

Wurster, Sebastian

Förderliche und hinderliche Bedingungen für die datengestützte Unterrichtsentwicklung mit Vergleichsarbeiten, zentralen Prüfungen und internen Evaluationen

Zeitschrift für Pädagogik 68 (2022) 1, S. 95-116



Quellenangabe/ Reference:

Wurster, Sebastian: Förderliche und hinderliche Bedingungen für die datengestützte Unterrichtsentwicklung mit Vergleichsarbeiten, zentralen Prüfungen und internen Evaluationen - In: Zeitschrift für Pädagogik 68 (2022) 1, S. 95-116 - URN: urn:nbn:de:0111-pedocs-292805 - DOI: 10.25656/01:29280; 10.3262/ZP2201095

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-292805>

<https://doi.org/10.25656/01:29280>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit this document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipt.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Sebastian Wurster

Förderliche und hinderliche Bedingungen für die datengestützte Unterrichtsentwicklung mit Vergleichsarbeiten, zentralen Prüfungen und internen Evaluationen

Zusammenfassung: Ergebnisse aus Vergleichsarbeiten, zentralen Abschlussprüfungen und internen Evaluationen stehen Lehrkräften als ein Ausgangspunkt zur Unterrichtsentwicklung zur Verfügung. Die Verwendung dieser Daten sowie dafür förderliche und hinderliche Bedingungen wurden auf Basis des IQB-Ländervergleichs 2012 ($N = 3099$ Lehrkräfte) vergleichend untersucht. Aspekte der Nutzung der drei Verfahren werden im Mittel moderat positiv bewertet und deren Ergebnisse intensiv rezipiert sowie von einem substantiellen Teil der Lehrkräfte zur datengestützten Unterrichtsentwicklung genutzt. Die bedeutsamsten Prädiktoren für die Ergebnisnutzung sind die eingeschätzte Nützlichkeit der Verfahren und der wahrgenommene Veränderungsbedarf sowie teilweise die Besprechung der Ergebnisse im Kollegium.

Schlagnote: Interne Evaluation, Vergleichsarbeiten, zentrale Abschlussprüfungen, Datennutzung, Lehrkräfte, Unterrichtsentwicklung

1. Einleitung

Datengestützte Schul- und Unterrichtsentwicklung ist ein wesentliches Ziel gegenwärtiger Bildungspolitik. Hintergrund sind Befunde aus internationalen Studien, die zeigen, dass datenbasierte Entscheidungen sehr bedeutsam für die Weiterentwicklung von Schul- und Unterrichtsqualität sein können. Erfolgreiche Datennutzung geht mit der Verbesserung von Schul- und Unterrichtsqualität und ausgehend davon mit einer Steigerung von Schülerleistungen einher (Creemers & Kyriakides, 2008; Kyriakides, Creemers, Antoniou & Demetriou, 2010; Schildkamp, Lai & Earl, 2013). Dementsprechend wurden in Deutschland verschiedene Verfahren zur Qualitätsentwicklung und -sicherung implementiert. Dazu gehören u. a. die bundesweit durchgeführten Vergleichsarbeiten und die auf Landesebene implementierten zentralen Abschlussprüfungen und internen Evaluationen, die Lehrkräften zur Unterrichtsentwicklung zur Verfügung stehen (KMK, 2015; Wurster, 2019). Ziel ist es, dass, nach Analyse der Ergebnisse dieser Verfahren, geeignete Maßnahmen zur Weiterentwicklung schulischer bzw. unterrichtlicher Prozesse initiiert werden.

Angesichts der mit den Verfahren verbundenen finanziellen und personellen Ressourcen muss empirisch überprüft werden, ob und unter welchen Bedingungen die intendierte datengestützte Unterrichtsentwicklung stattfindet. Internationale Überblicksarbeiten zeigen, dass die Rückmeldung von Daten nicht automatisch zu den intendierten

Entwicklungsprozessen führt (Altrichter, Moosbrugger & Zuber, 2016; Coburn & Turner, 2011; Schildkamp, Poortman, Luyten & Ebbeler, 2017). Neben Faktoren auf Ebene der Schule und der DatennutzerInnen können auch die Ziele und Ausgestaltung der Verfahren selbst Einfluss auf die Nutzung haben (Visscher & Coe, 2003). Bisherige Forschung beschränkte sich vorwiegend auf einzelne Verfahren. Deshalb wird gezielt die Konstellation verschiedener Verfahren in den Blick genommen, um kontrastierende Analysen zu ermöglichen. Es wird erstens den Fragen nachgegangen, inwiefern verschiedene Datenquellen durch Lehrpersonen tatsächlich zur Unterrichtsentwicklung verwendet und wie diese bewertet werden sowie ob sich die Verfahren untereinander dahingehend unterscheiden. Zweitens werden förderliche und hinderliche Bedingungen für die Datennutzung analysiert, um Anhaltspunkte für die zukünftige Weiterentwicklung von Verfahren und Unterstützungssystemen herauszuarbeiten.

2. Daten als Ausgangspunkt für Unterrichtsentwicklung

Ausgangspunkt für datengestützte Unterrichtsentwicklung können verschiedene Daten sein. In (fast) allen deutschen Bundesländern¹ stehen Lehrkräften dafür Ergebnisse aus Vergleichsarbeiten (VA), zentralen Abschlussprüfungen (ZP) und internen Evaluationen (IE) zur Verfügung (Thiel et al., 2019). VA und ZP erfassen standardisiert die SchülerInnenleistung. Die Testaufgaben werden auf Basis eines Referenzrahmens entwickelt, der die Lernziele definiert. Der Referenzrahmen für VA sind die nationalen Bildungsstandards in den Fächern Deutsch, Mathematik und erste Fremdsprache (Richter & Stanat, 2019). Die ZP beziehen sich auf die Rahmenlehrpläne der Bundesländer, die sich für die Hauptfächer an den Bildungsstandards orientieren (Kühn, 2013). Während VA explizit als Instrument zur datenbasierten Unterrichtsentwicklung implementiert wurden (KMK, 2010), ist das Hauptziel von ZP die Zertifizierung von SchülerInnenleistung. Damit erfüllen sie eine Selektions- und Allokationsfunktion, indem die in den ZP erbrachten Leistungen Einfluss auf das berufliche Fortkommen der Schülerinnen und Schüler haben (Maag Merki, 2016). Die unterschiedlichen Zielsetzungen führen zu verschiedenen Einsatzszenarien. Die ZP finden zum Abschluss der jeweiligen Schullaufbahn statt. Die VA werden hingegen bereits zwei Jahre vor dem in den Bildungsstandards festgelegten Zeitpunkt des erwarteten Erreichens der Kompetenzen durchgeführt. Die frühzeitige Bereitstellung von kriterialen Informationen über die Schülerleistung soll ein Impuls für die Schul- und Unterrichtsentwicklung sein.

Neben Daten zur SchülerInnenleistung stehen Lehrkräften auch Daten aus IE zur Verfügung, wenn diese in ihrer Schule bzw. von ihnen selbst durchgeführt werden. Bildungspolitisch haben IE einen hohen Stellenwert, wie die gesetzliche Verpflichtung zu ihrer Durchführung in vielen Bundesländern zeigt (Thiel et al., 2019). Zudem sind IE in den Schulqualitätsrahmen der meisten Bundesländer explizit als Teil einer ‚guten Schule‘ genannt. Unter IE werden zahlreiche Verfahren subsummiert, die auf Initia-

1 In Rheinland-Pfalz werden keine zentralen Abschlussprüfungen durchgeführt.

tive schulischer Akteure durchgeführt werden bzw. für deren Durchführung die Einzelschule verantwortlich ist und diese eigenständig gestaltet (Gärtner, 2013). Dies können beispielsweise SchülerInnenfeedback zum Unterricht, Bestandsaufnahmen im Rahmen von Entwicklungsprozessen oder des schulinternen Qualitätsmanagements sein (Berkemeyer, Müller & Van Holt, 2016). IE sind somit alle Formen der systematischen Informationsbeschaffung und -bewertung in der Verantwortung schulischer Akteure. IE sollen Informationen generieren, anhand derer der Erfolg schulischer Arbeit überprüft und Schul- bzw. Unterrichtsentwicklung initiiert werden kann (Thiel & Thillmann, 2012). Zur Unterstützung der Durchführung von IE wurden in den letzten Jahren in allen Bundesländern standardisierte Instrumente entwickelt und bereitgestellt (z. B. Selbstevaluationsportal in Berlin und Brandenburg, SEfU in Thüringen, Sachsen und NRW (Thiel et al., 2019)).

Die Ergebnisse aus VA, ZP und IE sollen Grundlage für die Reflexion und Weiterentwicklung von Unterricht sein (Helmke, 2010; KMK, 2010). Die Grundidee ist für alle Verfahren gleich: Nach der Bereitstellung der Daten soll deren Analyse und Interpretation erfolgen. Aus den Schlussfolgerungen dieser Datenanalyse werden idealtypisch mögliche Entwicklungsbereiche identifiziert. Helmke (2010) unterscheidet dabei die Rezeption und Reflexion der Daten, Coburn und Turner (2011, S. 230) die Schritte „noticing“ und „interpreting“. Diese Schritte sollen nicht nur individuell, sondern möglichst auch kollaborativ in Fachgruppen oder im Kollegium erfolgen (KMK, 2010). Die aus der Analyse gezogenen Schlussfolgerungen bzw. Interpretationen, ob eine (subjektive) Notwendigkeit (Helmke, 2010) besteht, Unterrichtsentwicklung zu initiieren, ist Voraussetzung für die Entwicklung und Realisierung von entsprechenden Maßnahmen. Somit lassen sich grob zwei Schritte der Datennutzung unterscheiden: 1. die Analyse von Daten (Rezeption, Reflexion und Ableitung von Maßnahmen) und 2. die Initiierung und Umsetzung von Maßnahmen (siehe auch Abb. 1 und Abschnitt 6.3).

