} = 2.58$, $SD_{\text{GYM}} = 0.87$, $M_{\text{RS}} = 2.84$, $SD_{\text{RS}} = 0.80$, $M_{\text{IGS}} = 2.82$, $SD_{\text{IGS}} = 0.84$; $F(2, 790) = 8.03$, $p < .001$, part. $\text{Eta}^2 = .02$). Die Akzeptanzunterschiede zwischen Gymnasial-

lehrkräften und anderen Lehrkräften werden bei VERA und bei SI unter Einbezug der empfundenen Bestätigung durch die Evaluationsergebnisse kleiner.

Diskussion

Zusammenfassung und Interpretation der Ergebnisse

In der aktuellen Studie sollten Zusammenhangsstrukturen und Wirkmechanismen untersucht werden, die Hinweise liefern könnten, warum Schulinspektionen trotz tendenziell höherer Akzeptanz und Ergebnisrezeption als VERA dennoch häufig nicht zu mehr Unterrichtsentwicklung führen. Stringente kausale Schlussfolgerungen waren aufgrund des Designs aber nicht möglich. In den Hypothesen 1 und 2 der Studie wurde angenommen, dass Akzeptanz und Auseinandersetzung mit Ergebnissen bei Schulinspektionen womöglich schwächere Prädiktoren für unterrichtliches Anschlusshandeln sind als bei VERA. Hypothese 1 konnte bestätigt werden: Akzeptanz klärt bei VERA ca. 14 % der Varianz im persönlichen Anschlusshandeln im Unterricht auf, bei Schulinspektionen sind es nur 4 %. In den Modellen von Helmke (2012) und von Ehren et al. (2013) ist Akzeptanz eine positive Voraussetzung für Anschlusshandeln. Die niedrigere Einflussstärke von Akzeptanz auf Unterrichtsveränderungen bei Schulinspektionen im Vergleich zu VERA könnte somit, formal betrachtet, bedingen, dass sich die tendenziell bessere Akzeptanz von Schulinspektio-

nen nicht in mehr Unterrichtsveränderungen im Vergleich zu VERA niederschlägt. Hypothese 2, dass die Auseinandersetzung mit den Evaluationsergebnissen bei Schulinspektionen ein schwächerer Prädiktor für unterrichtliches Anschlusshandeln ist als bei VERA, konnte hingegen nicht bestätigt werden. Sowohl bei VERA als auch bei Schulinspektionen hat die Dauer der Ergebnisauswertung einen eher schwachen Einfluss auf das Anschlusshandeln im Unterricht. Die empirisch häufiger berichtete intensivere Rezeption von Inspektions- im Vergleich zu VERA-Ergebnissen (z. B. Demski et al., 2012) könnte somit deshalb nicht zu mehr Änderungen in der Unterrichtsweise nach Schulinspektionen führen, da sich die Ergebnisauswertung generell als ein eher schwacher Prädiktor für unterrichtliches Anschlusshandeln gezeigt hat. Allerdings ist diese Schlussfolgerung nur mit Vorsicht zu ziehen, da sie in Widerspruch zu anderen Studien steht, bei denen die Auseinandersetzung mit Ergebnissen ein wichtiger Prädiktor für Anschlusshandeln ist (z. B. Demski, 2017; Wurster et al., 2016). Der schwache Einfluss von Ergebnisauswertung auf Anschlusshandeln könnte auch durch die Single-Item-Messungen der Ergebnisauswertung in der aktuellen Studie mitbedingt sein, die womöglich das Konstrukt nicht genügend reliabel erfasst haben. Eine sinnvolle Ergänzung zu einer einfachen Erhebung der Ergebnisauswertung durch Zeit könnten eventuell konkretere qualitative Informationen sein.

