

Der Beitrag greift Ergebnisse einer Forschungsarbeit auf, die unter dem Projekttitel „Zur Nutzung großflächiger Tests für die Schulentwicklung – Exemplarische Analyse der Erfahrungen in England, Frankreich und den Niederlanden“ im Auftrag des BMBF durchgeführt wurde. Aus dem Vergleich der bislang diskutierten Effekte großflächiger Testformen für die einzelne Schule werden Überlegungen zur Entwicklung bildungspolitischer Perspektiven sowie zum weiteren Forschungsbedarf abgeleitet.

Datengeleitete Schulentwicklung

Was fangen Schulen mit den
Ergebnissen aus überregionalen Tests an?

Von Isabell van Ackeren

Das Interesse an Leistungsvergleichen und ihrem Potenzial für Entwicklungsprozesse von Schulen ist aus deutscher Perspektive insbesondere durch die Vorlage der zwischen 1995 und 2001 veröffentlichten Ergebnisse internationaler Leistungsvergleichsstudien wie TIMSS, PISA und IGLU und die damit verknüpften empirischen Belege mangelnder Qualitätsstandards in Forschung und Öffentlichkeit deutlich in den Vordergrund getreten. Die Wahrnehmung der aufgezeigten Missstände lässt in der Folge eine ganze Nation anhaltend intensiv nach möglichen Gegenmaßnahmen fragen und durch den Blick auf erfolgreiche PISA-Nationen auf Rezepte hoffen. Entsprechende Handlungsoptionen werden vor allem in der so genannten Outputsteuerung von Schule gesehen, mit der schulische Arbeitsresultate als

Qualitäts- und Steuerungsdimension stark in den Blick geraten sind. Dementsprechend zeigt die Bundesrepublik nach Jahren der Abstinenz¹ wieder deutliche Präsenz bei internationalen Leistungsvergleichen und auch innerhalb der Bundesländer sind die Aktivitäten überregionaler Leistungsmessung zahlreich. Sie dienen einerseits der Bestandsaufnahme schulischer Arbeitsresultate und damit der nationalen wie internationalen Standortbestimmung, sollen andererseits aber auch – so eine neuere Hypothese – selbst Maßnahme der Qualitätsentwicklung sein und als solche bewusst eingesetzt werden: Auf der einen Seite werden Schulen klare Zielvorgaben gemacht², auf der anderen Seite wird die Erreichung dieser Ziele standardisiert erfasst und verglichen. Wie die einzelne Schule jedoch diese Zielvorgaben erreicht, bleibt ihr zuneh-

mend mehr selbst überlassen (ein zentrales Stichwort ist an dieser Stelle das der schulischen ‚Teilautonomie‘ beziehungsweise das der ‚Selbstständigen Schule‘)³.

Vielfältig: Typen und Varianten der Leistungsmessung

Dabei sind die Formen großflächiger Leistungsmessung nicht mehr nur stichprobenartig angelegt, um Entwicklungen auf der Ebene des Gesamtsystems Schule zu beobachten (so genanntes ‚Systemmonitoring‘) oder um eine wissenschaftsorientierte Grundlagenforschung zu betreiben, wie man die Ziele einer Studie wie PISA beschreiben könnte. Es setzen sich darüber hinaus mehr und mehr flächendeckende Formate durch, die für *alle* Schülerinnen und Schüler eines Bundeslan-



Isabell van Ackeren. Foto: Timo Bobert

Evaluationsform	Funktion			
	Systemmonitoring	Rechenschaftslegung	Zertifizierung / Selektion	Diagnose von Stärken und Schwächen / Entwicklung der Einzelschule
Schulleistungsstudien	●	●		
Parallelarbeiten		●	●	●
Vergleichsarbeiten	●	●	●	●
Zentrale Abschlussprüfungen		●	●	

(1) Übersicht über großflächige Lernstandserhebungen in Deutschland und die zugeschriebenen Funktionen.

Quelle: in Anlehnung an van Ackeren 2003

des verbindlichen und regelmäßigen Charakter haben, stärker an der Unterrichts-praxis anknüpfen und sie damit praxiswirksamer gestalten sollen. Folgende Maßnahmen werden dabei insbesondere diskutiert⁴:

– Bundesdeutsche Schulen werden derzeit mit – durch die Kultusministerkonferenz beschlossenen – so genannten *Vergleichs- und Orientierungsarbeiten* konfrontiert. Sie werden landesweit in Deutsch, Mathematik und überwiegend auch in der ersten Fremdsprache in bestimmten, von Land zu Land differierenden Jahrgangsstufen der Primar- und Sekundarstufe geschrieben. Sie sollen Standards in die Klassenzimmer transportieren und in umgekehrter Richtung Informationen über Defizitbereiche und Handlungsbedarfe liefern.

– Die Vergleichsarbeiten gehen über die bereits innerhalb von Schulen praktizierten Arbeiten zwischen parallelen Klassen (so genannte *Parallelarbeiten*) hinaus.

– *Zentrale Abschlussprüfungen* stellen schließlich einen weiteren wichtigen und traditionellen Evaluations-Strang dar, bei dem ebenfalls Externe von zentraler Stelle aus bereits seit Jahrzehnten in der Hälfte der Bundesländer – und es werden immer mehr, die entsprechende Maßnahmen beschließen – schulische Arbeitsergebnisse überregional beurteilen.