Der skizzierte Datennutzungsprozess ist anspruchsvoll und voraussetzungsreich. In Modellen und systematischen Reviews zur datengestützten Unterrichtsentwicklung wurden deshalb förderliche und hinderliche Bedingungen herausgearbeitet (Altrichter, Moosbrugger & Zuber, 2016; Coburn & Turner, 2011; Helmke, 2010; Hoogland et al., 2016; Schildkamp, Poortman, Luyten & Ebbeler, 2017; Visscher & Coe, 2003). Auch bildungspolitische Handreichungen formulieren Gelingensbedingungen (KMK, 2010). Förderliche und hinderliche Bedingungen werden in der genannten Literatur in vier Kategorien systematisiert: 1) die *Ergebnisse/Daten* selbst bzw. das Verfahren mit dem diese generiert werden (z. B. Verständlichkeit, Akkuratheit, Nützlichkeit, Ziele und Implementation des Instruments), 2) die *NutzerInnen* (z. B. Expertise im Umgang mit Daten, Akzeptanz des Verfahrens, Motivation), 3) die *Organisation Schule* (z. B. Routinen und Verbindlichkeit im Umgang mit Daten, Kooperation bei Verarbeitungs- und Planungsprozessen sowie der Umsetzung von Maßnahmen, zeitliche und räumliche Gelegenheit zur Auswertung und Planung, data-wise Leadership, Bezug zum Schulprogramm, Innovationskapazität, Schulkultur) und 4) auf den *schulischen Kontext* (z. B. Unterstützungssysteme). Diese Systematisierungen unterscheiden jedoch nicht nach den verschiedenen Verfahren.

3. Datengestützte Unterrichtsentwicklung in Deutschland

3.1 Daten zur Leistung von Schülerinnen und Schülern

Angesichts der Herausforderungen in der Umsetzung stellt sich die Frage, ob ausgehend von VA, ZP und IE datengestützte Unterrichtsentwicklung in der intendierten Form stattfindet und welche Faktoren darauf Einfluss haben. Zu VA liegen die meisten Studien vor, Ergebnisse aus VA werden insbesondere durch Lehrkräfte rezipiert (Groß Ophoff, Koch & Hosenfeld, 2019; Nachtigall & Hellrung, 2013; Wurster & Richter, 2016). Die Mehrheit der Lehrkräfte, Fachkonferenz- und Schulleitungen gibt an, ausgehend von VA-Ergebnisse Schul- und Unterrichtsentwicklung zu initiieren. Diese haben primär Bezug zum Unterricht, wie die gezielte Übung von VA-Aufgaben, die (Weiter-)Entwicklung von Aufgaben und Unterrichtsmaterialien oder die Einführung neuer Methoden. Ein Teil der Lehrkräfte nutzt VA-Ergebnisse für verstärkte Einzelförderung oder Kompetenzorientierung (Diemer & Kuper, 2011; Groß Ophoff, 2013; Nachtigall & Hellrung, 2013; Wurster, Richter & Lenski, 2017). Bei Grundschullehrkräften zeigt sich jedoch keine generelle Tendenz dazu (Richter, Böhme, Becker, Stanat & Pant, 2014). Schulentwicklungsmaßnahmen, wie Fortbildungen, Intensivierung der kollegialen Kommunikation und Arbeit an schulinternen Curricula, finden seltener statt (Bach, Wurster, Thillmann, Pant & Thiel, 2014; Diemer & Kuper, 2011; Groß Ophoff, 2013; Wurster, Richter & Lenski, 2017).

Förderliche Bedingungen der datengestützten Unterrichtsentwicklung mit VA sind mit Bezug auf das Verfahren und die NutzerInnen insbesondere die Akzeptanz/wahrgenommene Nützlichkeit des Verfahrens, die eingeschätzte Messqualität, die Motivation zur und Intensität der Ergebnisreflexion, eine kriteriale Bezugsnormorientierung (z.B. an den Bildungsstandards) und für die Rezeption der Ergebnisse soziale Bezugsnormorientierung, Innovationsbereitschaft sowie die Fortbildungsaktivität (Groß Ophoff, 2013; Groß Ophoff, Koch & Hosenfeld, 2019; Nachtigall & Hellrung, 2013; Wurster & Richter, 2016). Die Bedeutung der Bewertung des Verfahrens zeigt sich auch darin, dass in der Grundschule bessere Lese- und Mathematikleistungen erreicht werden, wenn Lehrkräfte VA als Mittel der Unterrichtsentwicklung sehen und dementsprechend (vermutlich) eher die Daten nutzen (Richter, Böhme, Becker, Stanat & Pant, 2014). Eigenschaften der Schule, die mit der Nutzung von VA-Ergebnissen zusammenhängen, sind die Unterstützung durch die Schulleitung, das Innovationsklima und kollegiale Kooperationen (Bach, Wurster, Thillmann, Pant & Thiel, 2014; Groß Ophoff, 2013; Maier, Metz, Bohl, Kleinknecht & Schymala, 2012; Stump, Zlatkin-Troitschanskaia & Mater, 2016).

Ausgangspunkt für Unterrichtsentwicklung können verschiedene Leistungsdaten sein (Wurster & Richter, 2016). Brandenburger Fachkonferenzleitungen berichten, dass Ergebnisse aus VA und den ZP zum mittleren Schulabschluss in der Mehrzahl der Fachkonferenzen intensiv rezipiert werden und z.B. Ausgangspunkt für die Weiterentwicklung von Lern- und Testaufgaben sind (Wurster & Richter, 2016). Im Vergleich zu VA

erfolgt die Initiierung datengestützter Unterrichtsentwicklung in Fachkonferenzen häufiger ausgehend von ZP-Ergebnissen. Diese werden auch als nützlicher bewertet und das, obwohl Rückmeldungen aus VA deutlich mehr Informationen enthalten. Jedoch erfolgt datengestützte Unterrichtsentwicklung am häufigsten auf Basis beider Datenquellen gemeinsam. In Deutschland gibt es keine weitere Forschung, jedoch zeigt sich international in den wenigen Studien ebenfalls, dass ZP Basis von Schul- und Unterrichtsentwicklung sind (Klein, Krüger, Kühn & Van Ackeren, 2014).

3.2 Daten aus internen Evaluationen

Trotz der großen bildungspolitischen Bedeutung liegen aus Deutschland nur wenige empirische Arbeiten zu IE vor, wenngleich der Verbreitungsgrad von IE als hoch einzuschätzen ist. Von den im Rahmen des IQB-Ländervergleich 2012 befragten Mathematiklehrkräften berichten 80%, IE-Ergebnisse als Ausgangspunkt für Unterrichtsentwicklung verwendet zu haben (Wurster, Richter & Lenski, 2017). Zwischen den PISA Tests 2000 und 2009 fand nach Angabe von Schulleitungen an 70% der Schulen eine Zunahme an IE statt, jedoch wurde an 18% der Schulen bis 2009 keine IE durchgeführt (Bischof, Hochweber, Hartig & Klieme, 2013). IE wird generell – und im Vergleich zu anderen Verfahren – von Schulleitungen und Lehrkräften als nützlicher bewertet. Die Ergebnisse aus IE werden zudem umfassend rezipiert (Demski, 2017). Schulleitungen und Lehrpersonen berichten sowohl von initiiertem Schulentwicklung, wie z. B. die Intensivierung von Kommunikation im Kollegium, als auch Unterrichtsentwicklung, wie z. B. der Einsatz neuer Unterrichtsmethoden und -materialien (Wurster, Richter & Lenski, 2017). Trotz des hohen Anteils an schulischen Akteuren, die von der Nutzung von IE für Entwicklungsprozesse berichten, zeigt sich für einzelne Aktivitäten keine hohe Ausprägung. Dies wird durch Studien zu konkreten Instrumenten der IE zur standardisierten Erfassung der Unterrichtsqualität bestätigt (Gärtner, 2014; Gärtner & Vogt, 2013). Wenn Unterrichtsentwicklung durchgeführt wird, werden bspw. Unterrichtsmethoden verändert oder es findet Einzelförderung statt. Lehrkräfte berichten zudem, dass sie sich vor allem auf die Bereiche der Unterrichtsqualität konzentrieren, die von der Schülerschaft positiv bewertet wurden. Hinsichtlich möglicher förderlicher oder hinderlicher Bedingungen liegt ein Desiderat vor.

4. Forschungsfragen

Trotz einiger Studien und systematischer Zusammenfassungen bleiben – insbesondere im deutschsprachigen Raum – zentrale Fragen zur Bewertung und Verwendung von Daten aus VA, ZP und IE offen. Dies gilt insbesondere für förderliche Bedingungen datenbasierter Unterrichtsentwicklung durch Lehrpersonen sowie der Frage, ob sich solche Bedingungen zwischen den Verfahren unterscheiden. Daraus ergeben sich die Forschungsfragen:

- 1) Wie werden VA, ZP und IE von Lehrpersonen hinsichtlich der Nützlichkeit, Verständlichkeit und Informiertheit über das Verfahren eingeschätzt und deren Ergebnisse innerschulisch verarbeitet? Welche Unterschiede bestehen zwischen den verschiedenen Verfahren? Unterscheidet sich die Bewertung der Verfahren zwischen Lehrpersonen, die bereits selbst Erfahrung mit einem Verfahren haben von Lehrpersonen, die keine Erfahrung damit haben?

