

Sebastian Wurster & Dirk Richter

Nutzung von Schülerleistungsdaten aus Vergleichsarbeiten und zentralen Abschlussprüfungen für Unterrichtsentwicklung in Brandenburger Fachkonferenzen

Zusammenfassung

Ergebnisse von Vergleichsarbeiten (VERA) und zentralen Abschlussprüfungen bieten eine bedeutsame Datengrundlage zur Unterrichtsentwicklung. Innerhalb der Schule stellt die Fachkonferenz eine Schlüsselrolle zur Verarbeitung der Ergebnisse und zur Einleitung von Entwicklungsmaßnahmen dar. In diesem Beitrag wird deshalb am Beispiel des Bundeslandes Brandenburgs untersucht, inwieweit Ergebnisse von VERA und den Prüfungen zum Mittleren Schulabschluss (MSA) in Fachkonferenzen (N = 98) für Unterrichtsentwicklung genutzt werden, ob es Unterschiede in der Einschätzung und Nutzung der beiden Testsysteme gibt, ob beide Datenquellen gemeinsam verwendet werden und welche Faktoren (Einstellungen und Organisationsmerkmale) die Nutzung der Ergebnisse vorherzusagen. VERA- und MSA-Ergebnisse werden in der Mehrzahl der Fachkonferenzen für die Weiterentwicklung von Aufgaben verwendet. Die Initiierung datengestützter Unterrichtsentwicklung erfolgt am häufigsten auf Basis beider Datenquellen gemeinsam. Dennoch zeigt sich, dass MSA-Ergebnisse im Vergleich zu VERA häufiger Ausgangspunkt für Entwicklungsaktivitäten sind und der MSA als Datengrundlage z. B. nützlicher wahrgenommen wird. Bedeutsame Prädiktoren für die Nutzung der VERA-Ergebnisse sind die eingeschätzte Nützlichkeit und die Ausrichtung des Curriculums an den Bildungsstandards, wohingegen die MSA-Nutzung primär mit dem Ausmaß kollegialer Unterrichtsentwicklung zusammenhängt.

Dr. Sebastian Wurster (corresponding author), Institut für Erziehungswissenschaft, Johannes Gutenberg Universität Mainz, Jakob-Welder-Weg 12, 55128 Mainz & Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin
E-Mail: wurster@uni-mainz.de

Prof. Dr. Dirk Richter, Department Erziehungswissenschaft, Universität Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam & Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin
E-Mail: dirk.richter@uni-potsdam.de

Schlagworte

Evaluation; Vergleichsarbeiten; Zentrale Abschlussprüfungen; Datennutzung; Fachkonferenz

Use of student performance data of state-wide comparison tests and centralized examinations for instructional improvement in Brandenburg

Abstract

Results of state-wide comparison tests (VERA) and centralized examinations provide an important data base for improving the quality of instruction. Within the school, the conference of teachers of the individual departments plays a key role in processing the test results and in initiating activities that lead to instructional change. This study investigates to what extent the results of VERA and the intermediate school-leaving certificate (MSA) are used in teacher conferences in the federal state Brandenburg (N = 98). We further examined if the use of data differs between the two tests, if teachers use the data of both tests simultaneously, and what factors (attitudes and organizational characteristics) predict the data use. VERA and MSA results are used in the majority of teacher conferences for the improvement of class assignments. The initiation of data-driven instructional improvement is carried out most often based on both data sources. Nevertheless, our findings show that MSA results were more frequently used and were considered more useful than the results of VERA. Significant predictors for data use of VERA results are perceived usefulness and whether the school's curriculum is based on national educational standards. In contrast, data use of MSA results is associated primarily with the extent of collegial instructional improvement.

Keywords

Evaluation; State-wide comparison tests; Centralized examinations; Data use; Data-based decision making

1. Einleitung

Vergleichsarbeiten (VERA) und zentrale Abschlussprüfungen wurden in der vergangenen Dekade im Zuge einer Reorganisation der Steuerung im deutschen Schulsystem als Verfahren der Qualitätssicherung und -entwicklung im Bildungswesen bundesweit eingeführt bzw. haben an Bedeutung gewonnen (Altrichter & Maag Merki, 2010). Zentrale Abschlussprüfungen sind ein Verfahren zur landesweit standardisierten Zertifizierung von Schülerleistungen. Sie verfügen in einigen Bundesländern zwar bereits über eine längere Tradition, allerdings fand erst ab 2003 (Mittlerer Schulabschluss) bzw. 2005 (Abitur) eine Überführung der Abschlussprüfungsmodalitäten von dezentralen in zentrale Prüfungen in fast al-

len Bundesländern statt (Klein, Kühn, van Ackeren & Block, 2009; Kühn, 2013). Parallel ist VERA in der Bundesrepublik als ein neues Evaluationsverfahren flächendeckend eingeführt worden, um ein auf Schülerleistungsdaten basierendes Verfahren zur datengestützten Schul- und Unterrichtsentwicklung bereitzustellen (KMK, 2006).

Ergebnisse aus VERA und zentralen Abschlussprüfungen liefern Schulen Informationen über die Leistungen ihrer Schülerschaft und bieten damit eine wichtige Grundlage zur Reflexion der eigenen Arbeit und datengestützten Unterrichtsentwicklung. Innerhalb der Schule stellt die Fachkonferenz eine Schlüsselrolle zur Auseinandersetzung mit den Ergebnissen und zur Einleitung von Entwicklungsmaßnahmen dar, da in diesem Gremium zentrale fachspezifische Entscheidungen getroffen werden (Maier, Metz, Bohl, Kleinknecht & Schymala, 2012). Datenbasierten Entscheidungen wird auch in der wissenschaftlichen Diskussion ein hoher Stellenwert für die Weiterentwicklung schulischer Qualität beigemessen. Es wird angenommen, dass empirisch fundierte Informationen zu besseren Entscheidungen und damit zu einer Verbesserung schulischer Prozesse führen (Lai & Schildkamp, 2013). Ein Desiderat besteht hinsichtlich der Fragen, ob Daten aus zentralen Abschlussprüfungen generell als Ausgangspunkt für datengestützte Unterrichtsentwicklung genutzt werden und ob datengestützte Unterrichtsentwicklung in Fachkonferenzen auf Basis von Ergebnissen aus VERA *und* zentralen Abschlussprüfungen stattfindet bzw. welche Datengrundlage dafür herangezogen wird. In diesem Beitrag wird deshalb untersucht, inwieweit Ergebnisse aus VERA und zentralen Abschlussprüfungen in Fachkonferenzen für datengestützte Unterrichtsentwicklung genutzt werden, ob es Unterschiede in der Einschätzung und Nutzung der beiden Evaluationsverfahren gibt, ob beide Informationsquellen simultan verwendet werden sowie welche Faktoren für datengestützte Unterrichtsentwicklung in Fachkonferenzen förderlich und hinderlich sind. Die parallele Betrachtung beider Verfahren erlaubt nicht nur die Analyse einer simultanen Nutzung verschiedener Datenquellen, sondern auch eine kontrastierende Analyse dahingehend, ob sich die Datennutzung aufgrund der Ausgestaltung und Implementation der Verfahren unterscheidet. Dieses Vorgehen ermöglicht ebenfalls, die Ausprägung der Einschätzung und Nutzung jeweils in Relation zum anderen Verfahren zu betrachten und somit eine Referenz für die Interpretation zu schaffen.

2. Theoretischer Hintergrund

2.1 Unterrichtsentwicklung in Fachkonferenzen

Fachkonferenzen wird sowohl seitens der Bildungsadministration als auch der Wissenschaft eine bedeutsame Rolle für datengestützte Schul- und Unterrichtsentwicklung zugeschrieben (KMK, 2010; Maier et al., 2012). Zur Unterrichts-

entwicklung gehören nach Helmke (2010) u. a. Aktivitäten zur Verbesserung des Unterrichts durch Veränderung der Lehrmethoden, Stärkung der Kompetenzen der Lehrkräfte und Optimierung des Lehrmaterials. Die Relevanz der Fachkonferenz für die Unterrichtsentwicklung spiegelt sich in ihrem gesetzlichen Auftrag wieder. Sie soll die Lernentwicklung und Leistungsbewertung beobachten, koordinieren und kollektiv Maßnahmen der Qualitätsentwicklung umsetzen (§87 BbgSchulG¹). Zudem erfassen die Ergebnisse aus VERA und dem MSA einen über mehrere Schuljahre andauernden fachspezifischen Kompetenzerwerb der Schülerinnen und Schüler. Daran sind oftmals mehrere Fachkollegen und nicht nur einzelne Lehrkräfte beteiligt (Maier et al., 2012). Außerdem sind Fachbereiche ein geeigneter Adressat für die Ergebnisse, da die Bildungsstandards fachspezifisch formuliert sind. Für die Auswertung von Ergebnissen aus VERA und zentralen Abschlussprüfungen sowie die darauf basierende (fachspezifische) Unterrichtsentwicklung spielen Fachkonferenzen folglich eine zentrale Rolle.

Ein idealtypischer Qualitätsentwicklungskreislauf zur Unterrichtsentwicklung wurde von der KMK (2010) formuliert: Nach der Überprüfung des Lernstands sollen die Ergebnisse in Fachgruppen ausgewertet werden. Es folgt der Austausch zwischen den Lehrkräften über mögliche Ursachen und daraufhin bei Bedarf die gemeinsame Festlegung von Zielen und Maßnahmen. Diese sollen anschließend im Unterricht umgesetzt werden. Im Falle von VERA erfolgt dies in derselben Kohorte von Schülerinnen und Schülern, bei zentralen Abschlussprüfungen in der folgenden Kohorte. Der Kreislauf schließt sich mit der erneuten Überprüfung des Lernstands. Diese Form von Qualitätskreisläufen findet sich in unterschiedlicher Ausgestaltung in der Literatur (z. B. Hellrung & Hartig, 2013).