Exemplarisch sei auf die Situation in Nordrhein-Westfalen verwie-

sen: Derzeit werden in den Klassen 3, 7 und 9 Parallelarbeiten in parallelen Klassen einer Schule in Sprache beziehungsweise Deutsch, Mathematik und Englisch geschrieben. Im Schuljahr 2004/2005 kommen Vergleichsarbeiten in den Klassen 4 und 9 in den gleichen Fächern auf die Schulen zu; sie werden die Parallelarbeiten vermutlich ablösen. Und schließlich wird es in Zukunft zentrale Abschlussprüfungen nach der 10. Klasse geben. Tabelle 1 fasst die in Deutschland derzeit zentral diskutierten beziehungsweise in einer Vielzahl der Bundesländer eingesetzten überregionalen Tests und Prüfungen und die Erwartungen, die mit ihnen verknüpft sind, zusammen.

Zunehmend gefragt: Schulische Datenrückmeldung und -nutzung

Die Fülle der so gewonnenen Daten verändern Schule kaum – es sei denn, es würden Wege der Rückmeldung gefunden. Dabei erweist sich die deutsche Praxis des Daten-Feedbacks – dies fällt insbesondere im internationalen Vergleich ins Auge – bislang als eher unterentwickelt. Das Problem, wie Formen standardisierter Leistungsmessung über Momentaufnahmen schulischer Qualität hinaus mit Hilfe bestimmter Rückmeldeformate zum Ausgangspunkt von Schulentwicklungsprozessen im Sinne von Qualitätsverbesserung und Vergleichbarkeitssicherung gemacht werden können, wird in Deutschland in jüngster Zeit von

vielen Bildungsforschern aufgeworfen: Was bringt das Vermessen von Schulen eben diesen Schulen?

Das Bestreben, über Evaluationsformen wie Schulleistungsstudien und Vergleichsarbeiten schulische Routinen zu durchbrechen, ist in zwei grundsätzlichen Modellen in seiner Umsetzung denkbar:

1) Zum einen können Einzelschulen direkt und vertraulich über ihre Ergebnisse entsprechender überregionaler Tests, Studien etc. unterrichtet werden. Ein solches Feedback soll den Ausgangspunkt eines informationsbasierten Entscheidungshandelns vor Ort zugunsten der Qualitätsentwicklung darstellen. Dabei ist mehr als fraglich, ob Rückmeldungen linear in Entscheidungshandeln überführt werden beziehungsweise überführt werden können.

2) Zum anderen existiert – in der Diskussion und Praxis bislang weniger in Deutschland als in anderen europäischen Ländern – die Idee, über die Etablierung marktähnlicher Strukturen in Bildungssystemen in den schulischen Institutionen einen von außen gesetzten Anreiz zur Qualitätsentwicklung wirksam werden zu lassen. Dahinter steht die zumeist bildungspolitische Annahme, dass standardisierte Testsysteme, die aus Informations- und Anreizsystemen bestehen, geradezu notwendig sind, damit sich Bildungsakteure dem Ziel der Schulentwicklung verpflichtet fühlen. Dabei spielt die Veröffentli-

chung einzelschulischer Daten sowie die freie elterliche Schulwahlentscheidung – beispielsweise auf der Grundlage von Schulrankings in der Presse – eine wesentliche Rolle: Je besser die Arbeits- beziehungsweise speziell die Leistungsresultate einer Schule sind, desto eher kann die Anwahl dieser Schule durch Eltern angenommen werden. Je mehr Schüler eine Schule jedoch anzieht, desto mehr finanzielle Ressourcen erhält sie in Systemen mit einer Pro-Kopf-Finanzierung. In der Konsequenz müssen sich ‚schlechte‘ Schulen um Verbesserung bemühen, um in einer radikalen Wirkung finanziell nicht auszubluten und sogar geschlossen zu werden.

Die Wirksamkeit solcher Quasi-Märkte im Bildungsbereich ist umstritten; Einigkeit herrscht hingegen über die mangelnde Fairness eines solchen Systems, das die unterschiedliche Ausgangslage der Schülerklientel, mit der Lehrkräfte an unterschiedlichen Schulstandorten arbeiten, nicht berücksichtigt. Bislang zeichnen sich ähnliche Entwicklungen, wie sie etwa aus England bekannt sind, in Deutschland jedoch nicht ab, obgleich das Bundesland Sachsen mit der einzelschulischen Veröffentlichung der Ergebnisse zentraler Abschlussprüfungen an Mittelschulen und Gymnasien begonnen hat.

In Deutschland setzen die Bundesländer derzeit auf den zuerst genannten Weg. Es wird zunehmend nach Verknüpfungsmöglichkeiten zwischen Tests und Schulentwicklung gefragt: Wie lässt sich das Verhältnis von System- und Schulebene über Tests, Prüfungen und Studien definieren und möglicherweise steuern? Veranlassen großflächig eingeführte Reformelemente wie standardisierte Leistungstests Bildungsakteure tatsächlich zu einer nachhaltigen Änderung ihrer Praxis? Und: Welche Einflussfaktoren sind dabei zu berücksichtigen? Oder: Handelt es sich vielleicht gänzlich um eine Steuerungsillusion (vgl. Abb. 2)?