Lehrpersonen und Fachkonferenzleitungen berichten eine positivere Einschätzung der Verfahren IE und ZP im Vergleich zu VA (Demski, 2017; Wurster & Richter, 2016). Deshalb wird die Hypothese formuliert, dass Lehrpersonen den Stellenwert von VA als Ausgangspunkt für ihre Unterrichtsentwicklung als niedriger ansehen als ZP und IE. Hinsichtlich der eigenen Erfahrung mit den Verfahren wird keine gerichtete Hypothese aufgestellt, da die eigene Involviertheit sowohl zu positiven als auch zu negativen Einstellungen führen kann.

- 2) Welche Unterrichtsentwicklungsaktivitäten werden aus Ergebnissen von VA, ZP und IE von Lehrpersonen abgeleitet? Bestehen Unterschiede zwischen den verschiedenen Datentypen?

Einzelne Studien berichten für Lehrpersonen (Demski, 2017) und Fachkonferenzleitungen (Wurster & Richter, 2016) Unterschiede in der Verwendung von Daten aus verschiedenen Verfahren. Es zeigt sich, dass Ergebnisse aus ZP und IE häufiger Ausgangspunkt von Unterrichtsentwicklung sind als VA-Daten. Befunde für Mathematiklehrpersonen aus dem IQB-Ländervergleich 2012 zeigen jedoch keine signifikanten Unterschiede (Wurster, Richter & Lenski, 2017). Deshalb wird die Hypothese formuliert, dass es keine Unterschiede in der Nutzung bei Lehrpersonen zwischen den verschiedenen Verfahren gibt.

- 3) Welche förderlichen und hinderlichen daten- und nutzerbezogenen sowie organisationalen Bedingungen stehen mit datengestützter Unterrichtsentwicklung in Beziehung? Welche Unterschiede bestehen zwischen den verschiedenen Verfahren?

Ausgehend vom Stand der Forschung (Altrichter, Moosbrugger & Zuber, 2016; Schildkamp, Poortman, Luyten & Ebbeler, 2017) wird angenommen, dass insbesondere Faktoren wie die Bewertung des Verfahrens sowie die Intensität der Verarbeitung der Ergebnisse wesentliche Prädiktoren für die Ableitung von Entwicklungsmaßnahmen sind. Die Verfahren haben zwar teilweise andere Funktionen (Entwicklungs- vs. Zertifizierungsfunktion) und beinhalten unterschiedliche Informationen mit denen gearbeitet werden kann. Es gibt jedoch keine Hinweise darauf, dass der grundlegende Mechanismus, wie er in Modellen zur datengestützten Unterrichtsentwicklung postuliert wird, zwischen den Verfahren variiert. Es wird deshalb die Hypothese aufgestellt, dass die Prädiktoren die gleiche Tendenz in der Richtung und Größe des Effekts auf datengestützte Unterrichtsentwicklung aufweisen.

5. Methode

5.1 Stichprobe

Die Studie ist eine Sekundäranalyse mit Daten der Lehrpersonenbefragung des IQB-Ländervergleichs 2012 (IQB-LV 2012: Pant et al., 2015) aus $N = 1171$ weiterführenden Schulen (ohne Förderschulen, $N = 515$ Gymnasien). Mit Ausnahme der Analysen zu Forschungsfrage 1, zur Rolle der Erfahrung mit den Verfahren, gehen insgesamt $N = 3028$ Lehrkräfte mit Erfahrung in mindestens einem Verfahren in die Analysen ein. Davon haben 60% Erfahrung mit VA, 46% mit ZP und 59% mit IE. In der Stichprobe sind 43% Gymnasiallehrkräfte, 56% weiblich, das Durchschnittsalter beträgt 48,7 Jahre ($SD = 9,4$) und die mittlere Berufserfahrung 21,3 Jahre ($SD = 12$). Es gehen wie in der Studie von Wurster, Richter und Lenski (2017) ebenfalls Mathematiklehrkräfte ($N = 1355$) und zusätzlich Lehrpersonen aus den Naturwissenschaften Biologie ($N = 707$), Chemie ($N = 736$) und Physik ($N = 872$) in die Analysen ein. Vergleiche zwischen VA, ZP und IE werden mit einer Teilstichprobe von $N = 1009$ Lehrpersonen mit Erfahrung in allen drei Verfahren durchgeführt. Es wurde geprüft, ob in der Stichprobe ausschließlich Mathematiklehrkräfte Erfahrung mit VA haben, da VA nicht in Naturwissenschaften durchgeführt werden. Von den Lehrkräften mit Erfahrung in VA haben 19,5% weder die Lehrbefähigung für Mathematik, noch im betreffenden Schuljahr Mathematik unterrichtet.

5.2 Operationalisierung

Im Rahmen des IQB-Ländervergleichs 2012 wurden verschiedene Einzelitems zur datengestützten Unterrichtsentwicklung erfasst (Lenski et al., 2016). Als ZP wurde konkret nach dem Mittleren Schulabschluss (MSA) gefragt. Die Items sind in den Tabellen 1 und 2 dargestellt und wurden den in der Literatur genannten Kategorien der Datennutzung und Einflussfaktoren zugeordnet. Zum Datennutzungsprozess (Coburn & Turner, 2011; Helmke, 2010; KMK, 2010) sind sowohl Items zur Analyse der Daten (individuelle Auseinandersetzung mit Ergebnissen, Ergebnisbesprechung im Kollegium) als auch Items zu den initiierten Maßnahmen enthalten (z. B. Einführung neuer Methoden, Maßnahmen zur Einzelförderung). Ergänzend kann das Item zum wahrgenommenen Veränderungsanlass als Indikator für die Interpretation der Daten und die subjektive Notwendigkeit Unterrichtsentwicklung zu initiieren (Helmke, 2010) eingesetzt werden. Als förderliche Bedingungen datengestützter Unterrichtsentwicklung können verschiedene Prädiktoren aus den Bereichen Eigenschaften der Daten und NutzerInnen analysiert werden (Zuordnung bei Helmke (2010): „Individuelle Bedingungen“; Schildkamp, Poortman, Luyten & Ebbeler (2017, S. 244): „data“ und „user“; Visscher & Coe (2003, S. 331): „implementation process“, school performance feedback systems features“). Dazu gehören die (1) Informiertheit über das Verfahren, (2) Verständlichkeit der Rückmeldung und (3) eingeschätzte Nützlichkeit der Daten für die Unterrichtsent-

wicklung. Als weitere potenzielle Einflussfaktoren wurde die (4) Berufserfahrung in Jahren und das (5) Geschlecht der Lehrpersonen sowie (6) die Schulform (Gymnasium vs. andere weiterführende Schulen), an der diese unterrichten, erfasst. Das Modell ist in Abschnitt 6.3 beschrieben.

5.3 Analysestrategie

Die ersten beiden Forschungsfragen zu Einschätzung, Verarbeitung und Verwendung von VA, ZP und IE sowie möglicher Unterschiede zwischen den drei Verfahren werden mittels deskriptiver Verfahren, T-Tests und Varianzanalysen (ANOVA) sowie der entsprechenden Effektstärkemaße η^2 und Cohens d analysiert (Cohen, 1992). Mittelwertvergleiche im Rahmen der ANOVA wurden mittels multipler Paarvergleiche (T-Tests) mit Bonferroni-Korrektur untersucht. Fächerunterschiede wurden nicht berücksichtigt, da die überwiegende Mehrheit der Lehrpersonen mindestens zwei der mathematisch-naturwissenschaftlichen Fächer unterrichtet und ein Vergleich aufgrund der zahlreichen Doppelungen nicht sinnvoll ist.

Zur Analyse förderlicher Bedingungen datengestützter Unterrichtsentwicklung werden Strukturgleichungsmodelle (SEM) eingesetzt. Zur Komplexitätsreduktion sind die Einzelitems der Unterrichtsentwicklungsaktivitäten separat für jedes Verfahren zu einem gemeinsamen latenten Konstrukt ‚datenbasierte Unterrichtsentwicklung‘ zusammengefasst. Diese weisen eine gute Reliabilität auf (VA: Cronbachs α /McDonald, 1999: $\omega = .83/.83$; MSA: $\alpha/\omega = .82/.82$; IE: $\alpha/\omega = .81/.81$). Die neu gebildeten latenten Variablen werden als abhängige Variablen in den SEM zur Identifikation förderlicher Bedingungen datengestützter Unterrichtsentwicklung untersucht. Die Analysen wurden mit der Software R und dem Paket lavaan (Rosseel, 2012) durchgeführt. Um die Mehrebenenstruktur der Daten (LehrerInnen geschachtelt in Schulen) zu berücksichtigen, wurden cluster robust standard errors verwendet. Zum Umgang mit fehlenden Werten wurde das Full-Information-Maximum-Likelihood-Verfahren (FIML) eingesetzt (Range fehlende Werte: 1%–8%). Die SEM wurden für jedes Verfahren separat mit der Substichprobe der jeweils mit dem Verfahren vertrauten Personen berechnet.

6. Ergebnisse

6.1 *Einschätzung und Verarbeitung von Ergebnissen aus Vergleichsarbeiten, zentralen Prüfungen und interner Evaluation*

Zur Klärung der Frage, wie VA, ZP und IE von Lehrpersonen eingeschätzt und deren Ergebnisse innerschulisch verarbeitet werden, wird zunächst analysiert, ob sich Lehrpersonen mit Erfahrung von Lehrpersonen ohne Erfahrung mit dem jeweiligen Verfahren unterscheiden (Tab. 1).