Ein solcher Kreislauf der datengestützten Unterrichtsentwicklung wird allerdings von einer Reihe von Faktoren beeinflusst. Diese werden z. B. im Modell zu School Performance Feedback Systems (SPFS) von Visscher und Coe (2003) und dem Zyklen-Modell zur pädagogischen Nutzung von Vergleichsarbeiten (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) beschrieben. Den Eigenschaften des Testverfahrens wird im SPFS-Modell eine wichtige Rolle für die Datennutzung zugeschrieben. Visscher und Coe (2003) unterscheiden dabei drei Bereiche von Eigenschaften des Testsystems. Der erste Bereich umfasst die Ziele und Funktionen des SPFS, z. B. Unterrichtsentwicklung. Dieser steht in direktem Zusammenhang mit dem zweiten Bereich, der Ausgestaltung und den Merkmalen des Testsystems, die vor allem die Qualität der erfassten Leistungsdaten und deren Rückmeldung umfassen. Der dritte Bereich beinhaltet die Merkmale der Implementation des SPFS. Hier werden u. a. Unterstützungsmaßnahmen für Schulen subsummiert. Auf Basis dieser Einteilung werden die beiden Verfahren VERA und MSA im folgenden Abschnitt 2.2 und Tabelle 1 vergleichend beschrieben).

1 Brandenburgisches Schulgesetz in der Fassung der Bekanntmachung vom 2. August 2002 (GVBl.I/02, [Nr. 08], S.78), zuletzt geändert durch Artikel 7 des Gesetzes vom 25. Januar 2016 (GVBl.I/16, [Nr. 5]); siehe http://bravors.brandenburg.de/gesetze/bbgschulg_2016 (Zugriff am 20.10.2016).

Der im SPFS-Modell (Visscher & Coe, 2003) beschriebene Zusammenhang der Eigenschaften der Testsysteme mit der Verarbeitung und Verwendung der Testergebnisse kann zudem über die Einschätzung der Testsysteme durch die betroffenen Akteure, wie z. B. Fachkonferenzleitungen, erfolgen. Dies wird im SPFS-Modell (Visscher & Coe, 2003) jedoch nicht explizit formuliert. Die Einschätzungen, wie die Akzeptanz, die wahrgenommene Güte und Nützlichkeit des Verfahrens, werden im Zyklen-Modell von Helmke (2004) aufgegriffen und unter individuellen Bedingungen subsummiert. Zu den individuellen Bedingungen werden zudem z. B. das Vorwissen und die Expertise der Lehrkräfte, aber auch das professionelle Selbstverständnis und die Motivation gezählt. Im Zyklen-Modell (Helmke, 2004) haben die individuellen Bedingungen, neben weiteren Einflussfaktoren, Einfluss auf die Rezeption und Reflexion der Testergebnisse und die Umsetzung (Aktion) und anschließende Evaluation von Entwicklungsaktivitäten.

In einer aktuellen Version des Zyklen-Modells (Hosenfeld & Groß Ophoff, 2007) werden wie auch im SPFS-Modell von Visscher und Coe (2003) zudem schulische Bedingungen als zusätzliche Dimension von Einflussfaktoren auf die Datennutzung beschrieben. Für eine Reihe von Faktoren wird ein positiver Einfluss auf die Ergebniskommunikation und Initiierung von Entwicklungsmaßnahmen angenommen. Dazu gehören z. B. die Innovationsbereitschaft des Kollegiums, das Kooperationsklima sowie die Schaffung von Verbindlichkeit durch die Verankerung im Schulprogramm (Hosenfeld & Groß Ophoff, 2007).

Zur letzten Dimension im Zyklen-Modell (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) zählen schulexterne Faktoren, die Einfluss auf datengestützte Unterrichtsentwicklung haben. Dazu gehört vor allem externe Unterstützung wie z. B. Fortbildungen zur Interpretation der Daten und zur Ableitung von adäquaten Entwicklungsvorhaben (Koch, 2011).

2.2 Vergleichsarbeiten und zentrale Abschlussprüfungen als Referenzsystem für Schülerleistung

Zentrale Abschlussprüfungen und VERA sind zwei unterschiedliche Verfahren, die Informationen zu Schülerleistungen liefern. Diese können zur Qualitätsentwicklung innerhalb von Schulen durch schulische Akteure wie Schulleitungen, Lehrkräfte oder in Gremien wie der Fachkonferenz verwendet werden. Da der Implementation und Ausgestaltung von Evaluationsverfahren ein Einfluss auf die Datennutzung zugeschrieben wird (Visscher & Coe, 2003), werden im Folgenden VERA und zentrale Abschlussprüfungen dahingehend miteinander verglichen (s. Tabelle 1). Der Vergleich wird am Beispiel der Verfahren Mittlerer Schulabschluss (MSA²) und VERA in der achten Jahrgangsstufe im Land Brandenburg durchgeführt, da es in

2 In Brandenburg heißen diese Prüfungen *P10* (Prüfungen in der Jahrgangsstufe 10). Im Text wird zur besseren Verständlichkeit der geläufige Begriff Mittlerer Schulabschluss (MSA) verwendet.

diesem Bundesland deutliche Unterschiede in den bereitgestellten Informationen zur Schülerleistung gibt.

Im Bereich des Schulsystemkontexts (A) zeigen sich zentrale Unterschiede. Im Gegensatz zu VERA, die vor allem der Unterrichtsentwicklung dienen sollen, ist das primäre Ziel des MSA die Zertifizierung von Leistung zum Ende der 10. Klasse (Maag Merki, 2010). Der MSA erfüllt somit eine Selektionsfunktion. Da die Prüfungen benotet werden, sind diese für die Schülerinnen und Schüler mit bedeutsamen Konsequenzen (*high stakes*) verbunden (Zugang zu weiteren Ausbildungsmöglichkeiten und dem Arbeitsmarkt). VERA ist im Gegensatz dazu ein *Low-Stakes*-Testsystem, in dem keine unmittelbaren Konsequenzen für schulische Akteure vorgesehen sind. VERA ist vor allem für die innerschulische Qualitätsentwicklung konzipiert.

Bei den Merkmalen des Testsystems (B) werden Unterschiede, aber auch Gemeinsamkeiten deutlich. Beide Verfahren überprüfen zu einem bestimmten Zeitpunkt basierend auf definierten Referenzrahmen den Leistungsstand der Schülerinnen und Schüler. Für VERA dienen die Bildungsstandards (Köller, 2009) als Referenzrahmen. Dem MSA liegen die Rahmenlehrpläne zugrunde, die sich ebenfalls an den Bildungsstandards orientieren (MBJS, 2008). VERA ist Teil der KMK-Gesamtstrategie (KMK, 2006) und dient dazu, festzustellen, inwiefern die in den Bildungsstandards definierten Kompetenzen bereits ein Jahr (Primarstufe) bzw. zwei Jahre (Sekundarstufe I) früher erreicht wurden. Die Ergebnisse aus VERA 8 sind ein Hinweis auf das spätere Abschneiden beim MSA, da sich die Bildungsstandards für die Sekundarstufe I auf die erwartete Leistung zum Zeitpunkt des MSA beziehen (Graf, Harych, Wendt, Emmrich & Brunner, in Druck; MBK, 2011). VERA bietet somit eine zeitlich vorgelagerte, durch den Bezug auf die Bildungsstandards kriteriale und durch standardisierte Testaufgaben objektive Außenperspektive auf Schülerleistung. Deshalb wird VERA ein hoher Stellenwert als Impulsgeber für Schul- und Unterrichtsentwicklung zugeschrieben (KMK, 2010; Maier et al., 2012; Pant, 2013).

Die Durchführungs- und Auswertungsmodalitäten unterscheiden sich zwischen beiden Verfahren (s. Tabelle 1). Für die Datennutzung bedeutsame Differenzen finden sich bei den Informationen, die schulische Akteure erhalten. So werden bei VERA z. B. Informationen zu einzelnen Aufgaben, Vergleichswerte derselben Schulform für einen sozialen Vergleich und Kompetenzstufen für einen kriterialen Vergleich zurückgemeldet (ISQ, 2012). Die damit verbundenen Analysemöglichkeiten sind deutlich vielfältiger als beim MSA, bei dem keine Rückmeldung an schulische Akteure im klassischen Sinne vorgesehen ist. Es sind nur die von den Schulen selbst eingegebenen Daten in einem Onlineportal einsehbar und die Vergleichsmöglichkeiten deutlich eingeschränkter (nur Landesmittelwerte). Neben den hier am Beispiel Brandenburgs beschriebenen Unterschieden zwischen den Verfahren sind VERA (Maier, 2008) und der MSA (Kühn, 2013) in den Bundesländern jeweils unterschiedlich ausgestaltet. Zum Beispiel wird in Schleswig-Holstein – im Gegensatz zu Brandenburg – explizit auf die Möglichkeit der Nutzung von Ergebnissen aus zentralen Prüfungen

Tabelle 1: Vergleich der Verfahren VERA 8 und Prüfungen zum Mittleren Schulabschluss am Beispiel des Landes Brandenburg

	VERA 8	Mittlerer Schulabschluss
A) Schulsystemkontext		
Primäre Funktion	Schul- und Unterrichtsentwicklung	Zertifizierung und Selektion der Schülerschaft (high stakes)
Rechenschaftspflicht	Schulen, Lehrkräfte, bildungspolitisch Verantwortliche (Pant, 2013)	Schülerschaft (high stakes), Schulen, Lehrkräfte, bildungspolitisch Verantwortliche
B) Merkmale des Testsystems		
Durchführung	Vollerhebung in Klasse 8; z. T. Verpflichtung zur Teilnahme (nicht alle Fächer), jährliche Veränderung der Testinhalte Fächer: Deutsch, Mathematik, 1. Fremdsprache (nur 8. Klasse) Testadministration durch Lehrkräfte	Vollerhebung in 10. Klasse; Verpflichtung zur Teilnahme, Test aller Inhalte der Bildungsstandards Fächer: Deutsch, Mathematik, Englisch, weitere durch Schüler gewählte Fächer Testadministration durch Lehrkräfte
Auswertung	Dezentral durch Lehrkräfte	Dezentral durch Lehrkräfte
Grundlage des Tests	Bildungsstandards	Rahmenlehrpläne/Bildungsstandards
Zeitpunkt des Zugangs zu Informationen	Sofortrückmeldung, Zeitnah (Wochen)	Sofort nach Eingabe, Zeitnah (Wochen)
Ausgestaltung der Informationen	Schul-, Klassen- und Individualrückmeldung: Rückmeldung auf Aufgaben- und Kompetenzebene	Nur durch Schulen selbst eingebene Prüfungsergebnisse
Darstellung der Ergebnisse	Lösungshäufigkeiten, Kompetenzstufen	Noten
Bezugsnormorientierung	Sozial (Schulform als Vergleichsgruppe), kriterial (Kompetenzstufen)	Sozial (Bundesland als Vergleichsgruppe), kriterial (bestanden/nicht bestanden)
C) Merkmale der Implementation		
Unterstützungssystem	Aufgabendatenbank, didaktische Handreichungen, Informationen zur Weiterarbeit, Informationsveranstaltungen	

zur Unterrichtsentwicklung hingewiesen und entsprechende Rückmeldungen werden bereitgestellt (MBK, 2011). Auch in Berlin gibt es Rückmeldungen zum MSA auf Schul-, Fach- und Klassenebene (Hoth & Wesselhöfft, 2011). Hinsichtlich der Ausgestaltung von VERA-Rückmeldungen gibt es ebenfalls Unterschiede in den Inhalten und der Darstellung zwischen Bundesländern. In einigen Bundesländern, wie beispielsweise in Berlin und Brandenburg (ISQ, 2012) oder Nordrhein-Westfalen³, enthalten diese Ergebnisse in Form von an den Bildungsstandards orientierten Kompetenzstufen sowie Lösungshäufigkeiten für einzelne Aufgaben. Im

³ Siehe <http://www.schulentwicklung.nrw.de/lernstand8/lehrerinformationen/ergebnis-rueckmeldung/rueckmeldeebenen/rueckmeldeebenen.html> (Zugriff am 20.03.2016).