Praktisch betrachtet, könnten sich die tendenziell besseren Voraussetzungen von Schulinspektionen im Vergleich zu VERA deshalb nicht in mehr Unterrichtsveränderungen niederschlagen, da in Inspektionsberichten nur aggregierte Unterrichtsbeobachtungsergebnisse zurückgemeldet werden. Dadurch könnten sich Lehrkräfte häufiger in ihrer Unterrichtsweise bestätigt sehen, da sie schwächere Inspektionsergebnisse auch auf die schlechte Unterrichtsqualität anderer Lehrkräfte attribuieren können (*self-serving attributional bias*,

s. Mezulis et al., 2004). Das könnte zu einer geringeren Änderungsmotivation nach Schulinspektionen führen. Schwächere VERA-Ergebnisse können zwar auch auf die mangelnde Kompetenz voriger Klassenlehrkräfte oder andere externe Ursachen zurückgeführt werden (Bonsen et al., 2006), allerdings können sie im Vergleich zu den Inspektionsergebnissen von der Lehrkraft eindeutiger auf sich bezogen werden und dadurch womöglich eher Änderungsmotivationen hervorrufen. Hypothese 3, dass Lehrkräfte sich mehr durch Inspektionsergebnisse als durch VERA-Ergebnisse in ihrer Unterrichtsweise bestätigt sehen, konnte bestätigt werden. Allerdings führt eine höhere wahrgenommene Bestätigung der eigenen Unterrichtsweise bei VERA zu mehr Unterrichtsveränderungen. Im multiplen Modell ist dieser Zusammenhang allerdings nicht signifikant. Bei Schulinspektionen besteht hingegen kaum ein Zusammenhang zwischen empfundener Bestätigung und Anschlusshandeln. Hypothese 4, dass eine höhere wahrgenommene Bestätigung zu weniger Anschlusshandeln führt, konnte somit nicht bestätigt werden. Bei VERA nehmen Lehrkräfte, die sich durch die Ergebnisse bestätigt fühlen, tendenziell mehr und, umgekehrt betrachtet, Lehrkräfte, die sich durch die Ergebnisse weniger bestätigt fühlen, tendenziell weniger Unterrichtsveränderungen vor. Dieses Ergebnis könnte dadurch begründet sein, dass Lehrkräfte, die sich durch die Ergebnisrückmeldungen nicht bestätigt und vielleicht auch kritisiert fühlen, zunächst mehr selbstwertregulierende als handlungsrelevante Kognitionen haben, wodurch sie die Ergebnisse womöglich eher ablehnen, weniger stark rezipieren und keine Handlungen aus den Ergebnissen ableiten. Dies könnte bei den individualisierten VERA-Rückmeldungen tendenziell stärker der Fall sein, bei denen die Lehrkräfte die Ergebnisse eindeutiger auf sich selbst attribuieren können. Ein negatives Feedback könnte hier zudem die intrinsische Motivation von Lehrkräften verringern (Weidinger, Spinath & Steinmayr, 2016). Bei den aggregierten Unterrichtsbeobachtungsergeb-

nissen könnte dieser Effekt bei Lehrkräften etwas abgemildert sein, worauf der niedrige Zusammenhang zwischen empfundener Bestätigung und Anschlusshandeln hinweist. Bei diesem Feedbackformat betrifft eine schwächere, nicht bestätigende Bewertung nicht nur die Lehrkräfte selbst, sondern das gesamte Schulkollektiv, wodurch sie in ihrem individuellen Selbstwert weniger betroffen werden. Lehrkräfte fühlen sich zwar durch Inspektionsergebnisse mehr bestätigt als durch VERA-Ergebnisse, da dies aber zu eher mehr Unterrichtsveränderungen führt, scheint dieser Faktor eher weniger dafür verantwortlich zu sein, dass die tendenziell bessere Akzeptanz und die intensivere Ergebnisauswertung bei Schulinspektionen im Vergleich zu VERA nicht zu mehr Anschlusshandeln im Unterricht führen.

Eine weitere Erklärungsmöglichkeit könnte darin bestehen, dass die Schulinspektion im Vergleich zu VERA mehr als ein Instrument zur Schulentwicklung und zur Verbesserung der Schulorganisation denn als ein Instrument zur Unterrichtsentwicklung angesehen wird. Darauf deuten auch die Befunde von Demski (2017) hin, bei denen vor allem Schulleitungen angaben, die Inspektionsergebnisse zu rezipieren und für ihre Arbeit, die auch aus administrativen und organisatorischen Tätigkeiten besteht, zu nutzen. Daher könnte bei der Rezeption und Nutzung der Inspektionsergebnisse mehr die Schulentwicklungs- und weniger die Unterrichtsentwicklungsebene im Vordergrund stehen.