Item	VA			MSA			IE		
	<i>M (SD) mit Erfahrung</i>	<i>M (SD) ohne Erfahrung</i>	<i>d</i>	<i>M (SD) mit Erfahrung</i>	<i>M (SD) ohne Erfahrung</i>	<i>d</i>	<i>M (SD) mit Erfahrung</i>	<i>M (SD) ohne Erfahrung</i>	<i>d</i>
Ich habe mich mit den Ergebnissen intensiv auseinandergesetzt.	3.00 (0.77)	1.72 (0.82)	1.65	3.32 (0.72)	1.80 (0.86)	2.03	2.90 (0.72)	1.88 (0.81)	1.37
Ich habe mit meinen Kolleg(inn)en die Ergebnisse besprochen	3.08 (0.76)	1.84 (0.91)	1.58	3.38 (0.70)	1.92 (0.94)	1.94	3.01 (0.71)	1.97 (0.87)	1.40
Ich fühle mich gut über die Evaluation informiert.	2.90 (0.80)	1.83 (0.89)	1.32	3.17 (0.76)	1.89 (0.92)	1.61	2.88 (0.73)	2.01 (0.83)	1.17
Die Rückmeldungen der Ergebnisse waren für mich verständlich.	2.98 (0.78)	1.97 (0.95)	1.26	3.20 (0.74)	2.01 (0.97)	1.52	2.99 (0.67)	2.13 (0.89)	1.18
Für die Unterrichtsentwicklung waren die Ergebnisse nützlich.	2.50 (0.80)	1.72 (0.81)	0.98	2.82 (0.83)	1.73 (0.82)	1.31	2.53 (0.75)	1.86 (0.78)	0.88
Es gab keinen Anlass dazu, aufgrund der Ergebnisse meinen Unterricht zu ändern.	2.56 (0.78)	2.19 (1.07)	0.45	2.58 (0.84)	2.17 (1.04)	0.46	2.54 (0.74)	2.38 (0.96)	0.20

Anmerkungen. Vierstufige Skala von 1 = trifft gar nicht zu bis 4 = trifft völlig zu; Alle Mittelwertsvergleiche (*t*-Tests) signifikant ($p < .001$). *N* Erfahrung mit Verfahren: IE = 2224 bis 2776, MSA = 1666 bis 1713, VA = 2084 bis 2139; *N* keine Erfahrung mit Verfahren: IE = 597 bis 665, MSA = 368 bis 422, VA = 283 bis 308.

Tab. 1: Vergleich der Einschätzung von Vergleichsarbeiten, zentralen Prüfungen und interner Evaluation zwischen Lehrpersonen mit und ohne Erfahrung mit diesen Verfahren.

Lehrpersonen mit Erfahrung schätzen alle Items bedeutsam positiver ein (IE: $.2 < |d| > 1.4$; MSA: $.5 < |d| > 2.0$; VA: $.5 < |d| > 1.7$). Dazu gehört erwartungsgemäß die individuelle Ergebnisreflexion ($1.4 < |d| > 2.0$) und Verständlichkeit der Rückmeldung ($1.2 < |d| > 1.5$), aber auch die Beteiligung an der Ergebnisbesprechung im Kollegium ($1.4 < |d| > 1.9$). Lehrpersonen ohne Erfahrung schätzen ihre Informiertheit über das Verfahren ($1.2 < |d| > 1.9$) aber auch die Nützlichkeit der Verfahren zur Unterrichtsentwicklung deutlich geringer ein ($0.9 < |d| > 1.7$). Analysen zur Zusammensetzung der Teilgruppen zeigen signifikante Unterschiede für alle drei Verfahren in der Berufserfahrung (Teilgruppe Erfahrung mit Verfahren: $21 < M > 23$ Jahre Berufserfahrung, Teilgruppe ohne Erfahrung mit Verfahren: $14 < M > 16$ Jahre Berufserfahrung; $.40 < |d| > .66$). Insgesamt zeigt sich, dass Lehrpersonen ohne Erfahrung mit den Verfahren deutlich weniger Wissen über die Ergebnisse und die Verfahren haben, diese aber trotz dieses Informationsdefizits als weniger nützlich bewerten. Deshalb werden in den folgenden Analysen nur Befragte einbezogen, die selbst Erfahrung mit den Verfahren haben, da diese Personengruppe validere Aussagen über die Verarbeitung und Verwendung von Ergebnissen aus VA, IE und ZP machen können.

Lehrpersonen mit Erfahrung mit den Verfahren berichten im Mittel eine eher intensive individuelle und kollaborative Auseinandersetzung mit allen Evaluationsergebnissen (Tab. 2). Es werden jedoch Unterschiede zwischen den Verfahren deutlich. Lehrpersonen setzen sich insbesondere mit ZP-Ergebnissen auseinander ($.40 < |d| > .59$). Signifikante Unterschiede zwischen VA und IE gibt es nicht. Hinsichtlich der Einschätzung der Verfahren zeigt sich ein ähnliches Bild. Lehrpersonen schätzen die Rückmeldungen als eher verständlich ein und fühlen sich eher gut über das Verfahren informiert. Die Nützlichkeit der Ergebnisse für Unterrichtsentwicklung wird in etwas geringerem Ausmaß moderat positiv bewertet. Im Verfahrensvergleich wird erneut eine positivere Bewertung des MSA deutlich ($.29 < |d| > .45$), jedoch keine bedeutsamen Unterschiede zwischen VA und IE. Die Hypothese, dass Lehrpersonen den Stellenwert von VA als Ausgangspunkt für ihre Unterrichtsentwicklung als niedriger ansehen als ZP und IE, trifft somit nur für den Unterschied zwischen VA und ZP zu, jedoch nicht für Unterschiede zwischen VA und IE. Der wahrgenommene Anlass zur Veränderung des eigenen Unterrichts rangiert für alle drei Verfahren in etwa auf dem Mittelwert der eingesetzten Skala.

6.2 *Unterrichtsentwicklung ausgehend von Ergebnissen aus Vergleichsarbeiten, zentralen Prüfungen und internen Evaluationen*

Zur Beantwortung der Forschungsfrage, ob ausgehend von Ergebnissen aus VA, MSA und IE Unterrichtsentwicklungsmaßnahmen abgeleitet werden, wird zunächst analysiert, wie viele Lehrpersonen mindestens eine der insgesamt fünf erfragten Maßnahmen berichten (Antwortkategorien: 4 (,trifft voll zu‘) und 3 (,trifft eher zu‘)). Alle drei Verfahren werden von der überwiegenden Mehrzahl der Lehrpersonen als Ausgangspunkt für datengestützte Unterrichtsentwicklung verwendet: 70% der Lehrpersonen

Item	Zuordnung Modell	VA			MSA			IE			Einfaktorielle ANOVA	
		N	M	SD	N	M	SD	N	M	SD	p	η^2
Ich habe mich mit den Ergebnissen intensiv auseinandergesetzt.	Individuelle Auseinandersetzung mit Ergebnissen	983	3,10 ^a	0,77	1001	3,40 ^b	0,67	1002	3,00 ^a	0,69	<.001	.05
Ich habe mit meinen Kolleg(inn)en die Ergebnisse besprochen	Besprechung Ergebnisse im Kollegium	983	3,17 ^a	0,74	1001	3,44 ^b	0,65	1002	3,11 ^a	0,68	<.001	.05
Ich fühle mich gut über die Evaluation informiert.	Informiertheit über Verfahren	983	2,98 ^a	0,78	996	3,25 ^b	0,72	991	2,93 ^a	0,72	<.001	.04
Die Rückmeldungen der Ergebnisse waren für mich verständlich.	Verständlichkeit Ergebnisse	979	3,05 ^a	0,79	985	3,27 ^b	0,72	995	3,04 ^a	0,66	<.001	.02
Für die Unterrichtsentwicklung waren die Ergebnisse nützlich.	Wahrgenommene Nützlichkeit	977	2,61 ^a	0,81	993	2,90 ^b	0,82	996	2,61 ^a	0,74	<.001	.03
Es gab keinen Anlass dazu, aufgrund der Ergebnisse meinen Unterricht zu ändern.	Anlass zur Veränderung	960	2,52 ^a	0,79	981	2,58 ^a	0,82	979	2,52 ^a	0,73	.26	.00
Aufgrund der Ergebnisse ...												
habe ich neue Unterrichtsmethoden eingeführt.	Unterrichtsentwicklung	948	2,22 ^a	0,75	981	2,28 ^a	0,76	993	2,30 ^a	0,72	.06	.00
habe ich Maßnahmen zur Einzelförderung entwickelt.	Unterrichtsentwicklung	949	2,30 ^a	0,77	980	2,35 ^a	0,78	991	2,36 ^a	0,72	.12	.00
habe ich an Fortbildungen teilgenommen.	Unterrichtsentwicklung	942	2,13 ^a	0,86	978	2,22 ^{ab}	0,88	987	2,3 ^b	0,83	<.001	.01
wurde am schulinternen Curriculum gearbeitet.	Unterrichtsentwicklung	929	2,45 ^a	0,89	961	2,51 ^a	0,88	980	2,67 ^b	0,86	<.001	.01
habe ich die Kommunikation mit meinen Kolleg(inn)en verbessert.	Unterrichtsentwicklung	943	2,35 ^a	0,83	978	2,42 ^a	0,84	987	2,38 ^a	0,78	<.001	.01

Anmerkungen. Vierstufige Skala von 1 = trifft gar nicht zu bis 4 = trifft völlig zu; ^{a/b}: ein ungleicher Buchstabe bedeutet einen signifikanten und ein gleicher Buchstabe keinen signifikanten Unterschied zwischen den Mittelwerten; Mittelwertsvergleich mittels multipler Paarvergleiche (t-Tests) mit Bonferroni-Korrektur.