Vergleich dazu werden beispielsweise in Thüringen⁴ und Sachsen (SBI, 2012) ausschließlich Lösungshäufigkeiten und keine Kompetenzstufen zurückgemeldet.

Die VERA-Ergebnisse enthalten somit nicht nur mehr nutzbare Informationen, sondern, als ein Merkmal der Implementation (C), auch ein Unterstützungssystem zum Umgang und zur Weiterarbeit mit den Daten. Dazu gehören z. B. didaktische Handreichungen oder Informationsveranstaltungen zum Verfahren (s. Tabelle 1) (ISQ, 2012). Einhergehend mit der primären Funktion ist VERA deutlich stärker an der praktischen Nutzbarkeit für datenbasierte Unterrichtsentwicklung orientiert.

Wie in diesem Vergleich deutlich wurde, unterscheiden sich VERA und der MSA hinsichtlich der im SPFS-Modell (Visscher & Coe, 2003) formulierten, auf das Testsystem bezogenen Einflussfaktoren. Inwieweit die anderen im Zyklen- (Hosenfeld & Groß Ophoff, 2007) und SPFS-Modell (Visscher & Coe, 2003) beschriebenen individuellen, schulischen und schulexternen Bedingungen für den Datennutzungsprozess mit Ergebnissen aus VERA oder zentralen Prüfungen relevant sind und ob der idealtypische Ablauf der Datennutzung stattfindet, wird nachfolgend im Forschungsstand dargestellt.

3. Forschungsstand zur Nutzung von Daten aus Vergleichsarbeiten und zentralen Abschlussprüfungen für Unterrichtsentwicklung

3.1 Vergleichsarbeiten

Der bedeutsamste Ort der Ergebnisreflexion und innerschulischen Kommunikation über VERA sind nach Angaben von Lehrkräften – neben ihren Kollegen – die Fachkonferenzen (Maier, Bohl, Kleinknecht & Metz, 2011). Zwischen 67 % (Maier, 2008) und 91 % (Kühle & Peek, 2007) der befragten Lehrkräfte und Fachkonferenzleitungen gaben an, dass die Ergebnisse in Fachkonferenzen ausgewertet werden. Dabei findet eine Beschäftigung mit den Bildungsstandards und Lehrplänen statt, welche ihnen als Orientierung zur Ergebnisreflexion dienen (Koch, Groß Ophoff, Hosenfeld & Helmke, 2006; Kühle & Peek, 2007). In der überwiegenden Zahl der Schulen (42 % bis 93 %) werden die Ergebnisse nach Angaben von Lehrkräften und Schulleitungen auch in Schul- und Lehrerkonferenzen besprochen (Kühle & Peek, 2007; Maier, 2008; Wurster, Richter, Schliesing & Pant, 2013).

Die Nutzung der Ergebnisse zur Unterrichtsentwicklung fand nach Angaben von Fachkonferenzleitungen vor allem durch eine Konzentration auf Testinhalte (50 %) und weniger häufig durch Intensivierung der Förderung (20 %) oder

4 Siehe https://www.kompetenztest.de/download/kt2011_beispiel_schuelerbericht_mk8.pdf und https://www.kompetenztest.de/download/kt2011_beispiel_ergebnisbericht_ek6.pdf (Zugriff am 20.03.2016).

Veränderung der Unterrichtsmethoden (13 %) statt (Kühle & Peek, 2007). Befragte Lehrkräfte bestätigen die Sicht der Fachkonferenzleitungen: In der Mehrheit der Schulen werden vor allem VERA-Aufgaben geübt, als Basis für Klassenarbeiten sowie Leistungskontrollen eingesetzt und sie dienen der Neuentwicklung eigener Aufgaben. Außerdem werden Unterrichtsmaterialien (weiter-)entwickelt (Groß Ophoff, 2013; Koch et al., 2006; Nachtigall & Jantowski, 2007). Trotz des starken Fokus auf Aufgaben wird kaum eine Verengung des Curriculums auf Testinhalte berichtet (Richter, Böhme, Becker, Stanat & Pant, 2014; Wacker & Kramer, 2012). In einer auf dem IQB-Ländervergleich 2011 beruhenden quantitativen Studie in Grundschulen wird berichtet, dass es im Mittel keine eindeutige Tendenz zu Unterrichtsveränderungen hin zu einer stärkeren Kompetenzorientierung oder Differenzierung gibt (Richter et al., 2014). Die deutliche Streuung der Ausprägungen zeigt jedoch, dass es in unterschiedlichem Maße diesbezügliche Veränderungen des Unterrichts gibt. Dieser Befund wird durch eine qualitative Interviewstudie bestätigt, die die Heterogenität hinsichtlich von Veränderungen im Unterricht zeigt (Diemer & Kuper, 2011).

Betrachtet man die Ebene der Schule, geben Schulleitungen und Lehrkräfte an, dass VERA-Ergebnisse auch für Schulentwicklungsmaßnahmen z. B. im Bereich Fortbildung und Personalentwicklung eingesetzt werden (Bach, Wurster, Thillmann, Pant & Thiel, 2014; Groß Ophoff, 2013; Wurster et al., 2013). Diese Aktivitäten werden jedoch wesentlich seltener genannt als Maßnahmen zur Unterrichtsentwicklung.

Als förderliche und hinderliche Faktoren für die Ergebnisnutzung wurden individuelle Bedingungen, wie Einstellungen schulischer Akteure dem Verfahren gegenüber, untersucht. Lehrkräfte und Schulleitungen akzeptieren VERA überwiegend und nehmen VERA im Mittel als eher nützlich wahr, wobei Schulleitungen in der Tendenz noch positiver eingestellt sind als Lehrkräfte und Fachkonferenzleitungen (Bonsen, Büchter & Peek, 2006; Kühle & Peek, 2007; Maier, 2008; Nachtigall & Hellrung, 2013). Über alle befragten Personengruppen hinweg ist die wahrgenommene Nützlichkeit sehr bedeutsam für die Ableitung und Umsetzung von Maßnahmen (Bonsen et al., 2006). Zudem spielt die Motivation der Lehrkräfte zur Auseinandersetzung mit den Rückmeldungen und die Intensität der Auseinandersetzung dafür eine wichtige Rolle (Groß Ophoff, 2013). Ebenso sind die Innovationsbereitschaft, eine hohe Fortbildungsaktivität sowie eine kriteriale Bezugsnormorientierung (Bezug der Rückmeldung auf die Bildungsstandards) und eine daran orientierte Auswertung wichtige Prädiktoren für die Ergebnisnutzung (Groß Ophoff, 2013; Kühle, 2010). In Bezug auf schulbezogene Merkmale, wie die kollegiale Kooperation, zeigen empirische Analysen uneinheitliche Befunde. Einerseits ist kollegiale Kooperation ein signifikanter Prädiktor für die Besprechung und Verwendung der Ergebnisse (Maier et al., 2012), in anderen Analysen wird dieser Faktor allerdings nicht signifikant (Maier et al., 2011). Andere organisationale Merkmale der Schule, wie das Innovationsklima (Maier et al., 2012) oder die Unterstützung durch die Schulleitung (Bach et al., 2014), hängen positiv mit der Nutzung der Ergebnisse zusammen. Weiterhin scheint auch die

Fachzugehörigkeit für die Ergebnisnutzung von Bedeutung zu sein. Lehrkräfte aus sprachlichen Fächern berichten im Vergleich zu Mathematiklehrkräften eher von Verständnisproblemen und etwas geringer ausgeprägter Ergebnisnutzung (Bonsen et al., 2006; Kühle, 2010; Maier et al., 2011).

3.2 Zentrale Abschlussprüfungen

Die Nutzung von Daten aus zentralen Abschlussprüfungen wurde bislang in internationalen Studien nur selten (Klein, Krüger, Kühn & van Ackeren, 2014) und in Studien aus Deutschland kaum thematisiert. Einen ersten Hinweis auf die Nutzung der Ergebnisse aus zentralen Abschlussprüfungen liefert eine explorative Interviewstudie mit Schulleitungen und Lehrkräften in vier Berliner Schulen (Hoth & Wesselhöfft, 2011). Aktivitäten im Zusammenhang mit dem MSA sind vornehmlich die Prüfungsvorbereitung, wie z. B. gezieltes Arbeiten mit MSA-Materialien und weniger auf die Nachbereitung der Ergebnisse ausgelegt. Die Rückmeldung wird als verständlich und die darin abgebildeten Werte für Vergleichsgruppen (Parallelklassen) als sinnvoll eingeschätzt. Dennoch findet in Fachkonferenzen keine umfangreiche Rezeption und Reflexion der Ergebnisse mit dem Ziel der Ergebnisnutzung für Schul- und Unterrichtsentwicklung statt. Ein Grund dafür ist, dass die Möglichkeit MSA-Daten als Ausgangspunkt für Entwicklungsmaßnahmen zu nutzen, nicht in Betracht gezogen wird. Eine Nutzung erfolgt in Form eines Vergleichs der Prüfungsleistungen über verschiedene Schuljahre und selten durch Konzentration der Reflexion auf Schwächen im Unterricht.