Implikationen

In der aktuellen Studie hat sich, wie auch beispielsweise in der Untersuchung von Wurster et al. (2016), gezeigt, dass Akzeptanz bei VERA und bei SI ein wichtiger Prädiktor für weiteres Anschlusshandeln ist. Zudem fällt auf, dass Gymnasiallehrkräfte VERA und SI weniger gut akzeptieren als andere Lehrkräfte. Als praktische Implikation folgt daraus, dass es entwicklungsförderlich ist, vor allem bei Gymnasial-

lehrkräften, akzeptanzfördernde Maßnahmen durchzuführen. Die Einstellung von Lehrkräften gegenüber den Verfahren könnte beispielsweise durch organisatorisch entlastende Maßnahmen, wie die computerbasierte Durchführung von VERA, verbessert werden. Ebenso brauchen Lehrkräfte eventuell mehr Informationen dazu, wie sie die Ergebnisse der Verfahren in konkrete Unterrichtsverbesserungen umsetzen können, damit ihnen die Verfahren sinnhafter erscheinen. Allerdings sind die Akzeptanzunterschiede zwischen Gymnasial- und anderen Lehrkräften womöglich dadurch mitbedingt, dass sich Gymnasiallehrkräfte weniger durch VERA- und SI-Ergebnisse bestätigt fühlen als andere Lehrkräfte. Ein Gefühl der Nicht-Bestätigung durch VERA-Ergebnisse führt zudem tendenziell zu weniger Anschlusshandeln. Daher wäre es sinnvoll, Lehrkräfte schon während ihrer Ausbildung verstärkt darin zu schulen, auch mit negativem Feedback konstruktiv umzugehen und schwächere Ergebnisse als eine Möglichkeit zur Verbesserung anzusehen. Eine Sensibilisierung der Schulleitungen für diese Thematik könnte dazu beitragen, dass Lehrkräfte mit schwächeren Evaluationsergebnissen auch motivational und emotional im Rezeptionsprozess bei Bedarf noch besser unterstützt werden könnten. Zudem wäre es sinnvoll, der Frage nachzugehen, warum sich Gymnasiallehrkräfte weniger häufig durch Evaluationsergebnisse bestätigt fühlen als andere Lehrkräfte.

Schulinspektionen sind ein Instrument der Organisationsdiagnose, und die Schulleitung ist der Hauptadressat des Inspektionsberichts (Wurster, Feldhoff & Gärtner, 2016). Dadurch entsteht vielleicht beim schulischen Personal die Wahrnehmung, dass Schulinspektionen mehr der Schul- als der Unterrichtsentwicklung dienen (Demski et al., 2012). Damit Inspektionsergebnisse mehr zur Unterrichtsentwicklung genutzt werden, wäre es womöglich sinnvoll, Lehrkräfte in den Inspektionsprozess stärker miteinzubeziehen und ihnen auf Wunsch auch individualisierte Rückmeldungen zu den Unterrichtsbeobachtungen zu geben. Aus diesen

Rückmeldungen könnten Lehrkräfte konkrete Anregungen zur Verbesserung ihres Unterrichts ableiten.

Methodische Kritik

An der aktuellen Studie sind die Single-Item-Messungen verschiedener Variablen als eine Schwäche zu nennen. Da mit dem Online-Fragebogen noch weitere Fragestellungen beantwortet werden sollten, waren Single-Item-Messungen teilweise notwendig, um den Fragebogen möglichst ökonomisch zu halten, damit die Teilnahme für Lehrkräfte attraktiver wurde. Single-Item-Messungen von konkreten Sachverhalten und von globaleren Konstrukten wie der individuellen Persönlichkeitsstruktur haben sich in empirischen Studien zwar als relativ reliabel und in ihrer prädiktiven Validität als vergleichbar mit Multiple-Item-Messungen erwiesen (Bergkvist & Rossiter, 2007; Rammstedt, Koch, Borg & Reitz, 2004). In der aktuellen Studie sind verminderte Reliabilitäten der Variablen jedoch nicht ganz auszuschließen. Dadurch könnten die teilweise geringeren Varianzaufklärungen mitbedingt sein. Zudem war es durch die Single-Item-Messungen nicht möglich, Rezeption und konkrete Anschlusshandlungen differenzierter zu erfassen.