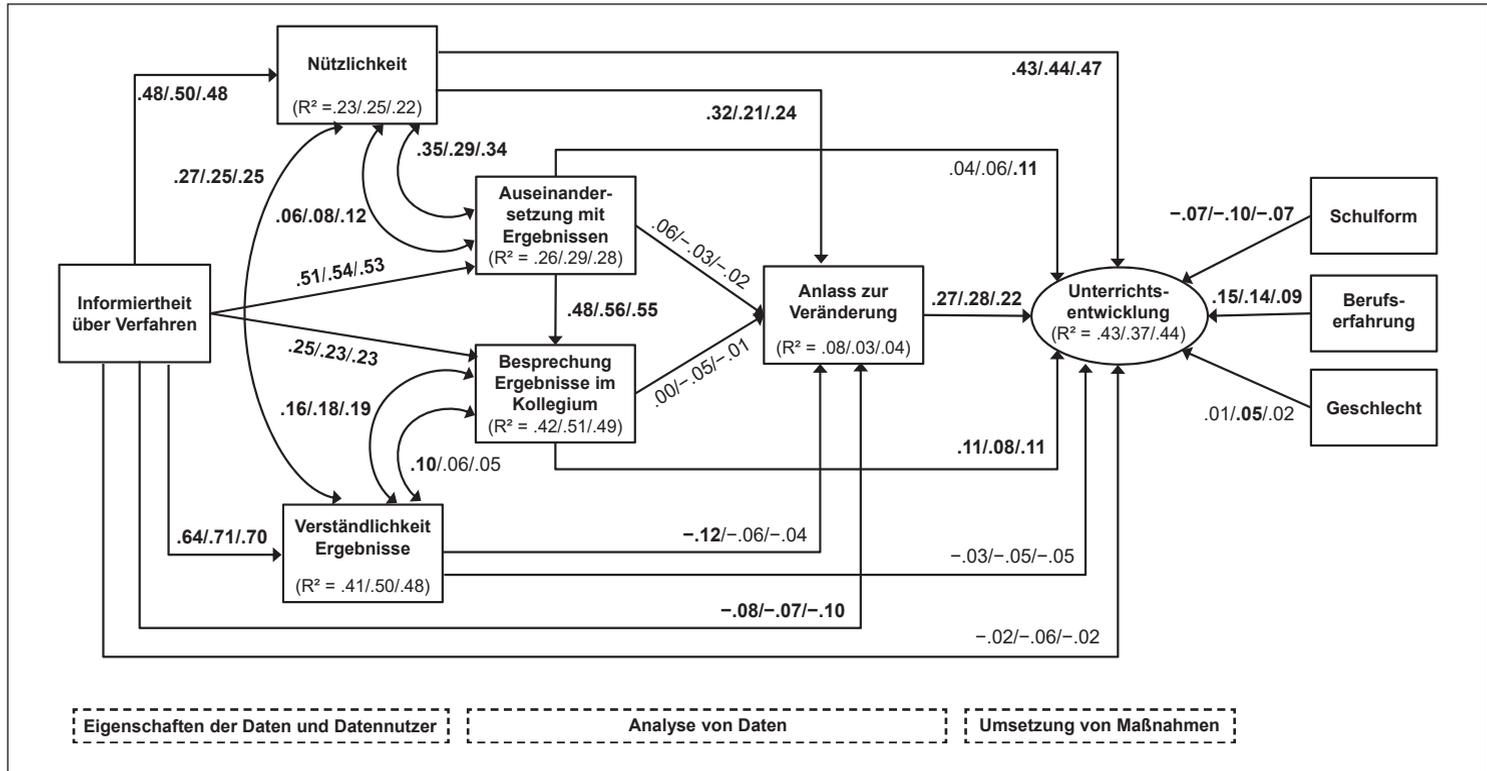
Tab. 2: Einschätzung, Verarbeitung und Verwendung von Ergebnissen aus internen Evaluationen, zentralen Prüfungen und Vergleichsarbeiten.

nutzen Ergebnisse aus VA, 75% Ergebnisse des MSA und 80% Ergebnisse aus IE. Die Ergebnisse weisen zudem darauf hin, dass wenige Lehrpersonen mehrere Entwicklungsmaßnahmen simultan durchführen. Die Mittelwerte der Einzelaktivitäten liegen alle Unterhalb des theoretischen Mittelwerts der Skala (Tab. 2). Einzige Ausnahme ist die Arbeit am schulinternen Curriculum auf Basis von IE, die jedoch auch eher moderat ausgeprägt ist. Die kontrastive Analyse der Maßnahmen zeigt lediglich für die Arbeit am schulinternen Curriculum und die Fortbildungsteilnahme kleine Unterschiede zwischen den Verfahren. Lehrpersonen nehmen eher ausgehend von Ergebnissen aus IE an Fortbildungen teil als aufgrund der VA-Ergebnisse ($d = .21$). Am schulinternen Curriculum wird ebenfalls eher aufgrund der Daten aus IE gearbeitet als ausgehend von VA- ($d = .26$) und MSA-Ergebnissen ($d = .18$). Die Unterschiede sind jedoch klein. Die Hypothese, dass es keine Unterschiede in der Nutzung bei Lehrpersonen zwischen den verschiedenen Verfahren gibt, trifft somit nur teilweise zu.

6.3 Einflussfaktoren datenbasierter Unterrichtsentwicklung

Die vorhandenen Items wurden anhand theoretischer Annahmen und des Forschungsstands (Coburn & Turner, 2011; Helmke, 2010; KMK, 2010; Schildkamp, Poortman, Luyten & Ebbeler, 2017) in einem Analysemodell systematisiert (Abb. 1). Der Prozess der Datennutzung kann mit mehreren Items operationalisiert werden. Die *Analyse von Daten* wird über die individuelle Auseinandersetzung mit den Daten und die Besprechung der Ergebnisse im Kollegium operationalisiert. Es wird angenommen, dass diese Verarbeitungsprozesse die Interpretation der Ergebnisse dahingehend, ob tatsächlich etwas am eigenen Unterricht geändert werden muss, vorhersagen. Diese Schlussfolgerung aus der Datenanalyse wird mit dem eingeschätzten Anlass zur Veränderung operationalisiert. Es wird in Anlehnung an den idealtypischen Verlauf der Datennutzungskreisläufe (Coburn & Turner, 2011; Helmke, 2010; KMK, 2010) ebenfalls angenommen, dass die Intensität der individuellen Auseinandersetzung die Besprechung der Ergebnisse im Kollegium voraussagt. Die *Umsetzung von Maßnahmen* wird durch das Konstrukt aus allen genannten Aktivitäten operationalisiert. Diese Unterrichtsentwicklungsaktivitäten werden durch die Items des Datenanalyseprozesses vorhergesagt.

Förderlichen und hinderlichen Bedingungen datengestützter Unterrichtsentwicklung können ebenfalls Items zugeordnet werden. Den Eigenschaften des Verfahrens bzw. der Daten und deren Bewertung durch die NutzerInnen lassen sich die Items zur Einschätzung der Informiertheit über das Verfahren, die Verständlichkeit der Ergebnisse und die Bewertung der Nützlichkeit zuordnen (Zuordnung bei Helmke (2010): „Individuelle Bedingungen“; Schildkamp, Poortman, Luyten & Ebbeler (2017, S. 244): „data“ und „user“; Visscher & Coe (2003, S. 331): „implementation process“, „school performance feedback systems features“). Ebenfalls in die Kategorie NutzerInnen fallen die Berufserfahrung und das Geschlecht der Lehrpersonen. Die Schulform ist Teil der Kategorie Organisation. Alle genannten Aspekte werden auf ihren Einfluss auf datengestützte Unterrichtsentwicklung hin analysiert. Die Informiertheit über das Verfahren ist ein Indika-



Anmerkungen. Dargestellt sind standardisierte Pfadkoeffizienten der drei Verfahren: β IE/ β MSA/ β VA; Fett gedruckte Koeffizienten (β) sind signifikant von 0 verschieden ($p < .05$); Manifeste Variablen sind in rechteckigen Kästen dargestellt und die jeweils latente Variable zur Unterrichtsentwicklung ist oval dargestellt. Für endogene Variablen wird zusätzlich die jeweils aufgeklärte Varianz (R^2) berichtet; Kodierung: Geschlecht (1 = weiblich, 0 = männlich), Schulform (0 = andere Schulformen, 1 = Gymnasium); Modellfit: IE: CFI = .96, TLI = .95, RMSEA = .05, SRMR = .04; MSA: CFI = .96, TLI = .94, RMSEA = .06, SRMR = .04; VA: CFI = .97, TLI = .95, RMSEA = .05, SRMR = .04.

Abb. 1: Einflussfaktoren datenbasierter Unterrichtsentwicklung.

tor für dessen Implementation. Deshalb dient dieses Item als Prädiktor für alle weiteren Einflussfaktoren. Die Beziehungen zwischen den verfahrens- und nutzerbezogenen Bedingungen sowie zu den Verarbeitungsprozessen wurden als Korrelationen modelliert. In Anlehnung an die *Theory of Planned Behavior* (Ajzen, 1991) wurde die wahrgenommene Nützlichkeit (Attitudes) als Prädiktor für den wahrgenommenen Veränderungsanlass (Intention) modelliert.