Erhellend: Der Blick zu den Nachbarn

Die vergleichenden Analysen der Erfahrungen anderer Länder zum Nutzungspotenzial überregionaler Lernstandserhebung stützen sich auf Literatur- und Materialrecherchen sowie auf Interviews mit Experten aus Wissenschaft und Administration in England, Frankreich und den Niederlanden. Diese Länder wurden ausgewählt, da sie in unterschiedlichen Verwaltungstraditionen über einen aus deutscher Perspektive vergleichsweise breiten Erfahrungshorizont und eine lange empirische Forschungstradition verfügen und

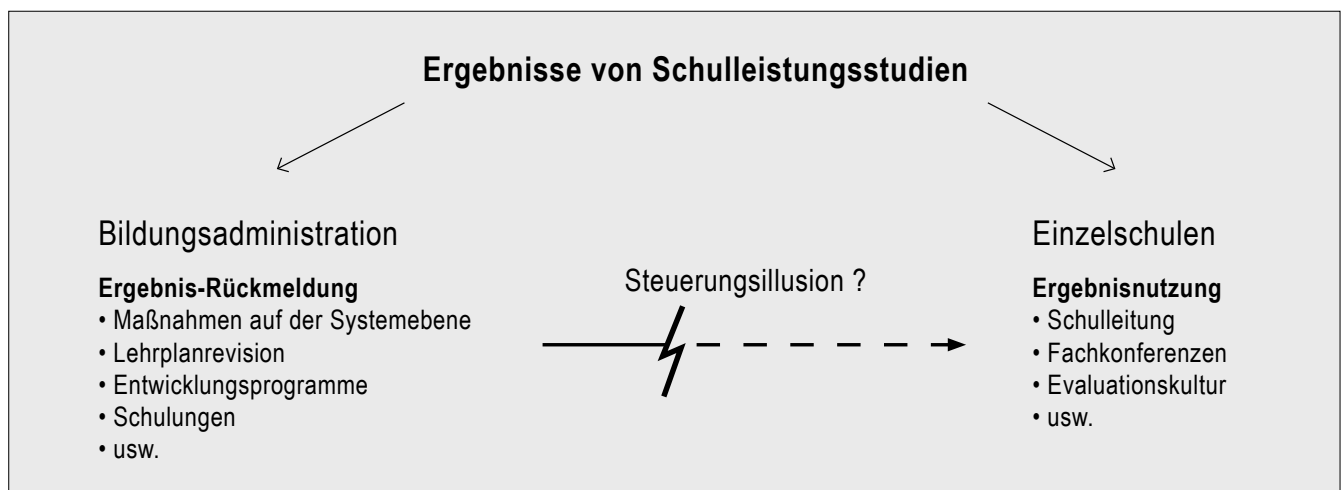
bei internationalen Vergleichsstudien eher gute bis sehr gute Leistungen erzielen. In Deutschland hingegen hat die empiriegestützte Forschung erst in den neunziger Jahren durch die aufgezeigten Defizite Auftrieb erhalten.

– Das Beispiel aus *England* verdeutlicht in diesem Kontext die Einbindung eines zentralisierten Prüfungswesens in eine marktorientierte Schulentwicklung.

– In *Frankreich* wird versucht, Gleichheit im zentral agierenden Staat in der Tradition der republikanischen Schulpolitik über zentrale Evaluationen zu stärken.

– Die *Niederlande* stehen schließlich für den Versuch, Qualität und Vergleichbarkeit in einem auf Schulautonomie setzenden Land zu sichern.

Großflächige Lernstandsmessungen sind in allen drei Untersuchungsländern somit eine omnipräsente Praxis, sozusagen ‚business as usual‘. Dabei setzen diese Länder vor allem auf diagnostische Tests während der Schullaufbahn als Ergänzung zu den Abschlussprüfungen, die den Leistungsstand nur noch abschließend feststellen aber entsprechende Rückmeldungen nicht mehr in das Unterrichtsgeschehen zur individuellen Förderung rückfließen lassen können. Darüber hinaus wächst in den gewählten Vergleichsländern mit der Ergebnismitteilung



(2) Datengeleitete Schulentwicklung – eine Steuerungsillusion?
 Quelle: Abbildung verändert nach van Ackeren / Rolff 2002

großflächiger Tests den bereits erwähnten Schulwahlentscheidungen seitens der Eltern eine wichtige Funktion zu. Angesichts der neuen Publizität von Evaluationsdaten gewinnt auch der Aspekt der fairen Darstellung von Leistungen unter Berücksichtigung der Lernausgangslage der Schüler/-innen zusehends an Bedeutung.

Die Frage der Test- und Prüfungspraxis dieser Länder soll an dieser Stelle nicht vertieft werden⁵, sondern es werden vielmehr die Wirkungen von Feedbackverfahren und einzelschulischen Nutzungsstrategien in den Blick genommen.