Ausblick

In der vorliegenden Studie wurde ein kontrastierender Vergleich von Zusammenhangsstrukturen bei VERA und bei SI durchgeführt, was in der bisherigen Forschungspraxis selten erfolgt ist. Jedoch konnten einige Variablen, die aufgrund theoretischer Modelle (z. B. Helmke, 2012) ebenfalls Einflussfaktoren für Anschluss handeln sind, nicht berücksichtigt werden. Auf individueller Ebene wären da beispielsweise das unterrichtete Fach der Lehrkräfte oder die empfundene individuelle Nützlichkeit von VERA und SI zu nennen. Die individuelle Nützlichkeit erfasst die von den Lehrkräften wahrgenommene Nützlichkeit von VERA und SI für die eigene pädagogische Arbeit (siehe z. B.

Wurster, Feldhoff & Gärtner, 2016). Die hier erhobene allgemeine Nützlichkeit hingegen beschreibt den empfundenen Nutzen der Verfahren für Schul- und Unterrichtsentwicklung im Allgemeinen. Auf Organisationsebene könnten ebenfalls Einflussfaktoren wie die Kooperation im Kollegium miteinbezogen werden. Durch Untersuchungen mit weiteren Einflussfaktoren könnten die in der vorliegenden Studie gefundenen Zusammenhänge noch validiert werden. Zudem sollte in weiteren Untersuchungen die Rezeption von Evaluationsergebnissen differenzierter untersucht werden. Da in der aktuellen Studie keine Aussagen über konkrete Anschluss handlungen gemacht werden konnten, wäre es sinnvoll, Unterrichtsveränderungen in weiteren Untersuchungen ebenfalls differenzierter zu erfassen. Schulinspektionen führen im Vergleich zu VERA vermutlich zu mehr typischen Schulentwicklungsaktivitäten (z. B. Arbeit am Schulprogramm). Es wäre daher sinnvoll, auch Zusammenhangsanalysen von Akzeptanz, Rezeption und Schulentwicklungsmaßnahmen bei Schulleitungen und auf Schulebene insgesamt durchzuführen. Bei Analysen mit diesem Kriterium könnten Akzeptanz und Rezeption bei SI stärkere Prädiktoren sein als bei dem Handeln auf Unterrichtsebene. Es ist zudem zu betonen, dass die hier gefundenen Ergebnisse sich zunächst auf den niedersächsischen Kontext von VERA und Schulinspektionen beschränken und durch Untersuchungen in anderen Bundesländern mit anderen Durchführungs- und Rückmeldespezifika validiert werden sollten.

Dennoch kann durch die aktuelle Studie die bislang spärliche Vergleichsforschung zwischen verschiedenen externen Evaluationsverfahren erweitert werden. Die öfters lediglich diagnostizierte mangelnde Effektivität von Schulinspektionen wurde näher beleuchtet, indem ein möglicher Wirkmechanismus durch wahrgenommene Bestätigung analysiert wurde, der in der bisherigen Forschung kaum fokussiert wurde. Aus den Ergebnissen lassen sich praktische Implikationen und Anregungen für weitere Forschung ableiten.

