

Stefan Hahn, Stanislav Ivanov & Roumiana Nikolova

Innovation und Transfer brauchen mehr als Daten zum Output

Erkenntnisse aus der begleiteten internen Evaluation mit KESS

Zusammenfassung

*In diesem Beitrag werden die Anlage und die Unterstützungsstrukturen der in Hamburg angebotenen schulinternen Evaluation mit KESS vorgestellt, und es wird anhand der Erfahrungen des Evaluationsteams exemplarisch aufgezeigt, in welcher Variationsbreite das längsschnittlich normierte Instrumentarium aus der Hamburger Schulleistungsstudie Kompetenzen und Einstellungen von Schülerinnen und Schülern (KESS) von Einzelschulen zu Zwecken der eigenverantwortlichen Schul- und Unterrichtsentwicklung genutzt wird. Die Erhebung primärer Daten mit dem KESS-Instrumentarium ermöglicht es, viele schulspezifische Fragestellungen in vielfältigen Evaluationsdesigns zu bearbeiten. Die Schulen folgen bei der Datennutzung häufig der kybernetischen Vorstellung einer funktionalen Optimierung bestehender Routinen. Den Schulen wird aber auch eine grundlegende Veränderung der Prozessmuster durch ein spezifisches Feedback des externen Evaluationsteams an schulische Akteure eröffnet, das die Rückmeldung von Kompetenz- und Einstellungsdaten von Schüler*innen mit einer gemeinsamen Interpretation dieser Ergebnisse verknüpft. Insbesondere die im Zuge dieser Interpretation erfolgende Kontextualisierung von Daten zum Output mit qualitativen Prozessdaten und dem Erfahrungswissen von Lehrkräften markiert den Kern einer wissenschaftlichen Begleitung von Prozessen des Transfers pädagogischer Prototypen im Sinne einer designbasierten Schulentwicklung.*

Schlüsselwörter: begleitete interne Evaluation, datengestützte Schulentwicklung, designbasierte Schulentwicklung, Transfer, KESS

Innovation and Transfer Need More than Output-Data

Insights from the Guided Internal Evaluation with KESS

Abstract

This article presents the design of the internal evaluation with KESS in Hamburg. It depicts in an exemplary manner how particular schools use the longitudinal standardized

assessment program from the KESS-study for the purpose of facilitating teaching and school quality development. Results show that surveying primary assessments with the KESS-instruments according to specific school-based research questions, and within a variety of evaluation designs, motivates general data use and is important for initiating school quality improvement processes. Furthermore, results show that the reported internal school evaluation does not automatically determine the mode of data use. Schools often do not just follow a cybernetic idea of functional optimization. Rather, a specific feedback system allows the different stakeholders to even change patterns of teaching and learning more fundamentally. Within this feedback system, the report of data regarding students' cognitive aptitudes, competences and attitudes is combined with a joint interpretation of these findings by professionals from the school and the external evaluation team. The contextualization of the output-data, which is accumulated and combined with qualitative process-data and background knowledge gained out of practice, proves to be especially valuable during the interpretation by the stakeholders. This process characterizes the genuine scientific support of transferring pedagogical prototypes in the sense of a design-based school improvement.

Keywords: guided internal evaluation, data-based school improvement, design-based school improvement, transfer, KESS

1. Einleitung

Daten aus Lernstandserhebungen sollen Schulen Informationen über schulisch erworbene Kompetenzen ihrer Schüler*innenschaft bereitstellen und bieten mit Referenzwerten und/oder Kompetenzmodellen externe Maßstäbe für die Einordnung der erzielten schulischen Leistungen an. Schulen bekommen also eine Rückmeldung darüber, inwiefern die von ihren Schüler*innen erzielten Ergebnisse einer bildungswissenschaftlich definierten Norm entsprechen, und sollen diese – so wünschen es sich die Auftraggeber – zum Anlass nehmen, die Lehr-Lern-Prozesse durch gezielte schulische und unterrichtliche Maßnahmen zu verbessern. Auch wenn die Datennutzung durch Schulen insgesamt noch als unbefriedigend angesehen werden muss (vgl. Altrichter, Moosbrugger & Zuber, 2016), wird aktuell diskutiert (vgl. Wiesner & Schreiner, 2019), ob einer regelmäßigen Bereitstellung von Schulleistungsdaten eher die kybernetische Vorstellung einer Funktionsoptimierung (Kruse, 2004) innewohnt und sich Schulentwicklung weniger an Modellen des Veränderungslernens (Argyris & Schön, 1999) orientiert und damit grundlegenden Innovationen in der Organisation und Durchführung des Unterrichts möglicherweise sogar im Wege steht. Ohne entweder die Funktionsoptimierung noch einen mit Veränderungslernen assoziierten Prozessmusterwechsel als in irgendeiner Weise überlegenen Modus der datengestützten Schulentwicklung darstellen zu wollen, wird in diesem Beitrag auf diese Idealtypen zurückgegriffen, um die Erfahrungen aus der intensiven Zusammenarbeit zwischen dem *Institut für Bildungsmonitoring*

und *Qualitätsentwicklung (IfBQ)* und jenen Hamburger Schulen zu beschreiben, die sich bei Vorhaben der internen Evaluation über mehrere Jahre hinweg vom IfBQ mit dem Instrumentarium aus der Schulleistungsstudie KESS (**K**ompetenzen und **E**instellungen von Schülerinnen und Schülern) begleiten lassen. Wie im Weiteren dargelegt wird, legt die Reflexion der Kooperationserfahrungen dreierlei nahe: (1) Schulen können durch die Rückmeldung von primären Leistungs-, Einstellungs- und sozialen Hintergrunddaten anhand schulexterner Kriterien erkennen, inwiefern sie ihren Bildungsauftrag erfüllen. (2) Sie können Schwerpunkte für die weitere Entwicklung legen, weil sie in einem kommunikativen Prozess durch eine gemeinsame Interpretation der Daten und deren Einordnung in die praktische Handlungstheorie der Schule unterstützt werden. (3) In einzelnen Fällen wird deutlich, dass eine datengestützte Schul- und Unterrichtsentwicklung auch zu einem grundlegenden Wechsel der Prozessmuster einer Schule führen kann.