Theoretisch: Denkbare Rückmelde- und Nutzungsstrategien

Die Anatomie des Verwendungszusammenhangs von Test- und Prüfungsergebnissen in Schulen erweist sich in vielen Aspekten als unklar. Sicher scheint, dass Ergebnisse unterschiedlicher Evaluationsformen nicht automatisch in Entscheidungshandeln transportiert werden. Vor diesem Hintergrund soll ein Modell zur Wirkung entsprechender Feedback-Systeme präsentiert werden. In einem weiteren Schritt sollen diesen theoretischen Überlegungen Erkenntnisse aus den Untersuchungsländern gegenübergestellt werden. Abbildung 3 skizziert die grundlegenden anzunehmenden Einflussgrößen der Nutzung großflächiger Tests in und durch Schulen beziehungsweise durch die Mitarbeiter in Schulen. Dazu gehören

- die Anlage beziehungsweise Konzipierung überregionaler Testung mit ihren Zielen und Verfahren,
- die Rückmeldebedingungen einschließlich der Datenqualität und
- die Datenempfänger mit ihren personalen Bedingungen im Kontext der Schule als Institution und Organisation.

Im Zentrum der Übersicht stehen denkbare Nutzungsstrategien. Diese Einflussgrößen werden in ihrer Bedeutung nun weiter differenziert.

– Bei der *Anlage großflächiger Evaluationen* geht es um die Frage, welchen Einfluss die Ziele und Verfahren unterschiedlicher Testformate auf den Gebrauch von Daten haben können⁶. Zu den möglichen Erwartungen an Evaluationen über ein für einzelschulische Entwicklung weniger hilfreiches Systemmonitoring hinaus gehört zum einen die Rechenschaftslegung der Schulen gegenüber der Öffentlichkeit. Die zentrale Frage lautet, ob Schulen unter externem Veränderungsdruck eher dazu neigen, sich einerseits zu verbessern oder andererseits Defizite zu verstecken als in einem System mit weniger stark ausgeprägten externen Qualitätskontrollen. Ein weiteres Evaluationsziel kann im direkten Zusammenhang die Unterstützung elterlicher Schulwahlentscheidungen im Rahmen bildungsmarktähnlicher Strukturen sein, ebenso wie die direkte Initiierung von Schulentwicklungsprozessen. Hier finden sich Diskussionen über die Wirkung öffentlich und vertraulich rückgemeldeter Daten.

– Bei *Rückmeldeverfahren und Datenqualität* ist die Frage zu stellen, welche Daten auf welche Weise und durch wen rückgemeldet werden und damit zu welchen Effekten führen. Wesentliche Charakteristika rückgemeldeter Daten können dabei sein: Rohdaten und adjustierte Daten, absolute und vergleichende Daten, Standarddaten und individuell zugeschnittene Informationen, die Aktualität der Informationen und die Darstellung von Zeitverläufen sowie die Präsentation der Daten und schließlich Unterstützungsmaßnahmen bei ihrer Interpretation.

– Die Effekte der Feedbackverfahren hängen ganz wesentlich auch mit den *Datenempfängern und -nutzern* zusammen. Hierzu gehören die Lehrkräfte mit unterschiedlichen Positionen und Aufgaben im multidimensionalen Arbeitszusammenhang der Organisation Schule sowie die Schülerschaft, die Schuladministration und Bildungspolitik sowie die breitere Öffentlichkeit