Literatur

- Altrichter, H. & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School effectiveness and school improvement*, 26(1), 32–56. <https://doi.org/10.1080/09243453.2014.927369>
- Altrichter, H. & Kemethofer, D. (2016). Stichwort: Schulinspektion. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-016-0688-0>
- Bach, A., Würster, S., Thillmann, K., Pant, H. A. & Thiel, F. (2014). Vergleichsarbeiten und schulische Personalentwicklung – Ausmaß und Voraussetzungen der Datennutzung. *Zeitschrift für Erziehungswissenschaft*, 17, 61–84. <https://doi.org/10.1007/s11618-014-0486-5>
- Bellmann, J., Schweizer, S. & Thiel, C. (2016). Nebenfolgen Neuer Steuerung unter Bedingungen von „low-stakes“ und „no-stakes“ – Qualitative und quantitative Befunde einer Untersuchung in vier Bundesländern. In Bundesministerium für Bildung und Forschung (Hrsg.), *Steuerung im Bildungswesen. Implementation und Wirkung neuer Steuerungsinstrumente im Schulwesen* (Bildungsforschung, Bd. 43, S. 208–237). Berlin: BMBF
- Bergkvist, L. & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, Vol. XLIV, 175–184. <https://doi.org/10.1509/jmkr.44.2.175>
- Böhm-Kasper, O. & Selders, O. (2013). „Schulinspektionen sollten regelmäßig durchgeführt werden?“ Ländervergleichende Analyse der Wahrnehmung und Akzeptanz von Schulinspektionsverfahren. *DDS – Die Deutsche Schule, Beiheft 12*, 121–153
- Bonsen, M., Büchter, A. & Peek, R. (2006). Datengestützte Schul- und Unterrichtsentwicklung – Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen und Lehrer. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Band 14. Daten, Beispiele und Perspektiven* (S.125–148). Weinheim/München: Juventa
- Chapman, C. (2001). Changing classrooms through inspection. *School Leadership & Management*, 21(1), 59–73. <https://doi.org/10.1080/13632430120033045>
- Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. (1989). User acceptance of computer technology. A comparison of two theoretical models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>
- Dedering, K. (2016). Der Umgang von Schulen mit Schulinspektionen – Zum Projekt „Externe Evaluation/ Schulinspektion und Qualitätsentwicklung von Schulen.“ In Bundesministerium für Bildung und Forschung (Hrsg.), *Steuerung im Bildungswesen. Implementation und Wirkung neuer Steuerungsinstrumente im Schulwesen* (Bildungsforschung, Bd. 43, S. 86–109). Berlin: BMBF
- Demski, D. (2017). *Evidenzbasierte Schulentwicklung. Empirische Analyse eines Steuerungsparadigmas*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-18078-2>
- Demski, D., Rosenbusch, C., van Ackeren, I., Clausen, M. & Schmidt, U. (2012). Steuerung von Schule durch evidenzbasierte Einsicht? Konzeption und erste Befunde des Forschungsvverbundes EviS. In S. Hornberg & M. Parreira do Amaral (Hrsg.), *Deregulierung im Bildungswesen* (S. 131–150). Münster: Waxmann
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht. Bericht über die Voruntersuchung in Bayern*. Zugriff am 10.1.2018 unter <http://www.quassu.net/BerichtI.pdf>
- 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), 3–43. <https://doi.org/10.1007/s11092-012-9156-4>
- Ehren, M. C. M. & Visscher, A. J. (2008). The relationships between school inspections, school characteristics and school improvement. *British Journal of Educational Studies*, 56(2), 205–227. <https://doi.org/10.1111/j.1467-8527.2008.00400.x>
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz
- Groß Ophoff, J. (2013). *Lernstandserhebungen: Reflexion und Nutzung*. Münster: Waxmann
- Groß Ophoff, J., Koch, U. & Hosenfeld, I. (2018). Trends in der Akzeptanz und Auseinandersetzung mit Rückmeldungen. In J. Zuber, H. Altrichter & M. Heinrich (Hrsg.), *Bildungsstandards zwischen Politik und schulischem Alltag* (205–228). Wiesbaden: VS Verlag für Sozialwissenschaften
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. Aufl.). Seelze: Klett-Kallmeyer
- Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation. Überblick zum Stand der Forschung. *Zeitschrift für Erziehungswissenschaft*, 14, 259–282. <https://doi.org/10.1007/s11618-011-0204-5>
- Kohler, B. (2004). Zur Rezeption externer Evaluation durch Lehrkräfte, Eltern sowie Beamte der Schulaufsicht. *Empirische Pädagogik*, 18(1), 18–39
- Kühle, B. & van Ackeren, I. (2012). Wirkungen externer Evaluationsformen für eine evidenzbasierte Schul- und Unterrichtsentwicklung. In M. Ratermann & S. Stöbe-Blossey (Hrsg.), *Governance von Schul- und Elementarbildung – Vergleichende Betrachtungen und Ansätze der Vernetzung* (S. 45–62). Wiesbaden: Springer VS Verlag. https://doi.org/10.1007/978-3-531-94241-4_3
- Kultusministerkonferenz (2016). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Köln, Berlin, Bonn: Wolters Kluwer & KMK
- Mezulis, A. H., Abramson, L. Y., Hyde, J. S. & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, 130(5), 711–747. <https://doi.org/10.1037/0033-2909.130.5.711>
- Niedersächsisches Kultusministerium (2014). *Schulische Qualitätsentwicklung in Niedersachsen*. Zugriff am 19.2.2019 unter www.mk.niedersachsen.de/download/5154
- Niedersächsisches Landesinstitut für schulische Qualitätsentwicklung (2014). *Zum Entwicklungsauftrag der Schulinspektion*. Zugriff am 27.8.2018 unter http://www.nibis.de/nibis3/uploads/2/nlq-a2/files/Materialien_ABS/Zum_Entwicklungsauftrag_der_Schulinspektion_November2014.pdf
- Penninckx, M., Vanhoof, J., De Maeyer, S. & van Petegem, P. (2016). Effects and side effects of Flemish school inspection. *Educational Management, Administration & Leadership*, 44(5), 728–744. <https://doi.org/10.1177/1741143215570305>