2. Datengestütztes Feedback auf allen Ebenen

Neuere Ansätze zur Governance im Schulsystem (Altrichter & Maag Merki, 2016) haben darauf aufmerksam gemacht, dass Schulen von vielen Akteuren mit multiplen Interessen „gesteuert“ werden. Diese Akteure agieren auf unterschiedlichen, jeweils durch spezifische Eigenlogiken und Eigendynamiken gekennzeichneten Handlungsebenen des Bildungssystems (die Handlungsebenen der Bildungspolitik, der Bildungsverwaltung, der Einzelschule, des Lehrens und Unterrichtens und der Schüler*innen). Diese Ebenen sind miteinander verschränkt, sodass Entscheidungen auf höherer Ebene den Handlungskontext auf der Ebene darunter mitbestimmen, diesen aber nicht vollständig determinieren, da Vorgaben den Handlungsbedingungen auf der unteren Ebene entsprechend immer re-kontextualisiert werden müssen. Mit dem Konzept der Re-Kontextualisierung macht Fend (2008) darauf aufmerksam, dass auf verschiedenen Ebenen des Bildungswesens jeweils eigene Handlungsaufgaben bestehen, die eigene Handlungsinstrumente, Kompetenzen und Verantwortungen erfordern.

Das IfBQ bearbeitet im Hamburger Bildungswesen die Mehrebenenproblematik der Steuerung durch datengestütztes Feedback auf verschiedenen Handlungsebenen. Vom Anspruch her dienen die vielfältigen Daten aus Schulinspektion, Lernstandserhebungen, diversen Monitorings und Angeboten zur Unterstützung der internen Evaluation unterschiedlichen Zwecken, z. B. der Bildungspolitik als Grundlage der Entscheidungsfindung und verschiedenen Akteuren an Einzelschulen zur Qualitätssicherung und -entwicklung in ihrem jeweiligen Handlungskontext. Ob und inwiefern die unterschiedlichen Adressaten von Ergebnissrückmeldungen tatsächlich Impulse für eine Veränderung ihres (ebenenspezifischen) Handelns erhalten, hängt von zahlreichen Faktoren ab – von Eigenschaften der Daten bzw. des Evaluationsdesigns, von

der Datennutzung, von organisationalen Merkmalen des Handlungsfeldes und von Eigenschaften der Datennutzer (vgl. Visscher & Coe, 2003).

Je nachdem, mit welchem primären Untersuchungsziel – neben Wissensgewinnung nennt Landwehr (2011) Rechenschaftslegung, Normendurchsetzung und Schulentwicklung – und für welche Grundgesamtheit Daten generiert werden, ob es sich um Daten zu Inputs, Prozessen, Outputs schulischer Lernprozesse und/oder um Daten zu Handlungskontexten verschiedener Akteure handelt, variiert der potenzielle Informationswert für unterschiedliche Adressaten deutlich. Über den potenziellen Informationswert eines Datums hinaus entscheiden über den Nutzen für die Schule auch Formate der kommunikativen Vermittlung von Ergebnissen sowie die Fähigkeiten und Bereitschaften der Adressaten, die evidenzgestützten Deutungsangebote der Datengeber in die je eigene Handlungslogik zu übersetzen und praktische Schlussfolgerungen daraus zu ziehen.

Um die Komplexität des Themenfeldes der „evidenzbasierten Steuerung“ für diesen Beitrag handhabbar zu machen, nehmen wir nur die einzelschulischen Akteure als Adressaten von Datenrückmeldungen in den Blick und zeigen anhand des vergleichsweise offenen Angebots der begleiteten internen Evaluation mit KESS beispielhaft, welchen idealtypischen Modi der datengestützten Schulentwicklung Schulen innerhalb ihrer Kooperation mit dem externen Evaluationsteam folgen. Nachdem im folgenden Abschnitt die beiden idealtypischen Modelle der Funktionsoptimierung und des Prozessmusterwechsels skizziert werden, stellen wir im vierten Abschnitt den Ansatz des Unterstützungsangebots der begleiteten Einzelschulevaluation mit KESS vor. Anhand von exemplarischen Fällen der schulseitigen Nutzung des Angebots wird anschließend kritisch diskutiert, inwiefern eine Öffnung von Lernstandserhebungen in Richtung einer designbasierten Evaluation von Schulen die Ausrichtung der datengestützten Schulentwicklung einer Schule verändern kann.

3. Idealtypische Modi der datengestützten Schulentwicklung

In der Literatur zur Schul- und Unterrichtsentwicklung wird zwischen dem Modus der Funktionsoptimierung und einem grundlegenden Prozessmusterwechsel unterschieden (Kruse, 2004; Schratz, 2009; Wiesner & Schreiner, 2019). Die Funktionsoptimierung zielt laut Bartz (2013) auf eine Verbesserung bestehender Handlungsroutrinen hinsichtlich ihrer Effektivität und Effizienz durch einen Abgleich von Ergebnissen und Zielen. Die Organisation oder in ihr handelnde Personen reagieren auf Datenrückmeldungen mit retrospektiven Analysen und dem Versuch, die eigene Performanz „im Rahmen bestehender Funktionalität zu verbessern“ (Kruse, 2004, S. 19). Dabei orientiert sich die Funktionsoptimierung an Ideen von *Best* bzw. *Good Practice*, d.h., die Akteure richten ihr Handeln an einem bestimmbar und kom-

munizierbaren Ziel aus und betreiben bei der Verbesserung bzw. der Optimierung eine Arbeit an Details im Sinne eines „Detailmanagements“ (Schley, 2017, S. 2). Obwohl kleinschrittige Optimierungen bestehender Praxen häufig zunächst erkennbare Verbesserungseffekte hervorrufen, stellen sich Wiesner und Schreiner (2019, S. 82) zufolge ab einem gewissen Punkt Deckeneffekte ein. Damit Organisationen und Personen sich dennoch weiterentwickeln, müsse die Entwicklungsarbeit nicht mehr nur „im System“, sondern „am System“ erfolgen (ebd.).