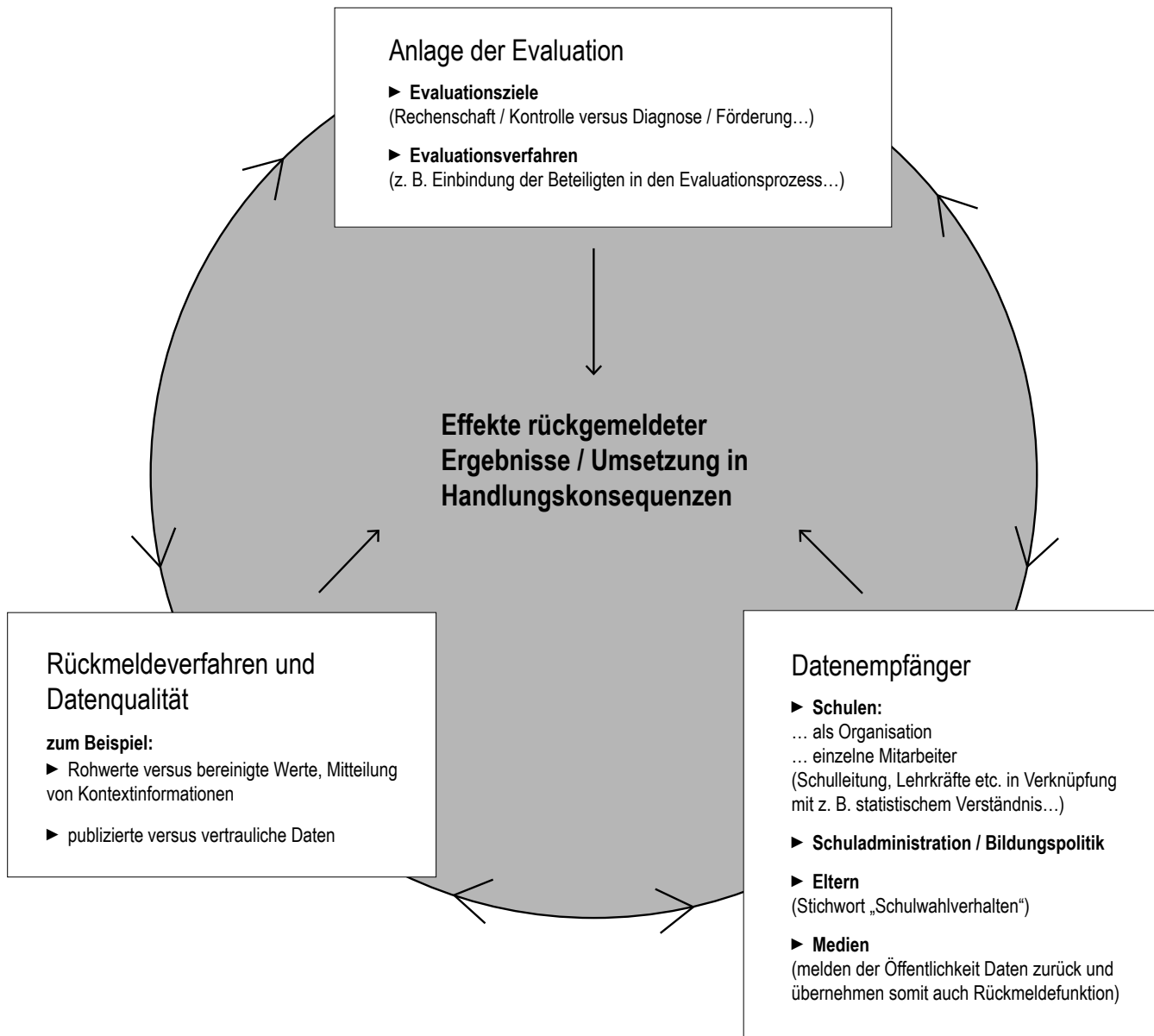
einschließlich der Medien. Innerhalb dieser Gruppen spielen wiederum individuelle und organisatorische Unterschiede der Einzelschule eine Rolle. Zentral erscheint vor allem der Einfluss der Sozialpsychologie von Lehrerkollegien: Lehrerarbeit stellt sich zumeist als Vereinzelung ohne effektive interne Kommunikationsstrukturen dar und ist nicht selten durch defensive Routinen (Abwehr, Verteidigung, Vertuschung) gekennzeichnet. Allerdings kann bezweifelt werden, ob derart strukturierte Lehrerkollegien geeignet sind, Probleme zu identifizieren und selbstständig zu lösen; Evaluation und Umgang mit Evaluationsdaten erfordert vielmehr Konfliktbereitschaft, Zeit und Arbeitsaufwand⁷.

Hinsichtlich möglicher Nutzungsstrategien sind vor allem folgende Formen denkbar: Bei der instrumentellen Nutzung werden Evaluationsergebnisse direkt als Ausgangspunkt von Handlungen angesehen. Bei der konzeptuellen Nutzungsform wird angenommen, dass Denkprozesse von Entscheidungsträgern längerfristig und akkumulativ geformt werden. Diese Form ist allerdings schwierig in eindeutigen Kausalketten nachzuweisen.

Festzuhalten bleibt, dass sich Schulentwicklung auf der Basis objektiver, rückgemeldeter Daten als ein zeitintensiver, komplexer Prozess interdependenter Komponenten erweist, der durch wechselnde menschliche Verhaltensweisen und soziale Strukturen beeinflusst wird. Auf der Grundlage der recherchierten Literatur wendet sich der nachfolgende Abschnitt exemplarisch den englischen, französischen und niederländischen Erfahrungen hinsichtlich dieser beschriebenen Einflussgrößen und Nutzungsstrategien zu.

Empirisch: Internationale Erfahrungen

Grundsätzlich ist zu berücksichtigen, dass auf der Basis der aktuellen, recherchierten Forschungslage keine Nettoeffekte großflächiger Test-



(3) Rahmenmodell möglicher Einflussgrößen auf die Wirkung bzw. gezielte Nutzbarmachung von Evaluationsergebnissen.

Quelle: van Ackeren 2003

formen beschrieben und bewertet werden können. Der Forschungsgegenstand ist international bislang nicht ausreichend repräsentativ erfasst und beschrieben und stellt sich als ‚work in progress‘ dar. Dennoch kann der Blick über die Grenzen einige wichtige Erkenntnisse befördern, die zum Aufbau handlungsrelevanten Wissens beitragen können und Forschungsbedarf identifizieren helfen. Nachfolgend werden die benannten Einflussgrößen – Anlage der Testform, – Wirkungen von Rückmeldeverfahren und Datenqualität sowie

– Wirkungen personengebundener Daten in ihrem Einfluss auf Nutzungsstrategien in den länderspezifischen Erfahrungen aufgegriffen.

Anlage der Testformen

In *England* hat man zum Ende der achtziger Jahre gezielt mit Hilfe laufbahnbegleitender Tests (Assessments) und entsprechender Ergebnisveröffentlichungen eine Wettbewerbssituation zwischen den Schulen initiiert, um schulische Leistungsfähigkeit zu entwickeln.