Die Modelle weisen einen guten bzw. vereinzelt akzeptablen Modellfit auf (Hu & Bentler, 1999). In Abb. 1 sind die standardisierten Pfadkoeffizienten aller drei Modelle dargestellt. Insgesamt lässt sich hypothesenkonform feststellen, dass die Prädiktoren bei allen Verfahren die gleiche Tendenz in Richtung und Stärke der Effekte aufweisen (max. Unterschied $\beta > .10$). Verfahrensübergreifend sind die bedeutsamsten Prädiktoren für *Maßnahmen datengestützter Unterrichtsentwicklung* die Nützlichkeit der Verfahren für Unterrichtsentwicklung (β IE = .43; β MSA = .44; β VA = .47) und der Anlass zur Veränderung (β IE = .27; β MSA = .28; β VA = .22). Entgegen der Modellannahmen ist die individuelle Auseinandersetzung nur für VA ein – wenn auch schwacher (β VA = .11) – signifikanter Prädiktor für die Umsetzung von Unterrichtsentwicklungsmaßnahmen. Die Besprechung der Ergebnisse im Kollegium ist zwar bei allen drei Verfahren ein signifikanter Prädiktor, der Einfluss ist jedoch ebenfalls klein (β IE = .11; β MSA = .08; β VA = .11). Das bedeutet, dass nicht die Intensität der individuellen Auseinandersetzung mit den Ergebnissen und Besprechung der Ergebnisse im Kollegium, sondern die daraus gezogenen Schlussfolgerungen (Anlass zur Veränderung) sowie die Nützlichkeit die bedeutsamsten Prädiktoren datengestützter Unterrichtsentwicklung sind. Schulform und Berufserfahrung zeigen ebenfalls für alle Verfahren kleine signifikante Effekte auf die Unterrichtsentwicklungsmaßnahmen. In Gymnasien wird etwas weniger datengestützte Unterrichtsentwicklung initiiert. Mehr Berufserfahrung geht mit einem höheren Maß an datengestützter Unterrichtsentwicklung einher. Das Geschlecht hat lediglich für den MSA einen sehr kleinen signifikanten Effekt.

Eine Analyse der Beziehung der Items zu den *Verarbeitungsprozessen* offenbart einerseits erwartungsgemäß, dass die Intensität der individuellen Auseinandersetzung mit den Ergebnissen die Besprechung dieser im Kollegium voraussagt (β IE = .48; β MSA = .56; β VA = .55). Andererseits hat das Ausmaß der individuellen Auseinandersetzung mit den Ergebnissen und die Besprechung der Ergebnisse im Kollegium keinen Effekt auf den wahrgenommenen Anlass zur Veränderung. Die Schlussfolgerung aus der Ergebnisinterpretation dahingehend, ob tatsächlich etwas am eigenen Unterricht geändert werden muss, ist im Modell unabhängig von der Intensität der Ergebnisrezeption und -reflexion. Insgesamt wird nur wenig Varianz des Items Veränderungsanlass erklärt (R^2 IE = .08; R^2 MSA = .03; R^2 VA = .04). Der relevanteste Prädiktor des Veränderungsanlasses ist die Nützlichkeit des Verfahrens für Unterrichtsentwicklung (β IE = .32; β MSA = .21; β VA = .24). Je nützlicher das Verfahren eingeschätzt wird, desto eher sehen Lehrpersonen einen Anlass, etwas an ihrem Unterricht zu verändern. Je informierter sich Lehrpersonen einschätzen, desto weniger Veränderungsanlass wird gesehen. Dies ist allerdings ein sehr kleiner Effekt (β IE = -.08; β MSA = -.07; β VA = -.10). Einen zwar signifikanten, aber sehr kleinen negativen Effekt hat die Verständlich-

keit der Ergebnisse aus IE auf den Veränderungsanlass ($\beta = -.12$). Zudem korreliert die Nützlichkeit der Verfahren mit der individuellen Auseinandersetzung mit den Ergebnissen ($r_{IE} = .35$; $r_{MSA} = .29$; $r_{VA} = .34$), jedoch nicht substantziell mit der Besprechung der Ergebnisse im Kollegium.

Die eingeschätzte *Informiertheit* über das Verfahren ist ein weiterer wesentlicher Prädiktor für die Verarbeitung der Ergebnisse. Ein höherer Grad an Informiertheit führt insbesondere zu einer intensiveren individuellen Auseinandersetzung mit den Ergebnissen ($\beta_{IE} = .51$; $\beta_{MSA} = .54$; $\beta_{VA} = .53$) und begünstigt die Besprechung der Ergebnisse im Kollegium ($\beta_{IE} = .25$; $\beta_{MSA} = .23$; $\beta_{VA} = .23$). Die zentrale Rolle der Informiertheit wird zudem durch den großen Effekt auf die Verständlichkeit der Ergebnisse ($\beta_{IE} = .64$; $\beta_{MSA} = .71$; $\beta_{VA} = .70$) und die Nützlichkeit ($\beta_{IE} = .48$; $\beta_{MSA} = .50$; $\beta_{VA} = .48$) deutlich. Die Relevanz der Informiertheit für datengestützte Unterrichtsentwicklung zeigt sich zwar nicht über einen direkten Effekt auf Entwicklungsmaßnahmen, es zeigen sich jedoch indirekte Effekte. Die Informiertheit hat für alle drei Verfahren einen indirekten Effekt auf datengestützte Unterrichtsentwicklung, insbesondere mediiert über die Nützlichkeit (indirekt: $\beta_{IE} = .16$; $\beta_{MSA} = .17$; $\beta_{VA} = .17$, total: $\beta_{IE} = .16$; $\beta_{MSA} = .17$; $\beta_{VA} = .17$). Die Mediation über die Kommunikation im Kollegium und den Anlass zur Veränderung zeigen zwar teilweise ebenfalls signifikante jedoch sehr kleine indirekte und nicht signifikante totale Effekte. Alle anderen indirekten Effekte sind nicht signifikant.

7. Diskussion

Lehrpersonen, die nicht in den jeweiligen Verfahren involviert waren, haben deutlich weniger Wissen über die Verfahren und Ergebnisse, bewerten diese aber trotz dieses Informationsdefizits als weniger nützlich (Forschungsfrage 1). Deshalb ist es sinnvoll, bei der Analyse von Verfahren datengestützter Unterrichtsentwicklung die Erfahrungshorizonte der Befragten zu ermitteln und in die Interpretation einfließen zu lassen. Die folgenden Ergebnisse beruhen demzufolge auf den Angaben von Lehrpersonen mit Erfahrung mit den jeweiligen Instrumenten.

Lehrpersonen bewerten VA, MSA und IE hinsichtlich ihres Potenzials zur datengestützten Unterrichtsentwicklung moderat positiv (Forschungsfrage 1). Sie berichten im Mittel eine eher intensive individuelle und kollaborative Auseinandersetzung mit allen Ergebnissen. Unterschiede zwischen den Verfahren bestehen dahingehend, dass der MSA positiver eingeschätzt und intensiver rezipiert wird als VA und IE. Bedeutsame Unterschiede zwischen VA und IE können nicht festgestellt werden. Daten aus VA, dem MSA und IE werden von der überwiegenden Mehrzahl der Lehrpersonen als Ausgangspunkt für datengestützte Unterrichtsentwicklung verwendet (Forschungsfrage 2): 70 bis 80 % der Lehrkräfte berichten mindestens eine datenbasierte Unterrichtsentwicklungsaktivität (VA: 70 %, MSA: 75 % und IE: 80 %). Nur wenige Lehrpersonen führen mehrere Entwicklungsmaßnahmen simultan durch. Die Mittelwerte der einzelnen Aktivitäten liegen alle unterhalb des theoretischen Mittelwerts der Skala. Substanzielle Un-

terschiede in der Ergebnisnutzung zwischen den drei Verfahren hinsichtlich der Einzelaktivitäten konnten nicht festgestellt werden.

Bedeutsame Prädiktoren für die Ergebnisnutzung sind für alle untersuchten Verfahren die eingeschätzte Nützlichkeit der Verfahren, der wahrgenommene Veränderungsbedarf und in geringerer Ausprägung die Besprechung der Ergebnisse im Kollegium (Forschungsfrage 3). Einen indirekten Effekt hat zudem die eingeschätzte Informiertheit über das Verfahren. Relevante Unterschiede zwischen den Verfahren hinsichtlich der Prädiktoren gibt es nicht.

Aufgrund der großen Überschneidung der Stichprobe, die in den hier berichteten Analysen um Lehrpersonen naturwissenschaftlicher Fächer ergänzt wurden, zeigen sich sehr ähnliche Befunde zu Unterrichtsentwicklungsaktivitäten bei Wurster, Richter und Lenski (2017). Die Befunde zur Auseinandersetzung mit den Ergebnissen und Bewertung der Verfahren finden sich in ähnlicher Form für VA (z. B. Groß Ophoff, Koch & Hosenfeld, 2019) und IE (z. B. Demski, 2017). Die Bedeutung von Ergebnissen aus ZP zur Initiierung von Unterrichtsentwicklung zeigt sich ebenfalls für Fachkonferenzleitungen (Wurster & Richter, 2016). Die Rolle von Ergebnissen aus ZP zur Unterrichtsentwicklung sollte deshalb von der Bildungsadministration unter Berücksichtigung des Stellenwerts der Abschlüsse für verschiedene Schulformen (z. B. MSA vs. Abitur) aufgegriffen und verstärkt werden (z. B. MBK, 2011). Unterschiede zwischen den Verfahren berichtet auch Demski (2017): Verfahren der IE werden im Vergleich zu VA von Lehrpersonen als nützlicher bewertet, intensiver rezipiert und sind häufiger Ausgangspunkt von Entwicklungsvorhaben. Das entspricht der Tendenz der hier präsentierten Befunde zu den Anteilen der Nutzung. Auf Ebene der Einzelaktivitäten zeigen sich jedoch keine substanziellen Unterschiede. Mögliche Erklärungen für die verschiedenen Befunde liegen in der Zusammensetzung der Stichprobe. Einerseits bleibt unklar, ob die befragten Lehrpersonen bereits Erfahrung mit dem Verfahren gesammelt haben. Andererseits wurden Lehrpersonen aller Fächer befragt und Lehrpersonen mit mathematisch-naturwissenschaftlichem Hintergrund schätzen VA tendenziell nützlicher ein (Maier, Metz, Bohl, Kleinknecht & Schymala, 2012).