Derartige Prozesse der grundlegenden Prozessmusterwechsel bzw. des Veränderungslernens (Argyris & Schön, 1999) orientieren sich an einer *Next Practice* und brauchen daher auch eine Vision einer wünschenswerten und anzustrebenden Zukunft, starke Änderungsimpulse und eine erhöhte Unsicherheitstoleranz aller Beteiligten (Schratz, 2009). Das retrospektive Analysieren tritt in diesem Modus zurück hinter ein prospektives Entwerfen. Neben explizitem Wissen über die eigene Praxis und aus der eigenen Praxis bekommt die Arbeit mit implizitem Wissen (Haltungen, Werten, Überzeugungen) eine größere Bedeutung.

Wiesner und Schreiner (2019) weisen dem Veränderungslernen „am System“ mehr Lösungsmöglichkeiten und damit Chancen auf grundlegendere Innovationen in der Organisation und Durchführung des Unterrichts zu als der Funktionsoptimierung „im System“ (ebd., S. 84). Weick und Quinn (1999) heben in ihrem Review von Analysen organisationalen Wandels jedoch deutlich hervor, dass auch ein kontinuierlicher Wandel, der durch wiederholte Aktivitäten, Routinen und augenscheinliche Beharrung an Bestehendem gekennzeichnet ist, immer auch fortlaufende Adaptionen und Anpassungen hervorbringt. Prozessmusterwechsel bzw. ein „episodic change“ erscheinen vor dem Hintergrund ihrer Analysen nur dann als der bessere Weg, wenn innerhalb der Organisation die Tiefenstruktur nicht mehr zu den wahrgenommenen Anforderungen aus der Umwelt der Organisation passt und es im Prozess gelingt, Unsicherheiten und Sorgen der handelnden Akteure in Motivation für Veränderung zu wandeln (ebd.).

Entsprechend dieser psychologischen Randbedingungen erfordern grundlegende Veränderungen am System Rolff (2019) zufolge ein Zusammenhandeln der schulischen Akteure in Entwicklungs- oder Lernwerkstätten. Durch eine solche Arbeit in Werkstätten haben sich mit gewisser Autonomie ausgestattete Schulen schon häufig in der Lage gezeigt, die vielfältigen Ideen der in ihnen tätigen Personen selbstverantwortet in eine neu strukturierte Praxis zu überführen, die Veränderungen auf den Ebenen der Organisation, des Unterrichts und der Professionalisierung umschließt und aufeinander bezieht. Akteure einer Schule arbeiten in Werkstätten im Idealfall als Expert*innen in (multi-)professionellen Lerngemeinschaften zusammen und integrieren aus einer geteilten Vision heraus alle Dimensionen, Aspekte und Facetten der Unterrichtsentwicklung möglichst kohärent in ein neues „Design“ (Mintrop, 2019, S. 36).

Wie Ansätze aus der designbasierten Schulentwicklungsforschung zeigen, sind solche Entwicklungsprozesse nicht linear:

„Diese Art von Design besteht aus einer Abfolge von Lernschritten oder -möglichkeiten, die in einem iterativen Versuch-und-Irrtum-Verfahren entdeckt und kreiert werden. Versuch und Irrtum ergeben sich jedoch nicht willkürlich, sondern erfolgen aus einer Handlungstheorie (*theory of action*), die auf der Basis von Bedarfsanalysen und wissenschaftlichen und praktischen Erkenntnissen gebildet wird“ (ebd.).

Wie das Zitat verdeutlicht, besitzen handlungspraktische Modelle, praktische Erfahrung und Erkenntnisse aus Evaluationen in verschiedenen Phasen der Entwicklung eines neuen Designs unterschiedliches Gewicht. Erkenntnisse aus dem Prozess müssen immer wieder in eine das Design kennzeichnende praktische Handlungstheorie integriert werden, bis im Idealfall ein pädagogischer Prototyp im Sinne eines gebrauchsfertigen Verfahrens entstanden ist, das aus explizitem und implizitem Wissen der beteiligten Akteure besteht (Rolf, 2019, S. 54).

Die Gestaltung komplexer Transformationsprozesse im Feld der Praxis erfordert im Kontext einer datengestützten Schulentwicklung also auch eine Re-Kontextualisierung von Daten zu Leistungen und schulbezogenen Einstellungen von Schüler*innen in einem Handlungsfeld, das auch durch Vorgaben des Designs (und der übergeordneten Handlungsebenen der Bildungspolitik und Bildungsverwaltung) sowie das Erfahrungswissen und die Werte, Haltungen und Überzeugungen der Lehrkräfte mitbestimmt ist.

Daten müssen durch geeignete Formate der Rückmeldung und dialogischen Auswertung im Zuge dieser Re-Kontextualisierung „zum Sprechen gebracht werden“ – dies gilt bei der Innovation eines pädagogischen Prototyps ebenso wie bei dessen Transfer in eine andere Schule, weil in jedem neuen Kontext die Re-Kontextualisierung anders verlaufen muss. Nach Kussau (2007, S. 302) müssen Innovationen innerhalb von Transferprozessen im neuen Kontext „nacherfunden“ werden, weil ihre Implementation nicht mit einer vollständigen Zerschlagung alter Routinen einhergehen kann. Wie im Folgenden dargestellt wird, werden die Daten aus der internen Evaluation mit KESS in einem kommunikativen Rückmeldeverfahren mit den individuellen und kollektiven Perspektiven der Lehrkräfte aus ihrer schulischen Praxis verschränkt, um auch komplexere Transformationsprozesse der pädagogischen Praxis zu ermöglichen.