In internationalen Studien wird allerdings berichtet, dass zentrale Abschlussprüfungen in Schulen durchaus als Grundlage für datenbasierte Entwicklungsmaßnahmen genutzt werden (Schildkamp, Rekers-Mombarg & Harms, 2012; Tymms, 1995). Die Rückmeldungen zu Ergebnissen aus zentralen Prüfungen werden in den Niederlanden und England als relevant, valide und reliabel betrachtet. Sie werden von Schulleitungen und Lehrkräften überwiegend rezipiert und diskutiert, wobei Fachkonferenzleitungen dies öfter tun als Lehrkräfte. Die Daten werden vor allem für Monitoringzwecke und Vergleiche mit anderen Referenzgruppen eingesetzt, um die Entwicklung der Schülerschaft zu beobachten (Klein, 2013; Schildkamp et al., 2012). In einer älteren Studie aus England wird beschrieben, dass drei Viertel der befragten Fachkonferenzleitungen und Lehrkräfte aufgrund der Ergebnisse aktiv wurden (z. B. Veränderung der Unterrichtspraxis, Diskussion der Ergebnisse; Tymms, 1995). Datennutzungsaktivitäten werden auch in anderen Studien berichtet. Zentrale Prüfungen werden von niederländischen Lehrkräften sowohl vereinzelt für eine umfassende Reflexion und Entwicklung des eigenen Unterrichts, als auch für unterrichtsbezogene Aktivitäten (Vorbereitung auf Examen, gezieltes Training von Schwächen) verwendet (Schildkamp et al., 2012). Allerdings findet diese Form der Nutzung nicht in allen an der Studie beteiligten Schulen statt und die Anzahl der ergriffenen Maßnahmen ist gering. Auch Klein (2013) berichtet, dass insbesondere irische, aber auch finnische und niederlän-

dische Lehrkräfte die Ergebnisse eher in der Vorbereitung auf die Prüfung, z. B. für mit den Prüfungen vergleichbare Assessments, und in geringerem Umfang für Unterrichtsentwicklung nutzen.

Einflussfaktoren auf die Nutzung im Fachbereich sind zum einen die generelle Kooperation im Kollegium wie auch die eingeschätzte Nützlichkeit (Klein, 2013). Die Einstellung gegenüber dem Verfahren ist bei Lehrkräften generell positiv, wobei Fachunterschiede deutlich wurden. Lehrkräfte aus sprachlichen Fächern sind negativer eingestellt als solche aus dem mathematisch-naturwissenschaftlichen Bereich (Tymms, 1995). Klare Ziele oder Visionen sind ebenfalls förderliche Bedingungen für die Datennutzung, während Schwierigkeiten bei der Ursachenattribution zum Abschneiden der Schülerinnen und Schüler und mangelnde Kompetenz im Umgang mit den Daten in einigen Schulen für die Ableitung von Entwicklungsmaßnahmen hinderlich sind (Schildkamp et al., 2012). Dementsprechend haben ein innerschulischer Experte, Training und Unterstützung bei der Datenanalyse sowie Erfahrung mit dem Verfahren positiven Einfluss auf die Datennutzung (Saunders, 2000). Die Teilnahme an einer Fortbildung zum Thema ist ein sehr bedeutsamer Faktor für die Ergebnisnutzung (Tymms, 1995).

4. Forschungsfragen

Sowohl Ergebnisse aus VERA als auch aus zentralen Prüfungen werden zur Unterrichtsentwicklung herangezogen. In Studien zu zentralen Prüfungen wird beschrieben, dass Anpassungen des Unterrichts eher im Vorfeld der Prüfungen stattfinden und dass die Ergebnisse zum Monitoring der Leistung verwendet werden. Bei VERA zeigte sich hingegen, dass sich die Unterrichtsentwicklung vor allem auf die Testinhalte konzentriert. In Teilen werden von den Schulen jedoch auch Personalentwicklungsaktivitäten ergriffen. Weiterhin wurden, wie bei zentralen Prüfungen, auch Maßnahmen der Einzelförderung berichtet. Hinsichtlich datengestützter Unterrichtsentwicklung in Fachkonferenzen liegt für beide Verfahren allerdings ein Desiderat vor. Außerdem ist bislang unbekannt, inwieweit beide Informationsquellen gemeinsam genutzt werden.

Ausgehend von den Desideraten, der Konstellation der beiden Datenbestände innerhalb von Schulen sowie den Modellannahmen, dass die Ausgestaltung und Implementation des Testsystems sowie schulische und individuelle Bedingungen Einfluss auf die Datennutzung haben (Hosenfeld & Groß Ophoff, 2007; Visscher & Coe, 2003), stellen sich die folgenden Forschungsfragen:

- 1) Wie werden VERA- und MSA-Ergebnisse in Fachkonferenzen verarbeitet, zur Ableitung von Unterrichtsentwicklungsmaßnahmen verwendet und hinsichtlich ihrer Akzeptanz und Nutzbarkeit eingeschätzt?
- 2) Unterscheidet sich die Einschätzung und Nutzung zwischen VERA und dem MSA?

- 3) Werden VERA- und MSA-Ergebnisse in Fachkonferenzen gemeinsam zur Ableitung von Unterrichtsentwicklungsmaßnahmen herangezogen?
- 4) Welche förderlichen oder hinderlichen Bedingungen stehen mit der Nutzung der Ergebnisse für Unterrichtsentwicklung in Fachkonferenzen im Zusammenhang?

Aufgrund des Forschungsstandes wird für die ersten beiden Fragen die Hypothese formuliert, dass beide Datenbestände positiv eingeschätzt und als Ausgangspunkt für Unterrichtsentwicklung verwendet werden. Außerdem wird die Hypothese aufgestellt, dass VERA-Ergebnisse häufiger für Unterrichtsentwicklungsmaßnahmen verwendet werden, da diese für die Nutzung gezielt aufbereitet werden, ein Unterstützungssystem und umfassende Vergleichsmöglichkeiten sowie Hinweise zur Weiterarbeit enthalten. Für die dritte Forschungsfrage wird die Hypothese formuliert, dass beide Datenbestände auch gemeinsam verwendet werden, da sie zu zwei unterschiedlichen Zeitpunkten in der Schulkarriere Schülerleistungen erfassen und VERA-Ergebnisse bereits Hinweise auf das spätere Abschneiden beim MSA geben können. Zur Analyse der vierten Forschungsfrage werden als individuelle Einflussbedingungen die Einschätzungen der beiden Verfahren herangezogen. Als schulische Bedingungen wird beispielhaft die kollegiale Unterrichtsentwicklung analysiert, da bislang der Forschungsstand diesbezüglich heterogen ist. Zudem wird die Ausrichtung des Curriculums an den Bildungsstandards betrachtet. Sie dient als Indikator dafür, ob innerhalb der Schulen die Bildungsstandards als mögliche Bezugsnorm bekannt und anerkannt sind. Eine Anbindung des Curriculums an den Bildungsstandards lässt vermuten, dass die Evaluation des Erreichens der Standards durch VERA und den MSA eher als Grundlage zur Ableitung von Entwicklungsmaßnahmen genutzt wird. Es wird die Hypothese formuliert, dass gemäß den Modellannahmen (Hosenfeld & Groß Ophoff, 2007; Visscher & Coe, 2003) alle genannten individuellen und schulischen Bedingungen mit der Datennutzung in einem positiven Zusammenhang stehen.

5. Methode

5.1 Stichprobe

Die vorliegende Studie basiert auf einer Onlinebefragung Brandenburger Fachkonferenzleitungen der Fächer Deutsch und Mathematik, die im Zeitraum zwischen November 2011 und Februar 2012 stattfand. Die Stichprobe setzt sich ausschließlich aus weiterführenden Schulen zusammen, da nur für diese Gruppe ein Vergleich zwischen VERA und zentralen Abschlussprüfungen, in diesem Fall nach der zehnten Klasse, möglich ist. Insgesamt haben sich $N = 98$ Fachkonferenzleitungen beteiligt (Mathematik: $n = 57$; Deutsch: $n = 41$). Die Fachkonferenzleitungen stammen aus den Schularten Gymnasium ($n = 28$; Rücklauf = 37 %), Gesamtschule mit gymnasialer Oberstufe ($n = 9$; Rücklauf = 45 %) und Oberschule (inkl. Oberschulen mit integriertem Grundschulteil; $n = 27$;

Rücklauf = 23 %). Die Rücklaufquote liegt insgesamt bei 30 %. Die prozentuale Verteilung der Schulformen in der Stichprobe entspricht in etwa der Verteilung der Schularten in Brandenburg. Aufgrund der Fallzahl wird in den Analysen nur zwischen Gymnasium und anderen weiterführenden Schulen unterschieden. Weitere soziodemographische Merkmale zur Beschreibung der Personen liegen aus Gründen des Datenschutzes nicht vor.

5.2 Operationalisierung

In Tabelle 2 sind Informationen über die eingesetzten Instrumente aufgeführt. Im Zentrum der Analysen steht die Nutzung der Ergebnisse aus VERA und MSA für Unterrichtsentwicklungsmaßnahmen. Zur Erfassung der Datennutzung wurden die Fachkonferenzleitungen danach gefragt, welche Maßnahmen nach der Rückmeldung der Ergebnisse von VERA bzw. dem MSA in den Fachkonferenzen initiiert wurden (Items s. Tabelle 3). Neben der Nutzung der Testergebnisse in der Fachkonferenz wurden die Fachkonferenzleitungen nach der eigenen individuellen Nutzung befragt.

Zudem wurde die Auseinandersetzung der Fachkonferenzleitung mit den Ergebnissen erhoben. Die Rezeption und Reflexion der Ergebnisse ist im Zyklen-Modell (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) Voraussetzung für datenbasierte Unterrichtsentwicklung. Die innerschulische Dissemination der Ergebnisse wird in Form der schulinternen Kommunikation der Ergebnisse operationalisiert. Hierfür wurde eine Skala gebildet, in die verschiedene Indikatoren einbezogen wurden. Zum einen integriert die Skala Angaben über die Kommunikation der Ergebnisse im Kollegium und gegenüber wichtigen anderen Stakeholdern wie den Eltern und der Schülerschaft. Zum anderen wird erfasst, ob es für den Austausch geregelte Verfahrensweisen gibt und ob die Schulleitung Wert auf die Ergebnisreflexion legt.