- Rammstedt, B., Koch, K., Borg, I. & Reitz, T. (2004). Entwicklung und Validierung einer Kurzskaala für die Messung der Big-Five-Persönlichkeitsdimensionen in Umfragen. *ZUMA Nachrichten* 28(55), 5–28
- Reezigt, G. J. & Creemers, B. P. M. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, 16(4), 407–424. <https://doi.org/10.1080/09243450500235200>
- Weidinger, A. F., Spinath, B. & Steinmayr, R. (2016). Why does intrinsic motivation decline following negative feedback? The mediating role of ability self-concept and its moderation by goal orientations. *Learning and Individual Differences*, 47, 117–128. <https://doi.org/10.1016/j.lindif.2016.01.003>
- Wurster, S., Bach, A., Schliesing, A., Thillmann, K., Pant, H.A. & Thiel, F. (2016). Schulen als Steuerungsakteure im Bildungssystem – datenbasierte Schul- und Unterrichtsentwicklung aus der Perspektive von Schulleitungen, Fachkonferenzleitungen und Lehrkräften. In Bundesministerium für Bildung und Forschung (Hrsg.), *Steuerung im Bildungswesen. Implementation und Wirkung neuer Steuerungsinstrumente im Schulwesen* (Bildungsforschung, Bd. 43, S. 178–207). Berlin: BMBF
- Wurster, S., Feldhoff, T. & Gärtner, H. (2016). Führen verschiedene Inspektionskonzepte zu unterschiedlicher Akzeptanz und Verwendung der Ergebnisse durch Schulleitungen und Lehrkräfte? *Zeitschrift für Erziehungswissenschaft*, 19, 557–575. <https://doi.org/10.1007/s11618-016-0693-3>
- Wurster, S., Richter, D. & Lenski, A. (2017). Datenbasierte Unterrichtsentwicklung durch Lehrkräfte und Zusammenhänge zur Schülerleistung. *Zeitschrift für Erziehungswissenschaft*. Published online. <https://doi.org/10.1007/s11618-017-0759-x>
- 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, Beiheft 12*, 19–50.

Dr. Inga Wagner

Prof. Dr. Ingmar Hosenfeld

Dr. Michael Zimmer-Müller

Zentrum für Empirische Pädagogische Forschung

Bürgerstr. 23

76829 Landau

E-Mail: iwagner@zefp.uni-landau.de

hosenfeld@zefp.uni-landau.de

zimmer-mueller@zefp.uni-landau.de