4. Die interne Evaluation mit dem KESS-Instrumentarium in Hamburg

Nachdem die *Behörde für Schule, Jugend und Berufsbildung* der Freien und Hansestadt Hamburg im Rahmen der längsschnittlich angelegten Schulleistungsstudie *Aspekte der Lernausgangslage und der Lernentwicklung* (LAU) bereits von 1996 an systematisch und kontinuierlich Leistungsdaten eines kompletten Schüler*innenjahrgangs generiert hat (vgl. Behörde für Schule und Berufsbildung, 2011), wurde zwischen 2003 und 2012 ein weiterer Jahrgang mit der KESS-Untersuchung vom Ende der vierten Klasse bis zum Abitur begleitet. Alle zwei Jahre wurden die Lernstände in den Kompetenzbereichen mathematische Grundbildung, naturwissenschaftliche Grundbildung, Deutsch-Leseverstehen, Deutsch-Rechtschreibung und Englisch (allgemeines Sprachverständnis), fachbezogene Einstellungen und Selbstkonzepte der Schüler*innen sowie umfangreiche Informationen zu schul- und unterrichtsrelevanten Hintergrundmerkmalen erfasst. Die KESS-Studie wurde mit dem Anspruch durchgeführt, Auskünfte über die Effektivität des Gesamtsystems und Steuerungswissen über alle Schulen und Schulformen hinweg zu generieren sowie innerhalb des Hamburger Schulsystems zur Standardsicherung im Bereich der Grundbildung beizutragen. Neben der Rechenschaftslegung wurde der Studie von den Auftraggebern deshalb zugleich die Funktion zugeschrieben, einen Qualitätsrahmen für datengestützte interne Schulevaluationen zu sichern.

Die im KESS-Jahrgang eingesetzten Kompetenztests, die an der Universität Hamburg, dem Institut für Schulentwicklungsforschung an der Technischen Universität Dortmund, der Humboldt-Universität zu Berlin und dem Max-Planck-Institut für Bildungsforschung in Berlin entwickelt und validiert wurden (Vieluf, Ivanov & Nikolova, 2011), orientieren sich an einer Idee von Grundbildung, die jene basalen Kompetenzen betont, die für das Weiterlernen in allen Fächern und allen Domänen elementar sind und für alle Schulabschlüsse vorausgesetzt werden. Die Tests lassen es aufgrund von Ankeritemdesigns zu, Testergebnisse aus allen Messzeitpunkten jeweils auf einer für alle Schüler*innen gemeinsamen domänenspezifischen Kompetenzskala zu verorten und somit individuelle Entwicklungsverläufe abzubilden. Die KESS-Studie liefert daher auch einen validierten und normierten Referenzrahmen mit Benchmarks zur Einordnung und Interpretation von Ergebnissen aus Folgeuntersuchungen an Einzelschulen (vgl. Benzing, Nikolova, Hunger, Hüskers & Wild, 2010).

Aufgrund des umfassenden Referenzrahmens der KESS-Studie lassen sich Lernstände anhand verschiedener Bezugsnormen einordnen – (1) kriterial durch Kompetenzstufenmodelle und abschlussbezogene Benchmarks, (2) sozial durch Vergleiche mit Schulen aus der KESS-Population mit gleicher Schulform sowie (3) ipsativ durch Vergleiche mit früheren Kohorten aus derselben Schule. Da die Kompetenztests auch längsschnittlich normiert sind, lässt sich schließlich auch (4) die individuelle

Bezugsnorm anlegen, sofern eine Schüler*innenschaft zu verschiedenen Zeitpunkten getestet wird.

Bereits ab dem Schuljahr 2004/05 wurde das KESS-Instrumentarium zu Zwecken der internen Evaluation eingesetzt. Im Rahmen des Hamburger Schulversuchs „d.18 – selbstverantwortete Schule“ ließen 18 Schulen auf Eigeninitiative über fünf Jahre hinweg die Auswirkungen und die Nachhaltigkeit ihres selbstverantworteten pädagogischen Handelns auf den Kompetenzerwerb der Schüler*innen empirisch prüfen (fünf Schulen aus diesem Projekt erhielten den Deutschen Schulpreis).

Als Reaktion auf die steigende Nachfrage der Schulen beauftragte die *Behörde für Schule und Berufsbildung* im Schuljahr 2013/14 eine Arbeitsgruppe am IfBQ, einzelne Schulen (die Oberstufen der Stadtteilschulen, vier sechsjährige Grundschulen, die katholischen Schulen und Schulen mit besonderem Profil) bei ihren Vorhaben der internen Evaluation mit dem KESS-Instrumentarium zu unterstützen.

Besondere Charakteristika der begleiteten internen Evaluation mit KESS werden im Folgenden herausgestellt.

Den Schulen gehören ihre Daten. Innerhalb des neuen Projektzusammenhangs sind die Schulen Auftraggeber der Evaluation und besitzen sämtliche Verfügungsrechte über die Daten. Dem Begriffsverständnis von Thiel und Thillmann (2012, S. 37) folgend ist das Angebot des IfBQ eines der internen Evaluation, da die Schulen ihre Fragestellungen selbst festlegen, ein dazu passendes Design in Auftrag geben und es dabei unerheblich ist, „ob die praktische Durchführung und Auswertung an Externe delegiert wird“. Da die Evaluierenden hier keine Mitglieder der Schule sind und somit unabhängig von kollegiumsinternen Befindlichkeiten objektivierte Wissen über die Schule bereitstellen, wird der Begriff der *begleiteten* internen Evaluation verwendet. Würdigend anzumerken ist an dieser Stelle, dass eine Behörde hier ein Angebot finanziert, auf dessen Ergebnisse sie letztlich nur mit freundlicher Genehmigung der Schulen zurückgreifen kann. Dies zeugt von großem Vertrauen in die Wirksamkeit einer datengestützten Schulentwicklung und das selbstverantwortliche Agieren der Schulen gleichermaßen.