Ebenso wurden individuelle und schulische Bedingungen des Zyklen-Modells (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) operationalisiert. Die Einschätzung der beiden Verfahren, als bedeutsame individuelle Bedingung, ist über die wahrgenommene Nützlichkeit und den Aufwand sowie über die Diagnosegüte erfasst. Um den Nutzen ins Verhältnis zu den Kosten bzw. dem Aufwand setzen zu können, wird der mit einem Verfahren assoziierte Aufwand erfasst, der bislang kaum betrachtet wurde. Die Diagnosegüte gibt an, inwieweit die Ergebnisse des Verfahrens mit den eigenen Erwartungen korrespondieren. Diese Skalen sind Weiterentwicklungen basierend auf der Arbeit von Gärtner und Wurster (2009), die anderen Items stellen Eigenentwicklungen dar.

Tabelle 2: Eingesetzte Instrumente

Item/Konstrukt	Beispielitem	n Items	Cronbachs α (VERA/MSA)
Nutzung ^a	„Konnten Sie an Ihrem Fachbereich die Ergebnisse für die gemeinsame Entwicklung von Unterrichtsentwicklungsmaßnahmen in den folgenden Bereichen nutzen? Entwicklung oder zielgerichtete Veränderung von Aufgaben“	5	Einzelitems bzw. Summenscore .78 / .69 (McDonald, 1999; ω : .77 / .78)
Individuelle Nutzung der VERA/MSA Ergebnisse ^a	„Haben Sie persönlich die Ergebnisse von VERA für eigene Unterrichtsentwicklungsmaßnahmen genutzt?“	1	Einzelitem
Auseinandersetzung der Fachkonferenzleitung mit Ergebnissen ^a	„Haben Sie sich mit den VERA-/MSA-Ergebnissen Ihres Fachbereichs auseinandergesetzt?“	1	Einzelitem
Kommunikation der Ergebnisse ^b	„In unserem Kollegium werden die Ergebnisse intensiv diskutiert.“	5	.87 / .83
Wahrgenommene Nützlichkeit ^b	„Für die pädagogische Arbeit waren die Ergebnisse sehr nützlich.“	3	.90 / .90
Wahrgenommener Aufwand ^b	„Der zeitliche Aufwand ist zu hoch.“	3	.88 / .85
Wahrgenommene Diagnosegüte ^b	„Die Stärken unserer Schule wurden in den Ergebnissen hinreichend abgebildet.“	3	.84 / .80
Kollegiale Unterrichtsentwicklung ^b	„Wie viel Prozent der Lehrkräfte Ihres Fachbereiches praktizieren regelmäßig mit den Kolleginnen oder Kollegen folgende Formen der kollegialen Unterrichtsentwicklung? Durchführung gegenseitiger Unterrichtsbesuche.“	3	.79
Ausrichtung Curriculum an Bildungsstandards ^b	„Das Curriculum ist streng an den Bildungsstandards ausgerichtet.“	1	Einzelitem

^a Antwortmöglichkeiten: dichotom, ja = 1, nein = 0; ^b Antwortmöglichkeiten: 6-stufig, 1 = *Stimme überhaupt nicht* zu bis 6 = *Stimme vollständig* zu.

Ausgehend vom Zyklen-Modell (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) wird als Merkmal der schulischen Bedingungen die kollegiale Unterrichtsentwicklung erfasst. Diese Skala beinhaltet die Durchführung gegenseitiger Unterrichtsbesuche, gemeinsame Unterrichtsvorbereitung und gemeinsame Auswertung von Evaluationsdaten. Als weiteres schulisches Organisationsmerkmal wird die Ausrichtung des schulinternen Curriculums auf die Bildungsstandards erhoben, das einen stärkeren Bezug zu den Leistungsdaten aufweist als das im Modell genannte Schulprogramm. Als Kontrollvariablen dienen die Schulart und das Fach.

5.3 Analysestrategie

Zur Analyse der Forschungsfragen werden deskriptive Maße, *t*-Tests, exakte Fisher-Yates-Tests zur Analyse von Vierfeldertafeln bei kleineren Stichproben sowie für die Prüfung von Unterschieden zwischen den Evaluationsverfahren

die Effektstärken d für metrische und w für dichotome Variablen (Cohen, 1992) berichtet. Zur Analyse der Zusammenhänge für dichotome Items wird das Korrelationsmaß Yules Q eingesetzt, das je nach Stärke und Richtung des Zusammenhangs ähnlich der Pearson-Korrelation Werte im Bereich von -1 bis 1 annehmen kann. Zur Beschreibung des Zusammenhangs zwischen der Nutzung von Leistungsdaten und den Einflussfaktoren werden Pearson-Korrelations- und multiple lineare Regressionsanalysen durchgeführt. Zur Komplexitätsreduktion werden dazu jeweils verfahrensspezifisch die Einzelitems zur Ergebnisnutzung in Fachkonferenzen zu einem Summenscore zusammengefasst. Die so aus einzelnen Nutzungsaktivitäten gebildeten Summenscores dienen als abhängige Variable in den Regressionsanalysen. Aufgrund der Stichprobengröße und fehlender Werte bei einigen unabhängigen Variablen wurden im Vorfeld der Regressionsanalysen multiple Imputationen durchgeführt, um die Daten aller Personen einzubeziehen. Dabei wurde das Predictive-Mean-Matching-Verfahren (Little, 1988) unter Einbezug aller in den Regressionsanalysen verwendeten Variablen angewendet und 50 imputierte Datensätze erstellt. Der Anteil imputierter fehlender Werte beträgt zwischen 4 % (Ausrichtung Curriculum an Bildungsstandards) und 21 % (Kommunikation).

6. Ergebnisse

6.1 Unterrichtsentwicklung in Fachkonferenzen auf Basis von VERA- und MSA-Ergebnissen

VERA- und MSA-Ergebnisse werden in Fachkonferenzen aus Sicht der Fachkonferenzleitungen am häufigsten für die Entwicklung oder zielgerichtete Veränderung von Aufgaben (88 %) genutzt, gefolgt von der Entwicklung neuer Materialien (66 %) und Maßnahmen der Einzelförderung (63 %) (Tabelle 3). Fortbildungsmaßnahmen und neue Unterrichtsmethoden werden hingegen deutlich weniger häufig in Betracht gezogen. Insgesamt werden Ergebnisse aus beiden Verfahren oftmals gemeinsam für Unterrichtsentwicklungsaktivitäten herangezogen bzw. beide Testergebnisse sind nicht Ausgangspunkt von Unterrichtsentwicklung, wie die hohen Zusammenhänge zeigen (Yules Q = .61 bis .88). Es zeigen sich jedoch auch zum Teil deutliche Unterschiede zwischen VERA und dem MSA mit mittleren bis großen Effektstärken ($w = .31$ bis .49). Im Falle datenbasierter Entwicklungsmaßnahmen dominiert zwar die gemeinsame Verwendung beider Datenquellen (Ausnahme: Fortbildungsmaßnahmen), im Vergleich der beiden Verfahren fällt die Nutzung von MSA-Ergebnissen zur Unterrichtsentwicklung in allen Bereichen jedoch höher aus. MSA-Ergebnisse werden auch von den Fachkonferenzleitungen selbst häufiger für die Entwicklung des eigenen Unterrichts eingesetzt, wobei 53 % der Fachkonferenzleitungen beide Datenquellen

heranziehen. Eine Auseinandersetzung mit den Ergebnissen beider Verfahren durch die Fachkonferenzleitung findet in über 90 % der Fachbereiche statt.

Tabelle 3: Nutzung von VERA- und MSA-Ergebnissen für Unterrichtsentwicklung in Fachkonferenzen (in %)

Aktivität	VERA	MSA		p	w	Yules Q
		Nein	Ja			
Persönliche Nutzung durch FKL	Nein	12	34	.001	.32	.89
	Ja	1	53			
Auseinandersetzung der FKL mit Ergebnissen	Nein	3	5	.002	.41	.96
	Ja	1	90			
Entwicklung oder zielgerichtete Veränderung von Aufgaben	Nein	13	37	.001	.31	.88
	Ja	1	50			
Einführung neuer Unterrichtsmethoden	Nein	51	20	< .001	.49	.87
	Ja	4	25			
Einführung oder Entwicklung neuer Materialien	Nein	33	29	< .001	.40	.80
	Ja	4	33			
Entwicklung gezielter Maßnahmen zur Einzelförderung	Nein	37	20	.002	.32	.61
	Ja	13	30			
Konzeption konkreter Fortbildungsmaßnahmen	Nein	61	23	< .001	.38	.83
	Ja	3	13			

Anmerkungen. N = 93, FKL = Fachkonferenzleitung.

6.2 Kommunikation und Akzeptanz von VERA- und MSA-Ergebnissen

Die Ergebnisse des MSA werden innerhalb der Schule umfangreicher kommuniziert als VERA-Ergebnisse (Tabelle 4). In allen erfragten Bereichen zeigt sich zudem, dass der MSA durch Fachkonferenzleitungen deutlich positiver beurteilt wird als VERA. Der Aufwand wird für VERA als deutlich höher wahrgenommen als für den MSA. Zudem werden die Nützlichkeit und Diagnosegüte des Verfahrens für den MSA im Vergleich zu VERA als deutlich höher eingeschätzt.