Die nachfolgend bei internationalen Leistungsvergleichen festzustellende deutliche Leistungssteigerung im Verlauf der neunziger Jahre wird vielfach als Beweis der Wirksamkeit dieser Steuerungsphilosophie herangezogen, ohne für einen solchen Kausalzusammenhang eindeutige Belege liefern zu können. In England geht man häufig von einer Schockwirkung in der Folge der radikalen Reformen aus, die aber aus Sicht der Kritiker nicht in langfristige Entwicklungen überführt werden können. Vor diesem Hintergrund

setzen die englische Forschung und die Schulentwicklungspolitik zunehmend auf unterstützende Maßnahmen bei festgestellten Defiziten.

In *Frankreich* hat man immer schon auf die Unterstützung von Schulen durch Datenrückmeldungen gesetzt. Die wesentlichen leistungsbezogenen, adjustierten Datensätze von Tests und Abschlussprüfungen werden zwar auch veröffentlicht, ohne dabei jedoch in Bildungsmarktstrukturen wie in England eingebunden zu sein; der Begriff der ‚Marktlogik‘ wird eher vermieden.

In den *Niederlanden* verfolgt man immer deutlicher einen ‚pressure and support‘-Ansatz, das heißt Schulen geraten durch die Verpflichtung zur öffentlichen Rechenschaftslegung unter einen gewissen Leistungsdruck, werden zugleich aber auch unterstützt, um Defizite zu beheben. Zwar war die Etablierung eines Bildungsmarktes nie eigentliches Ziel der Bildungspolitik, doch haben die Entwicklungen seit Ende der neunziger Jahre mit der vor Gericht durch die Medien erstrittenen Publikation schulischer Leistungsindikatoren im Kontext der bereits gegebenen Schulwahlfreiheit und der Pauschalfinanzierung zu einem verstärkten Wettbewerb der Schulen beigetragen. Auf der anderen Seite wird eine Vielzahl von optionalen Evaluationsinstrumenten entwickelt, die den Schulen als Diagnosemittel zur Verfügung stehen, um rechtzeitig individuelle Schwächen zu erkennen und angehen zu können.

Offensichtlich finden vor allem letztere Evaluationsformen Anwendung an niederländischen Schulen. Als zentrales Problem erweist sich die neue Verantwortungs- und Rechenschaftspflicht, welche die traditionell sehr autonomen Schulen in eine Spagatposition drängt, wollen sie an ihren Profilen und Visionen festhalten und zugleich den staatlichen Vorgaben und Anforderungen gerecht werden.

Wirkungen von Rückmeldeverfahren und Datenqualität

In *England* erstellen die lokalen Schulbehörden schulindividuelle Datenpakete, die von sehr unterschiedlicher Qualität sein können, da es keine standardisierten Verfahren des einzelschulischen Datenfeedbacks gibt. Die Unterstützung beim Umgang mit den vielfältigen Datensätzen erweist sich ebenfalls als noch sehr uneinheitlich, obgleich bereits spezielle Trainings angeboten werden oder Berater für die Dateninterpretation angefordert werden können. Ein Defizit wird mit Blick auf die Weiterbildung aller Lehrkräfte diagnostiziert. Meistens werden die Schulleitungen und Testkoordinatoren für entsprechende Maßnahmen berücksichtigt, die die Daten wiederum für ihre Kollegen aufbereiten, so dass es einen weiteren, schulinternen Rückmeldeprozess gibt. Es bleibt die Frage, ob auf diese Weise ein geteiltes Verständnis für externe Evaluationen und ihre Ergebnisse innerhalb der einzelnen Organisationen aufgebaut werden kann, für deren Nutzung sich alle Mitarbeiter verantwortlich fühlen.

Auch in *Frankreich* liegen offenbar sehr wenige Informationen zur Wirkung von Rückmeldeverfahren vor. Sehr deutlich wird lediglich ein starker Bedarf der Unterstützung im Umgang mit Daten formuliert. Hier richten sich die Überlegungen auf die Möglichkeiten der Lehreraus- und -fortbildung. Die Datenqualität stellt ebenfalls ein wichtiges Thema dar. Im Vordergrund steht die Vermittlung eines fairen Leistungsbildes, aber auch die externe Anregung des eigenen Innovationspotenzials mit Hilfe individuell zugeschnittener und verständlicher Informationen.

In den *Niederlanden* ist entsprechendes Forschungswissen ebenfalls sehr begrenzt. Bei den Rückmeldungen steht vor allem die Frage nach der öffentlichen versus vertraulichen Mitteilung von Indi-

katoren im Vordergrund. Hinsichtlich der Daten geht es vor allem um Datenqualität, um Fairness der Darstellungen und um ihre Verständlichkeit für unterschiedliche Zielgruppen. Ein wichtiger Diskussionspunkt ist zudem die Frage, wie sich schulische Qualität breiter als über kognitive Leistungsindikatoren abbilden lässt.

Wirkungen personengebundener Bedingungen

Die Organisation Schule und ihre Mitarbeiter stehen als Kernzelle von Innovation immer deutlicher im Vordergrund. Sie bewegen sich in Arbeitszusammenhängen, deren Komplexität eine Vielfalt möglicher positiver oder negativer Einflüsse auf die Wirkung von Lernstandsmessungen und rückgemeldeter Ergebnisse bedingt.