Die Einzelschule ist die Grundgesamtheit. Da der Einsatz aller Kompetenztests Aussagen über eine Einzelschule und nicht über ein ganzes Schulsystem zulassen soll, wird auf Rotationen in der Administration von Aufgaben (Multi-Matrixdesigns) verzichtet, d.h., verwendet wird ein fester Pool von normierten Aufgaben mit individualdiagnostischer Validität.

Es gibt mehr Bildungsziele als Kompetenzen in den Kernfächern. Neben psychosozialen Skalen zu schul- und fachbezogenen Einstellungen und Selbstkonzepten wird den Schulen auch ein um Instrumente aus den Bereichen Politische Bildung,

Gesellschaftswissenschaft und Religion erweitertes Spektrum von Kompetenztests angeboten.

Der Fokus liegt auf den individuellen Kompetenz- und Einstellungsprofilen. Den beiden vorstehend genannten Eigenschaften der internen Evaluation mit KESS entsprechend, werden den Schulen klassenweise die Lernstände und fachbezogenen Selbstkonzepte aller Schüler*innen in übersichtlichen Datenblättern zur Verfügung gestellt, aus denen auch hervorgeht, inwiefern die einzelnen Lernenden in unterschiedlichen Domänen oder Fächern den Benchmarks entsprechen, auf welchem Anforderungsniveau eine weitergehende Förderung erfolgen und inwieweit diese an der Bearbeitung negativer Selbstkonzepte ansetzen sollte. In diese Form der Ergebnisrückmeldung, die den Schulen spätestens vier Wochen nach der Datenerhebung zur Verfügung steht, wird explizit die individual-pädagogische Perspektive von Lehrkräften (vgl. Maier, 2009) berücksichtigt. Mit diesem individualdiagnostischen Screening ist auch ein gerichteter Impuls vom Evaluationsteam in die Schulen verbunden: hin zu einer größeren Sensibilisierung für die individuell je besonderen Bedürfnisse nach Lernunterstützung und hin zu einer Differenzierung des Lernangebots seitens der Lehrkräfte.

Leistungs- und damit förderrelevante Hintergrundmerkmale werden konsequent berücksichtigt. Sowohl die Datenblätter als auch die kommunikativen Rückmeldungen übergreifender Befunde enthalten grundsätzlich Informationen dazu, inwieweit es an der Einzelschule Hinweise auf Lernbarrieren gibt, die systematisch in einem Zusammenhang mit Merkmalen wie Geschlecht, Familiensprache oder soziokulturellem Hintergrund stehen. Berücksichtigt wird auch die nicht sprachgebundene kognitive Fähigkeit der Schüler*innen, allerdings nicht im Sinne einer Intelligenzdiagnostik. Es geht vielmehr darum, prüfen zu können, ob unterdurchschnittliche Testergebnisse möglicherweise auf die Sprachgebundenheit der Tests zurückzuführen sind und ob es Schüler*innen mit überdurchschnittlichem kognitivem Potenzial gelingt, dieses Potenzial auch in schulische Leistung zu übersetzen.

Individualdaten und systemisches Wissen sind die Grundlage einer umfassenden Analyse. Auf Wunsch mancher Schulen, die z. B. jahrgangsübergreifenden Unterricht, fächerübergreifende Profile, Projektunterricht oder selbstreguliertes Lernen in Lernbüros implementiert haben, werden auch qualitative Daten erhoben, um das Design der Schule als Evaluationsgegenstand genauer zu beschreiben. In der Regel werden dann leitfadengestützt und fokussiert auf die je besonderen Merkmale der Schule Gruppendiskussionen mit Lehrenden, Gruppendiskussionen mit Lernenden und/oder Interviews mit Schulleitungen geführt und mit der Methode des thematischen Kodierens ausgewertet. Das so gewonnene Kontextwissen wird dazu genutzt, konkrete Deutungsangebote zum Zusammenhang von Lernergebnissen, intendiertem und implementiertem Curriculum zu formulieren. Solche Deutungsangebote werden innerhalb eines Rückmeldesystems unterbreitet, indem zunächst die Testergebnisse auf Individual-, Klassen- und Schulformebene vor dem Kollegium oder

der Steuergruppe präsentiert und kommunikativ mit den Eindrücken aus der Praxisperspektive der Lehrkräfte validiert werden. Erst auf der Grundlage der geteilten Überzeugung, dass in den zurückgemeldeten Ergebnissen Evidenzen im Sinne legitimer „Mittel der Bestätigung und Rechtfertigung einer Annahme“ oder einer nachvollziehbaren „Grundlage einer Meinung“ (Jornitz, 2008, S. 207) enthalten sind, beginnt ein gemeinsamer Interpretationsprozess, in dem prospektiv pädagogische Entwicklungsfelder markiert werden und retrospektiv das Design der Schule möglichst unter Berücksichtigung von Input-, Prozess- und Kontextmerkmalen kritisch in den Blick genommen wird.