Die Ergebnisse verweisen auf zwei wichtige Dimensionen förderlicher Bedingungen datengestützter Unterrichtsentwicklung auf Basis von VA, ZP und IE: die Ausgestaltung des Verfahrens und die Analyse der zurückgemeldeten Ergebnisse. Die Bedeutung der Implementation und Ausgestaltung des Verfahrens (siehe auch Schildkamp, Poortman, Luyten & Ebbeler, 2017; Visscher & Coe, 2003) wird durch die zentrale Rolle der eingeschätzten Nützlichkeit der Verfahren deutlich. Sie hat den größten Effekt auf die Unterrichtsentwicklungsmaßnahmen und den wahrgenommenen Veränderungsanlass. Zudem wird die Nützlichkeit insbesondere vom Informationsgrad über das Verfahren prädiert. Die zentrale Rolle der Akzeptanz bzw. Bewertung der Nützlichkeit zeigt sich auch in anderen deutschsprachigen Studien zu VA und ZP (Groß Ophoff, 2013; Wurster & Richter, 2016) sowie internationalen Befunden zu IE (Schildkamp & Visscher, 2009; Vanhoof, Van Petegem & De Maeyer, 2009).

Die zweite wesentliche Dimension ist die Datenanalyse durch die Lehrpersonen. Die Ergebnisse aller drei Verfahren werden eher intensiv rezipiert. Allerdings sind nicht die

Intensität der individuellen und kollaborativen Analyse, sondern die daraus gezogenen Schlussfolgerungen (Anlass zur Veränderung) der bedeutsamste Prädiktor datengestützter Unterrichtsentwicklung im Bereich Verarbeitung der Ergebnisse. Diese Konstellation wirft die Frage auf, warum die Intensität der individuellen Ergebnisverarbeitung keinen direkten Einfluss auf das Interpretationsergebnis (Anlass zur Veränderung) und die Umsetzung von Maßnahmen hat. Mit den erfassten Variablen konnte nur ein kleiner Ausschnitt der Analyseprozesse untersucht werden. Möglicherweise ist der Auslöser datengestützter Unterrichtsentwicklung nicht allein in einer intensiven Auseinandersetzung mit den Daten zu sehen. Die Ergebnisse könnten durch Lehrpersonen möglicherweise ohne intensive Analyse ‚auf einen Blick‘ interpretiert werden. Ein weiterer Grund für den fehlenden Zusammenhang könnte auch sein, dass die Ergebnisse aus Sicht der Lehrpersonen so gut ausfallen, dass sie keinen Veränderungsanlass sehen. Auch die Attribution der Gründe für das jeweilige Ergebnis könnte Einfluss darauf haben. Lehrpersonen sehen weniger Veränderungsanlass, wenn sie die Ursache bspw. auf Seite der Lernenden und nicht auf den eigenen Unterricht attribuieren (Bertrand & Marsh, 2015).

Das verweist auf einen bislang weitgehend unerforschten Prozess der Entscheidungsfindung, der nicht nur die Daten, sondern auch Intuition (Vanlommel, Van Gasse, Vanhoof & Van Petegem, 2017), „Evidenzen qua Augenschein“ (Heinrich, 2015, S. 785) und komplexe Prozesse des „sensemaking“ (Bertrand & Marsh, 2015, S. 862) mit einbezieht. Hier könnte eine Schwäche der bisherigen Rezeptionsmodelle deutlich werden, deren Grundannahme ist, dass die Prozesse linear verlaufen und Entscheidungen primär rational auf Basis bereitgestellter oder zusätzlich herangezogener Informationen getroffen werden (Hyyryläinen & Viinamäki, 2008).

Weitergehende Forschung sollte sich deshalb a) der Analyse von Mikroprozessen der Dateninterpretation und Entscheidungsfindung in Schulen widmen (Little, 2012). Es ist unklar, wie die Schlussfolgerungen aus den Ergebnissen und der wahrgenommene Anlass zur Veränderung zustande kommen und welche Rolle die Ergebnisse dafür haben. Es bedarf ebenfalls b) der Analyse weiterer bedeutsamer Bedingungen der Datennutzung, wie z. B. Unterstützungssystemen und der Data Literacy (Mandinach & Gummer, 2016) von Schulleitungen und Lehrpersonen. Zudem fehlen Längsschnittanalysen zur c) Untersuchung des Zusammenhangs zwischen Datennutzungspraxis sowie organisationalen Merkmalen der Schulen und (Entwicklung der) Unterrichtsqualität sowie SchülerInnenleistung. Ein weiteres bedeutsames Forschungsdesiderat betrifft d) die Möglichkeiten der Integration verschiedener Datenbestände, da sich zeigt, dass zur Ableitung von Entwicklungsaktivitäten mehrere Datenquellen simultan genutzt werden (Wurster, Richter & Lenski, 2017; Wurster & Richter, 2016).

Die Ergebnisse legen Implikationen für die Durchführungspraxis der Verfahren nahe. Das Informationsdefizit bestimmter Lehrpersonengruppen sollte in der Aus- und Fortbildung sowie im Rahmen weiterer Implementationsbemühungen aufgegriffen werden. Dazu gehören die verbindliche Verankerung von Veranstaltungen zum Thema Data Literacy in allen Phasen der Aus- und Fortbildung von Lehrkräften und die gezielte Adressierung der Überzeugungen der Lehrpersonen, wie die Relevanz des Prädiktors Nützlichkeit verdeutlicht (siehe auch Datnow & Hubbard, 2016). Der Umgang mit Da-

ten muss als Teil professionellen Handelns von Lehrpersonen begriffen werden. Zudem sind gezielte Unterstützungsangebote bei der Arbeit mit Daten, z. B. durch die Etablierung kooperativer Strukturen zur Datennutzung oder die Unterstützung durch Schulleitungen sinnvoll (Datnow & Hubbard, 2016; Hoogland et al., 2016; Schildkamp, Poortman, Luyten & Ebbeler, 2017). Die Ergebnisse zeigen ferner, dass hier besonders Lehrkräfte mit weniger Berufserfahrung Unterstützung benötigen. Ein weiterer Ansatzpunkt ist die Gestaltung der Rückmeldungen, um die Rezeption von Informationen zu unterstützen (Merk, Poindl & Bohl, 2019).

Die Studie weist einige Limitationen auf. Für die Sekundäranalysen standen Querschnittsdaten basierend auf Selbstauskünften zur Verfügung, die keine Aussagen zur Relevanz der berichteten Maßnahmen zur Steigerung der Unterrichtsqualität ermöglichen. Sekundäranalysen gehen zudem mit Einschränkungen bezüglich der Auswahl der untersuchbaren Variablen einher. Dies zeigt sich zum einen in hier nicht analysierbaren potenziell wichtigen Faktoren (Hoogland et al., 2016; Schildkamp, Poortman, Luyten & Ebbeler, 2017). Zum anderen wurden viele Aspekte mit Einzelitems erfasst, obwohl eine Operationalisierung als latente Konstrukte vorteilhaft wäre. Dennoch haben auch Einzelitems bei klar definierten Sachverhalten eine ausreichend gute Aussagekraft (Bergkvist & Rossiter, 2007). Die Analysen ohne latente (messfehlerfreie) Konstrukte führen eher zu einer Unterschätzung der Effekte. Eine weitere Einschränkung ist die unberücksichtigte unterschiedliche Implementation der Verfahren in den Bundesländern, die zu unterschiedlichen Bewertungen führen können (Maier, 2008). Offen bleibt auch, inwiefern an Gymnasien der MSA im Vergleich zum Abitur einen anderen Stellenwert als Datengrundlage hat. In zukünftigen Studien sollten zudem Verfahren der IE differenziert erfasst werden. Dennoch ist die Datenbasis eines Large-Scale-Assessments mit den genannten Einschränkungen aufgrund der Datenqualität aussagekräftig.

Zusammenfassend lässt sich festhalten, dass die überwiegende Mehrheit der Lehrpersonen Unterrichtsentwicklung ausgehend von Ergebnissen aus VA, ZP und IE berichtet. Förderliche Bedingungen dafür sind die Bewertung der Verfahren hinsichtlich der Nützlichkeit, der aus den Ergebnissen geschlussfolgerte Veränderungsanlass und – als indirekte Bedingung – die Implementation bzw. Informiertheit über die Verfahren.

Literatur

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T).
- Altrichter, H., Moosbrugger, R., & Zuber, J. (2016). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (2. Aufl., S. 235–277). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18942-0_9.
- 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. <https://doi.org/10.1007/s11618-014-0486-5>.