In *England* lassen erste Forschungsergebnisse, wie zu vermuten, deutlich werden, dass die Datenrückmeldung nicht automatisch ihren Gebrauch initiiert. Die Wirkung externer Evaluationsformen wird entscheidend von der bereits vorhandenen Leistungs- und Evaluationskultur einer Schule beeinflusst. Dazu gehören beispielsweise die Akzeptanz von Innovation und Wandel, effektive interne Kommunikationsstrukturen sowie die Verteilung der Verantwortung der Datenanalyse auf alle Mitarbeiter einschließlich des Einbezugs der Schüler. Vieles hängt auch von Interessen und Kompetenzen einzelner engagierter Kollegen ab. Insgesamt wird für die englische Situation deutlich, dass sich die anfänglich beschriebenen Probleme auf der mikropolitischen Ebene schulischer Organisationen, die sich vor allem als Misstrauen gegenüber externen, quantitativen Daten beschreiben lassen, im Laufe der letzten Jahre offensichtlich abgeschwächt haben. Solche Effekte werden überlagert von einem vertrauteren Umgang mit Daten, mehr Diskussionen über Fragen, die die Daten aufwerfen, dem stärkeren Gebrauch der Ergebnisse für die curriculare Planung auf schu-

lischer und auf Klassenebene sowie vom Nutzen für Schüler hinsichtlich der klassenübergreifenden Wahrnehmung und Kenntnis ihres Leistungsstandes.

In *Frankreich* erscheint die Informationslage zum Thema deutlich magerer, obgleich es erste Belege dafür gibt, dass die Daten vor allem solche Schulen erreichen, die bereits über eine Evaluationskultur verfügen. In der entsprechenden *niederländischen* Literatur wird lediglich die Skepsis von Schulen gegenüber der verstärkten externen Kontrolle beschrieben. Schulen und ihre Mitarbeiter empfinden demnach kein Gleichgewicht zwischen den ihnen traditionell zugestandenem Freiheiten und der neuen Rechenschaftspflicht, was sich offensichtlich als Bremse für die Datennutzung erweist.

Auch die Eltern schulpflichtiger Kinder stellen wichtige Informationsempfänger dar: In allen drei Ländern spielen elterliche Schulwahlentscheidungen eine mehr oder weniger intendierte Rolle mit Blick auf die Schaffung von Konkurrenzsituationen zwischen Einzelschulen, um darüber Maßnahmen der Qualitätsentwicklung wirksam werden zu lassen. Die Schaffung von Konkurrenz im Kontext von einzelschulischer Datenpublikation und in den Ländern unterschiedlich stark gegebener elterlicher Schulwahlfreiheit folgt verschiedenen Intentionen:

- In *England*, wo Eltern den Wunsch für eine bestimmte Schule äußern können, ist die Etablierung marktähnlicher Strukturen zentral gewesen.

- In *Frankreich* geht es bei traditionell festen Schulbezirken, obgleich es Experimente mit Schulwahlfreiheit gibt, nicht um die Schaffung von Märkten, sondern um die Sicherung gleicher Chancen.

- In den *Niederlanden* schließlich ergibt sich die Schulwahlfreiheit, die in Amsterdam mittlerweile eingeschränkt gilt, aus der Tradition schulischer Autonomie.

Der Schwerpunkt der Analysen liegt in allen drei Ländern auf dem

Zusammenhang zwischen sozialem Hintergrund und Schulwahl und einer damit verbundenen möglichen sozialen Segregation. Solche Effekte sind jedoch schwierig nachzuweisen. Eltern höherer sozialer Schichten wissen zwar offensichtlich in allen drei Ländern besser, die ihnen mitgeteilten Informationen zu verstehen, doch wird tendenziell vor allem in England und den Niederlanden eine Schule in der Nähe des Wohnortes beziehungsweise mit einem besonderen pädagogischen Profil gesucht, und zwar vorrangig gegenüber leistungsbezogenen Aspekten.

Beschriebene Nutzungsstrategien

In den für *England* beschriebenen kleineren, regional begrenzten Leistungsstudien werden Informationen im Sinne der instrumentellen Nutzung für die innere Differenzierung, für ein gezieltes Schülermonitoring und die Reflexion von Lehr-Lern-Prozessen genutzt. Mit Blick auf die schullaufbahnbegleitenden National Curriculum Tests und die zentralen Abschlussprüfungen sind einerseits kurzfristige Strategien beschrieben, mit denen Schulen versuchen, ihren Listenplatz in öffentlichen Rankings durch gezielte Maßnahmen zu verbessern (zum Beispiel mit Marketingstrategien oder beschränkten Zulassungskriterien zu den Examina für schwächere Schüler/-innen). Hinsichtlich längerfristiger Nutzungsstrategien wird in England eine Einbindung in schulische Zielvereinbarungen, Planungs- und Bewertungsprozesse der Zielerreichung beschrieben, die sich auf Schülergruppen, Klassen und die Schule insgesamt beziehen und pädagogische Themen, curriculare Planungen und das Schulmanagement umfassen.