5. Beispiele der schulseitigen Nutzung des Angebots zur begleiteten internen Evaluation mit KESS

5.1 KESS in Stadtteilschulen

Mit der Schulreform von 2010 sind in Hamburg alle Schulformen der Sekundarstufe I außer dem Gymnasium in eine „zweite Säule“ überführt worden – die Stadtteilschulen, an denen sämtliche Schulabschlüsse erworben werden können. Obwohl auch einige Gesamtschulen mit eigener gymnasialer Oberstufe zu Stadtteilschulen wurden, fehlte es anderen Schulen häufig an Erfahrung, eine eigene Oberstufe für ihre Klientel aufzubauen und den Übergang in die Sekundarstufe II angemessen zu gestalten. Das Angebot der begleiteten internen Evaluation wurde, nachdem es sich herumgesprochen hatte, bereits im zweiten Jahr von der überwiegenden Mehrheit der Schulen genutzt (die Teilnahmequote liegt seither zwischen 76 und 87 Prozent). Das primäre Erkenntnisinteresse aller Stadtteilschulen liegt insbesondere darin, die Lernstände und fachbezogenen Selbstkonzepte zu erheben und den (Fach-)Lehrkräften und Schüler*innen der Jahrgangsstufe 11 möglichst früh zurückzumelden. Die Funktion eines diagnostischen Screenings dominiert somit. Etwa die Hälfte der Stadtteilschulen nutzt die Evaluation überdies zur Bilanzierung der Arbeit in der Sekundarstufe I, weil sie in den Analysen explizit eine Differenzierung zwischen Schüler*innen aus der eigenen Mittelstufe und den Neuzugängen wünschen und um Rückmeldungen zu systematischen Zusammenhängen zwischen Lernständen und individuellen sozialen Hintergrundmerkmalen der Schüler*innenschaft bitten. Ebenfalls ca. die Hälfte der Stadtteilschulen, die in Jahrgangsstufe 11 an der internen Evaluation mit KESS teilnehmen, möchten wissen, wie sich ihre Schüler*innen im Laufe der Oberstufe entwickeln, und lassen diese vor dem Abitur erneut testen.

Obwohl die Daten überwiegend prospektiv zur Unterstützung der individuellen Lernprozesse an die Lehrkräfte in Jahrgangsstufe 11 zurückgemeldet und demgegenüber seltener zur retrospektiven Analyse der Lernprozesse genutzt werden, lassen sich keine gesicherten Aussagen über den Modus der Datennutzung formulieren, da hier

allein aufgrund der mit der Schule abgesprochenen Fragestellungen und der bestellten Ergebnismeldungen auf den Modus geschlossen wird und dabei nicht ausgeschlossen werden kann, dass die Schulen auch jenseits der begleiteten internen Evaluation mit KESS datengestützte Schulentwicklung betreiben und das „KESS-Team“ nicht über jede implementierte Neuerung informieren.

Die engen Vorgaben durch Bildungspläne und Zentralabitur einerseits und die Erkenntnisse aus einem ersten KESS-Netzwerktreffen im Februar 2019 mit beteiligten Stadtteilschulen andererseits legen nahe, dass die Schulen ganz überwiegend in der Logik der Funktionsoptimierung geblieben sind. Denn im Austausch der Vertreter*innen von 20 Stadtteilschulen auf dem Netzwerktreffen wurde sehr deutlich, dass die Ergebnismeldungen den Schulen zwar helfen, pädagogische Entwicklungsbereiche zu markieren, konzeptionell neuartige Maßnahmen aber nur selten daraus abgeleitet werden konnten. Vom Netzwerk versprechen sich die teilnehmenden Schulen explizit, über gebrauchsfertige Verfahren anderer Schulen in ähnlicher Lage informiert zu werden. Eine der dort vorgestellten Maßnahmen – die an das Screening mit dem KESS-Test für mathematische Grundbildung anschließende individuelle Förderung mit einem web-basierten Lernportal (Hahn, Ivanov, Nikolova & Ehlers, 2019) – wird laut Angaben des Portals bereits fünf Monate nach dem Netzwerktreffen von ca. 100 Lehrkräften in ihren Lerngruppen genutzt. Auch die Vorstellung eines Coachings für Schüler*innen mit negativen fachbezogenen Selbstkonzepten stieß auf großes Interesse. Inwieweit derartige Impulse als den Unterricht ergänzende Zusatzangebote aufgegriffen werden oder mittelbar sogar grundlegende Prozessmusterwechsel initiieren, bleibt abzuwarten und müsste im Rahmen einer systematischen Evaluation von Transferprozessen innerhalb des Netzwerks untersucht werden.

Auch unabhängig vom Austausch innovativer Ideen im Netzwerk gibt es einzelne Stadtteilschulen, die nach wiederholter Rückmeldung wenig zufriedenstellender Lernstände ihrer Schüler*innenschaft z.B. mit der Einführung von Zeiten und Konzepten zur Leseförderung, der Einrichtung von Lernbüros oder fächerübergreifender Profile in der Sekundarstufe I bewusst neue Prozessmuster implementiert haben.

5.2 KESS in sechsjährigen Grundschulen

Anders als in den Stadtteilschulen stellt sich die Situation in den sechsjährigen Grundschulen dar, die im Rahmen eines Schulversuchs die längere gemeinsame Lernzeit an der Grundschule dazu nutzen wollten, grundlegende Prozessmusterwechsel einzuleiten. Mit Blick auf die je spezifische Klientel der Schulen wurden besondere Inklusionskonzepte, systematische Formen der Mitbestimmung und demokratischen Teilhabe und/oder jahrgangsübergreifende Lerngruppen ein-

geführt. Diese Schulen haben erst nach ihrem Antrag auf Teilnahme am Schulversuch den Wunsch geäußert, sich wissenschaftlich durch das „KESS-Team“ begleiten zu lassen. Die Erhebungen von Lernständen, fachbezogenen Selbstkonzepten und Hintergrundmerkmalen wurden hier ergänzt durch Analysen frei zugänglicher Dokumente wie den Schulprogrammen, den Schulhomepages, der Zusammenfassung des letzten Berichts der Schulinspektion und dem Antrag auf Teilnahme am Schulversuch, durch Unterrichtshospitationen sowie durch Interviews mit Schulleitungen und Lehrendenteams, um das Design der Schulen im Sinne des geplanten und implementierten Curriculums besser erfassen und in den Rückmeldesituationen berücksichtigen zu können. Überdies wurden im Fragebogen für Schüler*innen zusätzliche Skalen aufgenommen, mit denen auch die Lernenden-sicht auf die mit dem Design angestrebten überfachlichen Kompetenzen und Eigenschaften des pädagogischen Arbeitsbündnisses eingeschätzt wird. Die gemeinsame Wahrnehmung der Leistungsdaten kann so in einem gemeinsamen Interpretationsprozess auf die wahrgenommene Qualität des Lernangebots bezogen werden.