- Bergkvist, L., & Rossiter, J.R. (2007). The Predictive Validity of Multiple-Item versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, 44(2), 175–184. <https://doi.org/10.2307/30162466>.
- Berkemeyer, N., Müller, S., & Van Holt, N. (2016). Schulinterne Evaluation – nur ein Instrument zur Selbststeuerung von Schulen? In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (2. Aufl., S. 209–234). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18942-0_8.
- Bertrand, M., & Marsh, J.A. (2015). Teachers' Sensemaking of Data and Implications for Equity. *American Educational Research Journal*, 52(5), 861–893. <https://doi.org/10.3102/0002831215599251>.
- Bischof, L.M., Hochweber, J., Hartig, J., & Klieme, E. (2013). Schulentwicklung im Verlauf eines Jahrzehnts. Erste Ergebnisse des PISA-Schulpanels. In N. Jude & E. Klieme (Hrsg.), *PISA 2009. Impulse für die Schul- und Unterrichtsforschung* (59. Beiheft der Zeitschrift für Pädagogik, S. 172–199). Weinheim/Basel: Beltz Juventa.
- Coburn, C.E., & Turner, E.O. (2011). Research on Data Use. A Framework and Analysis. *Measurement: Interdisciplinary Research and Perspectives*, 9(4), 173–206. <https://doi.org/10.180/15366367.2011.626729>.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155–159.
- Creemers, B.P.M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness. A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge.
- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making. A literature review of international research. *Journal of Educational Change*, 17(1), 7–28. <https://doi.org/10.1007/s10833-015-9264-2>.
- Demski, D. (2017). *Evidenzbasierte Schulentwicklung*. Wiesbaden: Springer Fachmedien.
- Diemer, T., & Kuper, H. (2011). Formen innerschulischer Steuerung mittels zentraler Lernstandserhebungen. *Zeitschrift für Pädagogik*, 57(4), 554–571.
- Gärtner, H. (2013). Praxis und Verhältnis interner und externer Evaluation im Schulsystem im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 16(4), 693–712. <https://doi.org/10.1007/s11618-013-0444-7>.
- Gärtner, H. (2014). Effects of student feedback as a method of self-evaluating the quality of teaching. *Studies in Educational Evaluation*, 42, 91–99. <https://doi.org/10.1016/j.stueduc.2014.04.003>.
- Gärtner, H., & Vogt, A. (2013). Wie Lehrkräfte Ergebnisse eines Schülerfeedbacks verarbeiten und nutzen. *Unterrichtswissenschaft*, 41(3), 252–267.
- Groß Ophoff, J. (2013). *Lernstandserhebungen. Reflexion und Nutzung*. Münster: Waxmann.
- Groß Ophoff, J., Koch, U., & Hosenfeld, I. (2019). Vergleichsarbeiten in der Grundschule von 2004 bis 2015. In J. Zuber, H. Altrichter & M. Heinrich (Hrsg.), *Bildungsstandards zwischen Politik und schulischem Alltag* (S. 205–228). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-22241-3_9.
- Heinrich, M. (2015). Zur Ambivalenz der Idee evidenzbasierter Schulentwicklung. Das Beispiel „Schulinspektion“ – Fortschrittlicher Rückschritt oder Innovation? *Zeitschrift für Pädagogik*, 61(6), 778–792.
- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Hoogland, I., Schildkamp, K., Van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A.M. (2016). Prerequisites for data-based decision making in the classroom. Research evidence and practical illustrations. *Teaching and Teacher Education*, 60, 377–386. <https://doi.org/10.1016/j.tate.2016.07.012>.

- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hyyryläinen, E., & Viinamäki, O.-P. (2008). The Implications of the Rationality of Decision-Makers on the Utilization of Evaluation Findings. *International Journal of Public Administration*, 31(10-11), 1223–1240. <https://doi.org/10.1080/01900690801973311>.
- Klein, E. D., Krüger, M., Kühn, S. M., & Van Ackeren, I. (2014). Wirkungen zentraler Abschlussprüfungen im Mehrebenensystem Schule. Eine Zwischenbilanz internationaler und nationaler Befunde und Forschungsdesiderata. *Zeitschrift für Erziehungswissenschaft*, 7–33. <https://doi.org/10.1007/s11618-014-0479-4>.
- KMK (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Wolters Kluwer. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_00_00-Konzeption-Bildungsstandards.pdf [14.06.2021].
- KMK (2015). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf [14.06.2021].
- 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.
- Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors. Implications for theory and research. *British Educational Research Journal*, 36(5), 807–830. <https://doi.org/10.1080/01411920903165603>.
- Lenski, A. E., Hecht, M., Penk, C., Milles, F., Mezger, M., Heitmann, P., Stanat, P., & Pant, H.A. (2016). *IQB-Ländervergleich 2012. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen. doi:10.20386/HUB-42547.
- Little, J.W. (2012). Understanding Data Use Practice among Teachers. The Contribution of Micro-Process Studies. *American Journal of Education*, 118(2), 143–166. <https://doi.org/10.1086/663271>.
- Maag Merki, K. (2016). 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* (2. Aufl., S. 151–181). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18942-0_6.
- 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., 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 Verlag für Sozialwissenschaften.
- Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate. Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, 60, 366–376. <https://doi.org/10.1016/j.tate.2016.07.011>.
- MBK (2011). *Leitfaden zur Nutzung der Ergebnisse zentraler Prüfungen*. Kiel: Ministerium für Bildung und Kultur des Landes Schleswig-Holstein.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Mahwah, NJ: Erlbaum.

- Merk, S., Poindl, S., & Bohl, T. (2019). Wie sollten Rückmeldungen von quantitativ erfasstem Schülerfeedback (nicht) gestaltet werden? Wahrgenommene Informativität und Interpretationssicherheit von quantitativen Rückmeldungen zur Unterrichtsqualität. *Unterrichtswissenschaft, 47*, 475–494. <https://doi.org/10.1007/s42010-019-00048-5>.
- Nachtigall, C., & Hellrung, K. (2013). Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. *Empirische Pädagogik, 27*(4), 423–441.
- Pant, H. A., Stanat, P., Hecht, M., Heitmann, P., Jansen, M., Lenski, A. E., Penk, C., Pöhlmann, C., Roppelt, A., Schroeders, U., & Siegle, T. (2015). *IQB-Ländervergleich Mathematik und Naturwissenschaften 2012 (IQB-LV 2012). Version: 3*. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_LV_2012_v3.
- 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.
- Richter, D., & Stanat, P. (2019). Bildungsstandards und Kompetenzorientierung. In M. Haring, C. Rohlf & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (S. 698–708). Münster: Waxmann.
- Rosseel, Y. (2012). Lavaan. An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(1), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Schildkamp, K., Lai, M. K., & Earl, L. (2013) (Hrsg.). *Data-based Decision Making in Education. Challenges and Opportunities*. Dordrecht: Springer.
- Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement, 28*(2), 242–258. <https://doi.org/10.1080/09243453.2016.1256901>.
- Schildkamp, K., & Visscher, A. (2009). Factors influencing the utilisation of a school self-evaluation instrument. *Studies In Educational Evaluation, 35*(4), 150–159. <https://doi.org/10.1016/j.stueduc.2009.12.001>.
- Stump, M., Zlatkin-Troitschanskaia, O., & Mater, O. (2016). The effects of transformational leadership on teachers' data use. *Journal for Educational Research Online/Journal Für Bildungsforschung Online, 8*(3), 80–99.
- Thiel, F., Tarkian, J., Lankes, E.-M., Maritzen, N., Riecke-Baulecke, T., & Kroupa, A. (Hrsg.) (2019). *Datenbasierte Qualitätssicherung und -entwicklung in Schulen. Eine Bestandsaufnahme in den Ländern der Bundesrepublik Deutschland*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-23240-5>.
- Thiel, F., & Thillmann, K. (2012). Interne Evaluation als Instrument der Selbststeuerung von Schulen. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikation* (S. 35–55). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Vanhoof, J., Van Petegem, P., & De Maeyer, S. (2009). Attitudes towards school self-evaluation. *Studies In Educational Evaluation, 35*(1), 21–28. <https://doi.org/10.1016/j.stueduc.2009.01.004>.
- Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making. Data based or intuition driven? *International Journal of Educational Research, 83*, 75–83. <https://doi.org/10.1016/j.ijer.2017.02.013>.
- Visscher, A. J., & Coe, R. (2003). School Performance Feedback Systems. Conceptualisation, Analysis, and Reflection. *School Effectiveness and School Improvement, 14*(3), 321–349.
- Wurster, S. (2019). Evaluationsgestützte Schul- und Unterrichtsentwicklung. In M. Haring, M. Gläser-Zikuda & C. Rohlf (Hrsg.), *Handbuch Schulpädagogik* (S. 765–776). Münster: Waxmann.

- Wurster, S., & Richter, D. (2016). Nutzung von Schülerleistungsdaten aus Vergleichsarbeiten und zentralen Abschlussprüfungen für Unterrichtsentwicklung in Fachkonferenzen. *Journal for Educational Research Online*, 8(3), 159–183.
- Wurster, S., Richter, D., & Lenski, A. (2017). Datenbasierte Unterrichtsentwicklung durch Lehrkräfte und Zusammenhänge zur Schülerleistung. *Zeitschrift für Erziehungswissenschaft*, 20(4), 628–650. <https://doi.org/10.1007/s11618-017-0759-x>.

Abstract: Data from state-wide comparison tests, centralized examinations and self-evaluations are available to teachers as a starting point for data-based improvement of instruction. This data use, as well as fostering and hindering factors, were analyzed based on the German large-scale national assessment study 2012 ($N = 3099$ teachers). The three data sources are evaluated moderately positively, and their results are widely noticed, interpreted, and used by a substantial part of the teaching staff for data-based improvement of instruction. The most important predictors for data use are the perceived usefulness of the data sources and the perceived need for change as well as, in part, the discussion of the results in the teaching staff.

Keywords: Self-Evaluation, State-wide Comparison Tests, Centralized Examinations, Data Use, Teacher, Improvement of Instruction

Anschrift des Autors

Dr. Sebastian Wurster, Johannes Gutenberg-Universität Mainz,
Institut für Erziehungswissenschaft,
Jakob-Welder-Weg 12, 55128 Mainz, Deutschland
E-Mail: wurster@uni-mainz.de