Tabelle 4: Unterschiede zwischen VERA und MSA in der Kommunikation der Ergebnisse und Einschätzung der Verfahren

	N	M (SD) VERA 8	M (SD) MSA	t	p	d
Kommunikation	89	4.19 (1.01)	4.65 (0.99)	-3.05	.003	-0.46
Aufwand	88	4.27 (1.45)	2.89 (1.26)	6.75	< .001	1.02
Nützlichkeit	89	3.13 (1.22)	4.25 (1.25)	-6.04	< .001	-0.91
Diagnosegüte	87	3.72 (1.22)	4.44 (1.03)	-4.18	< .001	-0.63

6.3 Bedingungen der Nutzung von VERA- und MSA-Ergebnissen für Unterrichtsentwicklung

Zur Analyse der für die Nutzung von VERA- und MSA-Ergebnissen förderlichen und hinderlichen Bedingungen werden zunächst Korrelationen berichtet (Tabelle 5). Verfahrensübergreifend lassen sich mittlere bis starke Zusammenhänge zwischen der Datennutzung auf der einen und der Kommunikation, der wahrgenommenen Nützlichkeit und der kollegialen Unterrichtsentwicklung auf der anderen Seite feststellen. Große Unterschiede zwischen den Verfahren in der Höhe der Zusammenhänge zur Datennutzung bestehen nicht. Tendenziell ist die Korrelation zwischen kollegialer Unterrichtsentwicklung und der MSA-Nutzung etwas höher, während der Zusammenhang zwischen eingeschätzter Nützlichkeit und Diagnosegüte mit der Nutzung von VERA-Ergebnissen im Vergleich zum MSA stärker ist. Unterschiede zwischen den Verfahren bestehen ebenso darin, dass die kollegiale Unterrichtsentwicklung stärker mit den Einschätzungen und der Kommunikation von VERA zusammenhängt, als mit den entsprechenden MSA-Konstrukten. Allerdings sind keine der genannten Unterschiede zwischen den Korrelationskoeffizienten signifikant und entsprechen in der Effektgröße q zum Vergleich von Korrelationskoeffizienten mit einer Range von $q = .08$ bis $q = .17$ eher einer kleinen Effektgröße (Cohen, 1992).

Tabelle 5: Korrelationen Einschätzung der Verfahren und organisationale Bedingungen mit VERA- sowie MSA-Ergebnisnutzung

	VERA						MSA					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
(1) Ergebnisnutzung (Summenscore)	1						1					
(2) Kommunikation	.31	1					.27	1				
(3) Nützlichkeit	.43	.41	1				.36	.44	1			
(4) Aufwand	-.12	-.27	-.63	1			-.20	-.26	-.40	1		
(5) Diagnosegüte	.30	.38	.60	-.36	1		.21	.33	.56	-.41	1	
(6) Kollegiale Unterrichtsentwicklung	.35	.31	.29	-.23	.28	1	.44	.15	.16	-.10	.15	1
(7) Curriculum: Orientierung an Bildungsstandards	.20	.18	-.07	.20	.16	-.01 ^{n.s.}	.19	.19	.13	-.17	.23	-.01 ^{n.s.}

Anmerkung. Alle Korrelationen mit $p < .001$ signifikant, außer entsprechend gekennzeichnete Koeffizienten (^{n.s.}).

Die Vorhersage von Entwicklungsmaßnahmen, die auf den VERA- und MSA-Ergebnissen basieren, wird für jedes Verfahren mit separat durchgeführten Regressionsanalysen untersucht. Als abhängige Variable wird jeweils der Summenscore der Nutzungssitems verwendet. Als Prädiktoren kommt aufgrund der Stichprobengröße nur eine Auswahl der bisher analysierten Variablen zum Einsatz. Bei einer höheren Anzahl an Prädiktoren können selbst mittlere Effekte nicht si-

gnifikant werden. Um die Akzeptanz der Verfahren abzubilden, wurde die Variable Nützlichkeit ausgewählt und auf die Variablen Diagnosegüte und Aufwand verzichtet, da diese hoch mit der Nützlichkeit korrelieren. Zudem ist die Nützlichkeit nach bisherigem Forschungsstand der bedeutsamste Prädiktor (Bonsen et al., 2006). Um Eigenschaften der Schule zu berücksichtigen, wurden die kollegiale Unterrichtsentwicklung und der Bezug auf die Bildungsstandards im schulischen Curriculum sowie als Kontrollvariablen die Schulform und das Fach in die Modelle aufgenommen. Die Ergebnisse sind in Tabelle 6 dargestellt.

Modell 1 enthält nur verfahrensunspezifische schulische Faktoren. Signifikante Prädiktoren in Modell 1 sind bei der MSA-Nutzung die kollegiale Unterrichtsentwicklung und der Bezug zu den Bildungsstandards im Curriculum sowie für die VERA-Nutzung die Schulart und das Fach. In allen weiteren Modellen sind das Fach und die Schulform für die Nutzung von VERA-Ergebnissen jedoch keine signifikanten Prädiktoren mehr. In Modell 2a wird der zusätzliche, jeweils verfahrensspezifische Prädiktor Kommunikation der Ergebnisse sowohl für die VERA- als auch MSA-Nutzung signifikant. Die MSA-Ergebnisnutzung wird in Modell 2a allerdings erneut vor allem durch kollegiale Unterrichtsentwicklung vorhergesagt. Wenn anstelle der Kommunikation der Ergebnisse die wahrgenommene Nützlichkeit als Prädiktor in das Modell aufgenommen wird (Modell 2b), ist dieser ebenfalls für die VERA- und MSA-Nutzung statistisch bedeutsam. Für die VERA-Nutzung ist zudem der Bezug zu den Bildungsstandards im Curriculum ein signifikanter Prädiktor. Für die MSA-Nutzung ist erneut die kollegiale Unterrichtsentwicklung bedeutsam. Im Gesamtmodell mit allen Prädiktoren (Modell 3) sind für die VERA-Nutzung ebenfalls die wahrgenommene Nützlichkeit sowie der Bezug zu den Bildungsstandards im Curriculum signifikante Prädiktoren. Die MSA-Nutzung wird dagegen primär durch kollegiale Unterrichtsentwicklung vorhergesagt.

In der Regressionsanalyse sind die zusätzlichen Prädiktoren (Modell 1 zu Modell 2a/2b) jeweils signifikant. Auch die Hinzunahme von zwei Prädiktoren (Modell 1 zu Modell 3) erklärt, gemäß dem für die gepoolten Ergebnisse der multipel imputierten Datensätze notwendigen Wald-Test (van Buuren & Groothuis-Oudshoorn, 2011), signifikant mehr Varianz (VERA: $p < .001$; MSA: $p = .001$). Insgesamt wird im Modell 3 etwa ein Drittel der Gesamtvarianz durch die Prädiktoren erklärt.

Die Ergebnisse der Regressionsanalysen zeigen, dass die Nutzung von VERA-Ergebnissen vor allem durch die wahrgenommene Nützlichkeit des Verfahrens und der innerschulischen Implementation der Bildungsstandards vorhergesagt wird. Im Gegensatz dazu ist kollegiale Unterrichtsentwicklung ein Merkmal der Schulorganisation, das die Nutzung der MSA-Ergebnisse vorhersagt.

Tabelle 6: Regressionsanalysen mit imputierten Variablen zur Vorhersage der Nutzung von VERA- und MSA-Ergebnissen

	Modell 1				Modell 2a Kommunikation			
	VERA		MSA		VERA		MSA	
	β	SE	β	SE	β	SE	β	SE
Intercept	-.25	.17	-.22	.16	-.30	.17	-.25	.16
Schulart (nicht Gymnasium) ^a	.20*	.10	.26	.20	.30	.20	.23	.19
Fach Deutsch ^a	.32**	.12	.17	.19	.30	.19	.28	.19
Bezug zu Bildungsstandards im Curriculum	.24	.20	.19*	.09	.16	.09	.15	.09
Kollegiale Unterrichtsentwicklung	.26	.20	.44***	.10	.24	.12	.42***	.10
Kommunikation					.25**	.10	.24*	.10
R ²		.19		.28		.24		.34
	Modell 2b Nützlichkeit				Modell 3			
	VERA		MSA		VERA		MSA	
	β	SE	β	SE	β	SE	β	SE
Intercept	-.21	.16	-.24	.16	-.24	.16	-.26	.16
Schulart (nicht Gymnasium) ^a	.19	.19	.18	.20	.23	.19	.19	.19
Fach Deutsch ^a	.23	.18	.32	.20	.25	.18	.35	.19
Bezug zu Bildungsstandards im Curriculum	.21*	.09	.17	.09	.19*	.09	.14	.09
Kollegiale Unterrichtsentwicklung	.19	.11	.44***	.09	.17	.11	.43***	.09
Kommunikation					.12	.10	.17	.11
Nützlichkeit	.40***	.10	.26*	.10	.35***	.10	.20	.11
R ²		.33		.35		.34		.37

Anmerkungen. N = 98; Abhängige Variable = Summenscore Nutzung.

^aRegressionskoeffizienten sind standardisiert mit Ausnahme dichotomer Variablen.

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

7. Diskussion

In diesem Beitrag wurde untersucht, wie VERA- und MSA-Ergebnisse in Brandenburger Fachkonferenzen hinsichtlich verschiedener Aspekte bewertet, schulintern verarbeitet und für die Weiterentwicklung des Unterrichts genutzt werden. Des Weiteren haben wir in den Blick genommen, welche lehrerseitigen und schulischen Faktoren mit der Ergebnisnutzung zusammenhängen. Die parallele Betrachtung beider Verfahren ermöglicht einerseits die Analyse einer simultanen Nutzung beider Datenquellen und andererseits eine Überprüfung, ob – neben anderen Faktoren – die Ausgestaltung und Implementation der Verfahren im Sinne einer kontrastiven Analyse mit datenbasierter Unterrichtsentwicklung zusammenhängen.

Bezüglich der ersten Forschungsfrage berichten die befragten Fachkonferenzleitungen eine weitreichende Dissemination der Ergebnisse aus VERA und des

MSA innerhalb der Schulen, was mit bisherigen Forschungsbefunden zu VERA korrespondiert (Kühle & Peek, 2007; Wurster et al., 2013). Nahezu alle befragten Fachkonferenzleitungen geben an, sich mit den Ergebnissen aus VERA und dem MSA auseinanderzusetzen. Entgegen der aufgestellten Hypothese zur zweiten Forschungsfrage nutzen deutlich mehr Fachkonferenzleitungen MSA-Ergebnisse für die Entwicklung ihres Unterrichts als VERA-Ergebnisse. Dies zeigt sich auch in der kollektiven Ergebnisnutzung in Fachkonferenzen. Auch in internationaler Forschung wurde festgestellt, dass Ergebnisse aus zentralen Abschlussprüfungen in den Niederlanden und England zur datengestützten Unterrichtsentwicklung eingesetzt werden, wobei allerdings insgesamt eine eher gering ausgeprägte Nutzung beschrieben wird (Schildkamp et al., 2012; Tymms, 1995). Trotz der Unterschiede zwischen den beiden Verfahren kann hinsichtlich der dritten Forschungsfrage hypothesenkonform festgestellt werden, dass datengestützte Unterrichtsentwicklung vor allem simultan auf Basis beider Datenquellen stattfindet; und zwar häufiger als von einer einzelnen Datenquelle ausgehend, die in der überwiegenden Mehrzahl der Fachkonferenzen der MSA ist.