Obleich auch in den sechsjährigen Grundschulen die Rückmeldungen aus der internen Evaluation mit KESS überwiegend zur Funktionsoptimierung genutzt werden, handelt es sich hier um den speziellen Fall, dass diese Schulen mit der Evaluation eine Funktionsoptimierung neuer Prozessmuster anstreben, da sie mit Beginn des Schulversuchs z.B. jahrgangsgemischte Lerngruppen implementiert haben, einen pädagogischen Prototypen also, der aufgrund des zwangsweise hohen Grades an Selbstregulation einerseits ein breites Angebot an Lernmaterialien für unterschiedliche Anforderungsniveaus und andererseits Instrumente zur Planung, Begleitung und Reflexion der individuellen Lernprozesse umfassen muss. Ein solch radikaler Bruch mit den Routinen aus dem Unterricht in Jahrgangsklassen erfordert im Kollegium Mut und Zuversicht, die auch durch Rückmeldungen von Lernständen und deren Orientierung an geeigneten Benchmarks und unterschiedlichen Bezugsnormen (s. o.) verstärkt werden können. Ergänzende Rückmeldungen zur Wahrnehmung ihrer überfachlichen Kompetenzen und der pädagogischen Prozessqualität seitens der Schüler*innen vermitteln den schulischen Akteuren potenziell Sicherheit im Umbruch.

Da die sechsjährigen Grundschulen die Mitglieder des „KESS-Team“ auch außerhalb der regulären Ergebnisrückmeldungen als Beratende in Schulentwicklungsfragen konsultieren, wird u.E. deutlich, dass beim Transfer bzw. beim Nacherfinden pädagogischer Prototypen einer begleiteten internen Evaluation insbesondere dann ein hoher Stellenwert seitens der Schulen beigemessen wird, wenn das Evaluationsteam auf Wunsch der Schule Deutungsangebote zu den Kompetenz- und Einstellungsdaten vornehmen kann, die sich auch auf die spezifischen pädagogischen Prozesse einer Schule beziehen. Dazu muss sich das Evaluationsteam auch Expertise über den Prototypen aneignen, d.h., hierzu die Theorie und möglichst viele Varianten aus der

Praxis erarbeiten, und das prototypische pädagogische Design auch angemessen in einem Evaluationsdesign operationalisieren können.

6. Fazit

Die regelmäßig erfolgende Nachfrage nach einer begleiteten internen Evaluation mit KESS zeigt zunächst, dass ein von Externen generiertes, objektiviertes Wissen über Kompetenzstände und Einstellungen von Schüler*innen den Schulen einen differenzierten Realitätsbezug und eine zusätzliche Wahrnehmungsperspektive auf die schulische Praxis eröffnet. Die kommunikative Auseinandersetzung mit den Ergebnissen regt die individuellen und kollektiven Reflexionsprozesse sowie eine Konkretisierung von Entwicklungsmaßnahmen innerhalb der Schule an. Die Praxen unterschiedlicher Schulen im Umgang mit Daten aus der begleiteten internen Evaluation mit KESS zeigen, dass die Bereitstellung evaluativer Daten grundsätzlich keine kausalen Rückschlüsse auf die Form der Nutzung durch die Schule zulässt. Viele Schulen nutzen die Daten im Sinne der Funktionsoptimierung. Es gibt aber durchaus auch Fälle, die ausgehend von den zurückgemeldeten KESS-Ergebnissen die Entscheidung für eine grundlegende Neuorganisation des Unterrichts (hin zu mehr Individualisierung) getroffen haben. Daneben findet sich eine ganze Reihe von Schulen, die den Transfer von Prototypen einer pädagogischen Praxis in den eigenen Kontext – das Nacherfinden (Kussau, 2007) oder die Re-Kontextualisierung (Fend, 2008) einer pädagogischen Innovation – durch die interne Evaluation mit KESS wissenschaftlich begleiten lassen.

Es zeigt sich insbesondere in der Arbeit mit diesen Fällen, dass die Unterstützung dieser Transferprozesse über das möglichst genaue Eingrenzen zu bearbeitender Entwicklungsfelder in der Schul- bzw. Unterrichtsentwicklung hinausgehen kann. Gemeinsam mit schulischen Akteuren werden mitunter auch konkrete Veränderungsmöglichkeiten für die Praxis entwickelt und evaluiert. Dazu müssen jedoch die Evidenzen aus der begleitenden Evaluation mit Erfahrungswissen aus dem Transferprozess der jeweiligen Schule und aus vergleichbaren Transferprozessen an anderen Schulen in Beziehung gesetzt werden. Dieses „stellvertretende Erfahrungswissen“ aus anderen Schulen kann erstens von inhaltlich spezialisierten wissenschaftlichen Evaluationsteams eingebracht und zweitens in Netzwerken von Schulen mit vergleichbaren Problem- und Fragestellungen bzw. mit identischer Orientierung an einer prototypischen pädagogischen Praxis bereitgestellt werden. Die Kombination aus beidem verknüpft eine höhere Akzeptanz der Evaluationsergebnisse (aufgrund der Unabhängigkeit des Evaluationsteams) und verspricht zudem einen Austausch über praktische Konzepte und Erfahrungen. Wichtig ist, dass derartige Formen der Zusammenarbeit in einem geschützten, d. h. nicht öffentlichen Kontext erfolgen können.

Sowohl die begleitete interne Evaluation als auch die Unterstützung von Netzwerken lassen sich kaum mit verantwortbarem Ressourceneinsatz im Rahmen einer flächen-deckenden externen Evaluation realisieren. Regelmäßige Lernstandserhebungen können aber durchaus ihren kybernetischen Impetus überwinden, wenn Landesinstitute bzw. die Qualitätseinrichtungen der Länder den Schulen ergänzend zu ihnen Unterstützungsangebote für eine begleitete designbasierte Schulentwicklung vorhalten. Ein solches Angebot könnte adressatengerecht für bestimmte Cluster von Schulen auf- und ausgebaut werden, die ihre Schulentwicklung an bestimmten pädagogischen Prototypen ausrichten (wollen).