Analog zur bisherigen Verwendungsforschung zu VERA, zeigt sich auch in der vorliegenden Studie, dass sich Unterrichtsentwicklung auf Basis von VERA- und MSA-Ergebnissen am häufigsten in der Entwicklung oder zielgerichteten Veränderung von Aufgaben manifestiert. Weniger häufig werden Materialien und Methoden weiterentwickelt oder Fortbildungen konzipiert. Fraglich bleibt, ob der Aufgabenentwicklung eine tiefgehende Reflexion bzw. Ursachenanalyse vorausgeht. Diemer und Kuper (2011) beschreiben, dass dies auch ohne vorherige Ursachenanalyse durchgeführt wird.

Eine Unterrichtsentwicklung, die vornehmlich auf Aufgaben, aber auch Materialien ausgerichtet ist, könnte einerseits Ausdruck unerwünschter Nebenfolgen wie einer Verengung des Curriculums sein. Allerdings zeigen einige Studien, dass dies zumindest bei VERA nicht der Fall ist (Richter et al., 2014; Wacker & Kramer, 2012). Unterrichtsentwicklung mit Aufgaben kann andererseits eine wichtige Strategie zur Implementation von Neuerungen wie den Bildungsstandards sein, indem im Unterricht die auf Basis der Bildungsstandards entwickelten Aufgaben Einzug finden (Klein et al., 2014; Maier et al., 2011).

Vergleicht man die beiden Verfahren hinsichtlich der wahrgenommenen Nützlichkeit, des Aufwands und der Diagnosegüte, so zeigt sich eine positivere Bewertung des MSA. Andere Studien bestätigen, dass VERA im Vergleich zu anderen Evaluationsverfahren wie der Schulinspektion oder interner Evaluation als weniger nützlich eingeschätzt wird und dass die Ergebnisse weniger zur Schul- und Unterrichtsentwicklung verwendet werden (Demski, Rosenbusch, van Ackeren, Clausen & Schmidt, 2012; Wurster et al., 2013).

Ausgehend davon, dass VERA, als explizit für Unterrichtsentwicklung vorgesehenes Verfahren, im Vergleich zum MSA dahingehend weniger verwendet und als nützlich eingeschätzt wird, stellt sich die Frage nach den Ursachen. Mögliche Erklärungen lassen sich in der Implementationsforschung und den mit den Verfahren verbundenen Konsequenzen finden. Aus der Implementationsforschung

ist bekannt, dass eine erfolgreiche Veränderung Zeit benötigt und zahlreiche Einflussfaktoren darauf einwirken (Rogers, 2003). Bei der Nutzung von Ergebnissen aus Leistungstests für Unterrichtsentwicklungsmaßnahmen muss von einer doppelten Innovation ausgegangen werden. Die erste Innovation ist die Einführung eines neuen Testverfahrens. Die zweite damit verbundene Innovation ist die aus der Datennutzung abgeleitete Einführung bzw. Umsetzung von Maßnahmen und Änderungen innerhalb der Schule. VERA ist vergleichsweise neu und Lehrkräfte müssen sich zunächst mit dem Verfahren und den Ergebnissen vertraut machen. Abschlussprüfungen sind, zumindest in dezentraler Form, lange Bestandteil schulischer Arbeit. Der Umgang mit Ergebnissen wird deshalb viel routinierter sein und die Bewertungspraxis nachvollziehbarer, da sie in Form von Noten und Bestehensquoten näher am Alltag der Lehrkräfte ist. Im Vergleich zu VERA schreiben Fachkonferenzleitungen dem MSA eine höhere Diagnosegüte und Nützlichkeit zu, was möglicherweise auch damit zusammenhängt, dass der MSA länger und damit stärker in der Schulpraxis verankert ist. Zudem ist der MSA an den landesspezifischen Rahmenlehrplänen ausgerichtet und VERA an den ebenfalls vergleichsweise neuen Bildungsstandards.

Ein weiteres Merkmal in dem sich VERA und MSA unterscheiden, sind die mit den Verfahren verbundenen Konsequenzen für schulische Akteure. VERA ist ein *Low-Stakes*-Testsystem, wohingegen der MSA auf Ebene der Schülerschaft ein *High-Stakes*-Testsystem darstellt. Ein Literaturreview kommt bei der Analyse der Konsequenzen von Testverfahren zu dem Schluss, dass in Testsystemen mit moderaten *stakes* die Daten eher verwendet werden als in jenen mit *low* oder *high stakes* (Hellrung & Hartig, 2013). *Low-Stakes*-Verfahren scheinen kaum zur Ergebnisnutzung zu animieren und *High-Stakes*-Verfahren sind mit einer großen Bandbreite an nicht intendierten Nebenfolgen verbunden. Möglicherweise hat der MSA innerhalb der Schulen eine größere Bedeutung als VERA, weil die Konsequenzen für die Schülerschaft auch von Lehrkräften ernster genommen werden, so dass der MSA insgesamt für schulische Akteure stärkere Konsequenzen aufweist als VERA.

Signifikante Prädiktoren für die Verwendung von MSA-Ergebnissen sind vor allem Merkmale der Schulorganisation, wie die kollegiale Unterrichtsentwicklung, jedoch nicht der Bezug zu den Bildungsstandards im Curriculum. Der Bezug zwischen zentralen Abschlussprüfungen und den Bildungsstandards ist in Brandenburg nur indirekt z. B. in den Rahmenlehrplänen (MBSJ, 2008) kommuniziert. Die Abschlussprüfungen basieren auf den Rahmenlehrplänen, die sich wiederum an den Bildungsstandards orientieren. Die Verwendung von VERA-Ergebnissen hängt stärker mit verfahrensbezogenen Faktoren zusammen, wie der wahrgenommenen Nützlichkeit von VERA. Ein weiterer Prädiktor ist die Ausrichtung des Curriculums an den Bildungsstandards, die Grundlage für die in VERA erfassten Kompetenzen sind. Eine kriteriale Bezugsnormorientierung in der Rückmeldung und bei der Ergebnisreflexion wurde auch in anderen Studien als wichtiger Prädiktor für die VERA-Nutzung identifiziert (Groß Ophoff, 2013; Kühle, 2010). Das bedeutet, dass Elemente der neuen Steuerung wie Bildungsstandards und VERA nicht ge-

trennt voneinander im schulischen Kontext Bedeutung erlangen. Hier bietet sich ein wichtiger Ansatzpunkt für die Ausgestaltung von Unterstützungsmaßnahmen. In Bezug auf die Hypothese, dass die Prädiktoren Nützlichkeit, innerschulische Kommunikation der Ergebnisse und kollegiale Unterrichtsentwicklung einen positiven Einfluss auf die Ergebnisnutzung haben, zeigt sich, dass diese im Gesamtmodell nicht bei jedem Testsystem signifikant sind. In Modellen mit weniger Prädiktoren zeigen sie allerdings einen bedeutsamen Einfluss. In zukünftigen Studien mit einer größeren Stichprobe sollte deshalb mit Bezug auf theoretische Modelle (z. B. Hosenfeld & Groß Ophoff, 2007) die Beziehung der Prädiktoren untereinander mittels Strukturgleichungsmodellen modelliert werden. Dies ermöglicht, Fragen zu möglichen Mediations- bzw. Moderationseffekten nachzugehen.

Die Ergebnisse bieten Hinweise zur Weiterentwicklung der Testsysteme. Fortbildungen zum Umgang mit den Daten sind ein probates Mittel, da sie einen positiven Einfluss auf das Verständnis und die Nutzung von Daten aus VERA (Koch, 2011) und zentralen Abschlussprüfungen haben (Tymms, 1995). In Fortbildungen kann auf die Nutzbarkeit von Ergebnissen aus zentralen Prüfungen zumindest für das Monitoring und, bei entsprechendem Bedarf, für Entwicklungsprozesse hingewiesen werden. Um die wahrgenommene Nützlichkeit und damit auch Nutzung von VERA zu erhöhen, besteht nach wie vor die Notwendigkeit, das Potential von VERA aufzuzeigen. Das kann nicht nur über Fortbildungen erfolgen. Weitere Möglichkeiten sind beispielsweise gezielte Informationen zur Weiterarbeit mit den Ergebnissen oder der Bereitstellung von fachdidaktisch fundierten und zum getesteten Inhalt und den Bildungsstandards passenden Lernaufgaben, wie es in Ansätzen bereits z. B. in Brandenburg für VERA umgesetzt wird (ISQ, 2012). Für zentrale Prüfungen besteht diesbezüglich Potential zur Weiterentwicklung der Rückmeldungen. Aufgaben stehen in den Fachkonferenzen der Stichprobe im Mittelpunkt von Unterrichtsentwicklungsaktivitäten und erscheinen deshalb als ein wichtiger Anknüpfungspunkt. In Fortbildungen und Beratungen sollte zudem vor allem die gemeinsame Verwendung von VERA- und MSA-Ergebnissen im Mittelpunkt stehen. In einem Teil der Fachkonferenzen wird dies bereits praktiziert. Zudem zeigt sich, dass die VERA-Ergebnisse die MSA-Ergebnisse prognostizieren können (Graf et al., in Druck). Auf diese Weise könnte die positivere Wahrnehmung des MSA für eine Akzeptanzsteigerung von VERA eingesetzt werden. In diesem Zusammenhang erscheint auch eine stärkere Kommunikation des Bezuges zwischen den Abschlussprüfungen und den Bildungsstandards sinnvoll.