Literatur und Internetquellen

- Altrichter, H., & Maag Merki, K. (Hrsg.). (2016). *Handbuch Neue Steuerung im Schulsystem* (2., überarb. und aktual. Aufl.). Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-531-18942-0>
- 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., überarb. und aktual. Aufl.) (S. 235–277). Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-531-18942-0>
- Argyris, C., & Schön, D. A. (1999). *Die lernende Organisation. Grundlagen, Methode, Praxis*. Stuttgart: Klett-Cotta.
- Bartz, A. (2013). Innovationen an Schulen gestalten. Zwischen Optimierung und Innovationen unterscheiden. *Schulverwaltung. Bayern*, 36 (7–8), 207–210.
- Behörde für Schule und Berufsbildung der Freien und Hansestadt Hamburg (Hrsg.). (2011). *LAU – Aspekte der Lernausgangslage und der Lernentwicklung. Klassenstufen 5, 7 und 9*. Münster: Waxmann.
- Benzing, M., Nikolova, R., Hunger, S., Hüskers, H., & Wild, H. (2010). Standardisierte Schulleistungstests. Individuelle Diagnosemöglichkeit und Unterstützung zur Unterrichtsentwicklung. *Hamburg macht Schule. Zeitschrift für Hamburger Lehrkräfte und Elternräte*, 22 (3), 22–23.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS.
- Hahn, S., Ivanov, S., Nikolova, R., & Ehlers, B. (2019). Von der internen Evaluation mit KESS in die individuelle Förderung. *Hamburg macht Schule. Zeitschrift für Hamburger Lehrkräfte und Elternräte*, 31 (1), 47–49.
- Jornitz, S. (2008). Was bedeutet „evidenzbasierte Bildungsforschung“? Über den Versuch, Wissenschaft für Praxis verfügbar zu machen. *DDS – Die Deutsche Schule*, 100 (2), 206–216.
- Kruse, P. (2004). *Next practice – erfolgreiches Management von Instabilität. Veränderung durch Vernetzung*. Offenbach: GABAL.
- Kussau, J. (2007). Schulische Veränderungen als Prozess des „Nacherfindens“. In J. Kussau & T. Brüsemeister (Hrsg.), *Governance, Schule und Politik* (S. 287–304). Wiesbaden: VS. https://doi.org/10.1007/978-3-531-90497-9_8
- Landwehr, N. (2011). Thesen zur Wirkung und Wirksamkeit der externen Schulevaluation. In C. Quesel (Hrsg.), *Wirkungen und Wirksamkeit der externen Schulevaluation* (S. 35–69). Bern: hep.

- Maier, U. (2009). Testen und dann? – Ergebnisse einer qualitativen Lehrerbefragung zur individualdiagnostischen Funktion von Vergleichsarbeiten. *Empirische Pädagogik*, 21, 191–207.
- Mintrop, R. (2019). Designbasierte Schulentwicklung – ein kurzer Abriss. In C. Schreiner, C. Wiesner, S. Breit, P. Dobbstein, M. Heinrich & U. Steffens (Hrsg.), *Praxistransfer Schul- und Unterrichtsentwicklung* (S. 35–48). Münster & New York: Waxmann.
- Rolff, H.-G. (2019). Transfer von Innovationen im Schulbereich. In C. Schreiner, C. Wiesner, S. Breit, P. Dobbstein, M. Heinrich & U. Steffens (Hrsg.), *Praxistransfer Schul- und Unterrichtsentwicklung* (S. 49–60). Münster & New York: Waxmann.
- Schley, J. (2017). Führung im Z: Wie meine Führungsqualitäten den Einsatz der Z-Strategie unterstützen können. *LEA News*, 13 (1), 2–4.
- Schratz, M. (2009). „Lernseits“ von Unterricht. Alte Muster, neue Lebenswelten – was für Schulen? *Lernende Schule*, 12 (46/47), 16–21.
- 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* (S. 35–55). Wiesbaden: VS. https://doi.org/10.1007/978-3-531-94183-7_2
- Vieluf, U., Ivanov, S., & Nikolova, R. (Hrsg.). (2011). *KESS 10/11. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der Sekundarstufe I und zu Beginn der gymnasialen Oberstufe*. Münster: Waxmann.
- Visscher, A. J., & Coe, R. (2003). School Performance Feedback Systems. Conceptualisation, Analysis, and Reflection. *School Effectiveness and School Improvement*, 14 (3), 321–349. <https://doi.org/10.1076/sesi.14.3.321.15842>
- Weick, K. E., & Quinn, R. E. (1999). Organizational Change and Development. *Annual Reviews of Psychology*, 50, 361–386. <https://doi.org/10.1146/annurev.psych.50.1.361>
- Wiesner, C., & Schreiner, C. (2019). Implementation, Transfer, Progression und Transformation: Vom Wandel von Routinen zur Entwicklung von Identität. Von Interventionen zu Innovationen, die bewegen. Bausteine für ein Modell zur Schulentwicklung durch Evidenz(en). In C. Schreiner, C. Wiesner, S. Breit, P. Dobbstein, M. Heinrich & U. Steffens (Hrsg.), *Praxistransfer Schul- und Unterrichtsentwicklung* (S. 79–140). Münster & New York: Waxmann.

Stefan Hahn, Dr., geb. 1974, Wissenschaftlicher Referent im Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ), Hamburg.

E-Mail: stefan.hahn@ifbq.hamburg.de

Stanislav Ivanov, geb. 1969, Wissenschaftlicher Referent im Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ), Hamburg.

E-Mail: stanislav.ivanov@ifbq.hamburg.de

Roumiana Nikolova, Dr., geb. 1971, Wissenschaftliche Referentin und Leiterin des Arbeitsbereichs datengestützte Schulentwicklung mit KESS im Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ), Hamburg.

E-Mail: roumiana.nikolova@ifbq.hamburg.de

Korrespondenzadresse: IfBQ, Beltgens Garten 25, 20537 Hamburg