Bei der Interpretation der Befunde ist zu beachten, dass die jeweilige Implementation der SPFS zwischen den Bundesländern variiert (Kühn, 2013; Maier, 2008). Es zeigt sich, dass in Thüringen VERA von Lehrkräften positiver eingeschätzt wird als in Baden-Württemberg (Maier, 2008) und in Brandenburg von Schulleitungen positiver als in Berlin (Wurster et al., 2013). Eine Generalisierbarkeit der Befunde über Brandenburg hinaus ist deshalb nicht ohne weiteres möglich. Obwohl der Rücklauf für diese Art von Studien zwar üblich ist und die Verteilung der Schulformen in der Stichprobe und der Population in etwa übereinstimmt, kann eine mögliche Verzerrung der Stichprobe nicht ausgeschlos-

sen werden, zumal keine soziodemographischen Merkmale zu den Lehrkräften erfasst werden konnten.

Für zukünftige Forschung ist die Fragestellung interessant, ob und unter welchen Bedingungen datenbasierte Unterrichtsentwicklung ausgehend von VERA- und MSA-Ergebnissen zu einer Verbesserung der Schülerleistung führt. Ein weiteres Forschungsfeld ist die Ausgestaltung von Unterstützungssystemen, wie beispielsweise die Gestaltung von Fortbildungen, Rückmeldungen oder anderen Maßnahmen.

Literatur

- Altrichter, H. & Maag Merki, K. (2010). *Handbuch Neue Steuerung im Schulsystem*. Wiesbaden: VS.
- Bach, A., Wurster, S., Thillmann, K., Pant, H. A. & Thiel, F. (2014). Vergleichsarbeiten und schulische Personalentwicklung – Ausmaß und Voraussetzungen der Datennutzung. *Zeitschrift für Erziehungswissenschaft*, 17(1), 61–84. doi: 10.1007/s11618-014-0486-5
- Bonsen, M., Büchter, A. & Peek, R. (2006). Datengestützte Schul- und Unterrichtsentwicklung. Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen und Lehrer. In H. G. Holtappels, K. Klemm, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung* (Bd. 14, S. 125–148). Weinheim: Juventa.
- Buuren, S. van & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- Cohen, J. (1992). A Power primer. *Psychological Bulletin*, 112(1), 155–159.
- Demski, D., Rosenbusch, C., Ackeren, I. van, Clausen, M. & Schmidt, U. (2012). Steuerung von Schule durch evidenzbasierte Einsicht? Konzeption und erste Befunde des Forschungsverbundes EviS. In S. Hornberg & M. Parreira do Amaral (Hrsg.), *Deregulierung im Bildungswesen* (S. 131–150). Münster: Waxmann.
- Diemer, T. & Kuper, H. (2011). Formen innerschulischer Steuerung mittels zentraler Lernstandserhebungen. *Zeitschrift für Pädagogik*, 57(4), 554–571.
- Gärtner, H. & Wurster, S. (2009). *Befragung zur Wirkung von Schulinspektion in Berlin. Ergebnisbericht*. Berlin: Institut für Schulqualität der Länder Berlin und Brandenburg.
- Graf, T., Harych, P., Wendt, W., Emmrich, R. & Brunner, M. (in Druck). Prognostische Validität von VERA: Wie gut können VERA-8-Testergebnisse den schulischen Erfolg am Ende der Sekundarstufe I vorhersagen? *Zeitschrift für Pädagogische Psychologie*.
- Groß Ophoff, J. (2013). *Lernstandserhebungen: Reflexion und Nutzung*. Münster: Waxmann.
- Hellrung, K., & Hartig, J. (2013). Understanding and using feedback – A review of empirical studies concerning feedback from external evaluations to teachers. *Educational Research Review*, 9, 174–190. doi: 10.1016/j.edurev.2012.09.001
- Helmke, A. (2004). Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule. *Seminar*, 2/2014, 90–112.
- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Hosenfeld, I. & Groß Ophoff, J. (2007). Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht. *Empirische Pädagogik*, 21(4), 352–367.

- Hoth, K. & Wesselhöfft, K. (2011). *Mittlerer Schulabschluss Berlin 2010. Urteile aus den Schulen zu den Ergebnisrückmeldungen an die Schulen. Eine explorative Studie*. Berlin: Institut für Schulqualität der Länder Berlin und Brandenburg e.V.
- ISQ – Institut für Schulqualität der Länder Berlin und Brandenburg e.V. (2012). *Vergleichsarbeiten der Jahrgangsstufe 8 in Brandenburg als Beitrag zur Schul- und Unterrichtsentwicklung*. Berlin: Institut für Schulqualität der Länder Berlin und Brandenburg.
- Klein, E. D. (2013). *Statewide exit exams, governance, and school development. An international comparison*. Münster: Waxmann.
- Klein, E. D., Krüger, M., Kühn, S. M. & Ackeren, I. van (2014). Wirkungen zentraler Abschlussprüfungen im Mehrebenensystem Schule. Eine Zwischenbilanz internationaler und nationaler Befunde und Forschungsdesiderata. *Zeitschrift für Erziehungswissenschaft*, 17(7), 7–33. doi: 10.1007/s11618-014-0479-4
- Klein, E. D., Kühn, S. M., van Ackeren, I. & Block, R. (2009). Wie zentral sind zentrale Prüfungen? Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich. *Zeitschrift für Pädagogik*, 55(4), 596–621.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: LinkLuchterhand.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Wolters Kluwer.
- Koch, U. (2011). *Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnisrückmeldungen aus Vergleichsarbeiten*. Münster: Waxmann.
- Koch, U., Groß Ophoff, J., Hosenfeld, I. & Helmke, A. (2006). Qualitätssicherung: Von der Evaluation zur Schul- und Unterrichtsentwicklung – Ergebnisse der Lehrerbefragung zur Auseinandersetzung mit den VERA-Rückmeldungen. In F. Eder, A. Gastager & F. Hofmann (Hrsg.), *Qualität durch Standards? Beiträge zur 67. Tagung der Arbeitsgruppe der Empirischen Bildungsforschung (AEPF), Salzburg* (S. 187–199). Münster: Waxmann.
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529–548). Wiesbaden: VS.
- Kühle, B. (2010). *Zentrale Lernstandserhebungen – Ergebnisorientierte Unterrichtsentwicklung? Schulische Strategien beim Umgang mit Ergebnissen aus den Schulrückmeldungen im Kontext der ersten Lernstandserhebungen 2004/2005 in Nordrhein-Westfalen*. Berlin: Köster.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. *Empirische Pädagogik*, 21(4), 428–447.
- Kühn, S. M. (2013). Vergleichbarkeit beim Mittleren Schulabschluss? Ein Überblick über die Vielfalt schulstrukturell möglicher Bildungswege und Prüfungsverfahren in den deutschen Ländern. *Die Deutsche Schule*, 105(1), 86–101.
- Lai, M. K., & Schildkamp, K. (2013). Data-Based decision making: An overview. In K. Schildkamp, M. K. Lai & L. Earl (Hrsg.), *Data-based decision making in education. Challenges and opportunities* (S. 9–21). Heidelberg: Springer.
- Little, R. J. A. (1988). Missing data adjustments in large surveys. *Journal of Business Economics and Statistics*, 6(3), 287–301.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 145–169). Wiesbaden: VS.
- Maier, U. (2008). Vergleichsarbeiten im Vergleich – Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. *Zeitschrift für Erziehungswissenschaft*, 11(3), 453–474.

- Maier, U., Bohl, T., Kleinknecht, M. & Metz, K. (2011). Einflüsse von Merkmalen des Testsystems und Schulkontextfaktoren auf die Akzeptanz und Rezeption von zentralen Testrückmeldungen durch Lehrkräfte. *Journal for Educational Research Online*, 3(2), 62–93.
- Maier, U., Metz, K., Bohl, T., Kleinknecht, M. & Schymala, M. (2012). Vergleichsarbeiten als Instrument der datenbasierten Schul- und Unterrichtsentwicklung in Gymnasien. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikation* (S. 197–224). Wiesbaden: VS.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale, NJ: Erlbaum.
- MBJS – Ministerium für Bildung, Jugend und Sport. (2008). *Rahmenlehrplan für die Sekundarstufe I. Mathematik*. Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- MBK – Ministerium für Bildung und Kultur. (2011). *Leitfaden zur Nutzung der Ergebnisse zentraler Prüfungen*. Kiel: Ministerium für Bildung und Kultur des Landes Schleswig-Holstein.
- Nachtigall, C. & Hellrung, K. (2013). Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. *Empirische Pädagogik*, 27(4), 423–441.
- Nachtigall, C. & Jantowski, A. (2007). Die Thüringer Kompetenztests unter besonderer Berücksichtigung der Evaluationsergebnisse zum Rezeptionsverhalten. *Empirische Pädagogik*, 21(4), 401–410.
- Pant, H. A. (2013). Wer hat einen Nutzen von Kompetenzmodellen? *Zeitschrift für Erziehungswissenschaft*, 16(1), 71–79. doi: 10.1007/s11618-013-0388-y
- Richter, D., Böhme, K., Becker, M., Stanat, P. & Pant, H. A. (2014). Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten: Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen der Schülerinnen und Schüler. *Zeitschrift für Pädagogik*, 60(2), 225–244.
- Rogers, E. M. (2003). *Diffusion of innovations* (5. Aufl.). New York, NY: Free Press.
- Saunders, L. (2000). Understanding schools' use of 'value added' data: The psychology and sociology of numbers. *Research Papers in Education*, 15(3), 241–258. doi: 10.1080/02671520050128740
- SBI – Sächsisches Bildungsinstitut. (2012). *Kompetenztests an sächsischen Schulen*. Radebeul: Sächsisches Bildungsinstitut.
- Schildkamp, K., Rekers-Mombarg, L. T. M. & Harms, T. J. (2012). Student group differences in examination results and utilization for policy and school development. *School Effectiveness and School Improvement*, 23(2), 229–255. doi: 10.1080/09243453.2011.652123
- Tymms, P. B. (1995). Influencing educational practice through performance indicators. *School Effectiveness and School Improvement*, 6(2), 123–145. doi: 10.1080/0924345950060204
- Visscher, A. J. & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School Effectiveness and School Improvement*, 14(3), 321–349.
- Wacker, A. & Kramer, J. (2012). Vergleichsarbeiten in Baden-Württemberg. *Zeitschrift für Erziehungswissenschaft*, 15(4), 683–706. doi: 10.1007/s11618-012-0326-4
- Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2013). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. In I. van Ackeren, M. Heinrich, & F. Thiel (Hrsg.), *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund* (Die Deutsche Schule. Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis, Beiheft 12, S. 19–50). Münster: Waxmann.