In *Frankreich* sind Leistungsstudien wie die so genannten ‚panels d’élèves‘ weniger auf einzelschulische Rückmeldungen hin angelegt, als vielmehr auf Steuerungswissen auf der Systemebene. Den Abschlussprüfungen bleibt

in erster Linie eine zertifizierende Funktion vorbehalten, obgleich es Forderungen hinsichtlich einer stärkeren Nutzbarmachung der Ergebnisse gibt. Die schullaufbahnbegleitenden diagnostischen Tests, die ihrer Intention nach für die Schul- und Unterrichtsentwicklung genutzt werden sollen, werden von schulischer Seite überwiegend als Rechenschaftsinstrument gegenüber den Eltern sowie für die Bildung möglichst homogener Lerngruppen, hingegen kaum für die Schulprogrammentwicklung genutzt. Auch das Indikatorensystem IPES, das Schulen aus unterschiedlichen Bereichen schulindividuelle Daten bietet, hat sich als Maßnahme der Qualitätssicherung durch die kombinierte Mitteilung standardisierter und schulspezifisch zugeschnittener Indikatoren in der Praxis etwa auf Grund der sonstigen Arbeitsbelastung im schulischen Alltag als wenig erfolgreich erwiesen. Ein nicht unerheblicher Effekt bezieht sich jedoch auf die Testinhalte: So wird beschrieben, dass national festgestellte Defizite in einem bestimmten Bereich eines abgetesteten Fachs zu deutlich verbesserten Leistungsergebnissen im darauf folgenden Jahr führten.

In den *Niederlanden* gehen Resultate national angelegter Stichprobenuntersuchungen (etwa der Projekte PRIMA und PPO) vor allem an das Schulinspektorat, an Schulbuchverlage, Beratungsorgane und Lehrerbildungseinrichtungen und weniger an einzelne Schulen. Die zentralen Abschlussprüfungen in den Niederlanden dienen vor allem der Sicherung von Vergleichbarkeit und der Kontrolle der schulinternen Benotungen. Die Ergebnisse der Prüfungen, die auf so genannten ‚Qualitätskarten‘ öffentlich einzusehen sind, werden weniger für einzelschulische Prozesse genutzt, da die Daten zu global erscheinen. Dennoch habe die Einführung der Qualitätskarte im Zusammenspiel mit den neuen Aufgaben der Schulaufsicht und ihren systematischen Schulinspektionen für viele Schulen den Anstoß gegeben, sich grundsätzlich stärker um die erreichten Arbeitsergebnisse zu kümmern. Die niederländischen ‚leer-

lingvolgsystemen‘ (Schülerbegleitsysteme) hingegen sind als diagnostische Tests gezielt auf die Nutzung der Daten für die Schul- und Unterrichtsentwicklung hin angelegt. Ihr Gebrauch in den Einzelschulen zeigt sich in der Analyse von Testergebnissen auch im nationalen Vergleich, dem Weiterreichen von Ergebnissen beim Übergang in die nächste Klassenstufe und bei der Planung von Prozessen, die die Schule insgesamt betreffen.

Zusammenfassend: Schulen sind bislang wenig datengeleitet

Die skizzierten Erfahrungen zeigen, dass die Nutzung großflächiger Tests und ihrer Ergebnisse durch viele Faktoren beeinflusst wird. Zentrale systemische Strategien können zwar die Bedingungen für Lernergebnisse schaffen, diese aber nicht direkt beeinflussen. Es wird vielmehr in der Einzelschule entschieden, wie externes Intervenieren beantwortet wird. Immer wieder findet sich jedoch die Annahme, dass Initiativen bei ihrer Einführung zu einem gewissen Teil ‚top-down‘ implementiert werden müssen, gefolgt von einem anschließenden Investment in den Aufbau schuleigener Kapazitäten durch nachhaltige, unterstützende Programme. Dieses Modell sieht in der Kombination von ‚pressure‘ und ‚support‘ den effektivsten Motor von Schulentwicklung. Danach erhalten Schulen ohne Einsichtsmöglichkeiten der Öffentlichkeit schulbezogene Informationen zu Leistungen und ihren Bedingungen sowie Unterstützungsmöglichkeiten, beispielsweise durch externe Berater. Sie werden aber zugleich verpflichtet, über ihre Leistungsergebnisse, die durch Mehrwertberechnungen möglichst fair wiedergegeben werden sollen, nach außen Rechenschaft abzulegen. Die Veröffentlichung von Daten erzeugt einen Wettbewerbsdruck der Institutionen untereinander, der durch bestimmte Elemente im Bildungswesen, etwa eine freie Schulwahl

und die Pro-Kopf-Finanzierung von Schulen, potenziert werden kann. In einem ‚pressure and support‘-System werden die Schulen aber nicht sich selbst überlassen, sondern sie erhalten bei erkannten Defiziten besondere Unterstützung, etwa in Form finanzieller Mittel oder durch professionelle Beratung und Unterstützung.

Zumindest thesenhaft lässt sich formulieren, dass sich großflächige Lernstandsmessungen bilanzierender Art offensichtlich gegenüber schulkarrierebegleitenden diagnostischen Tests als unproduktiv für die Qualitätsentwicklung erweisen. Testformate mit dynamisch-entwickelndem Charakter liefern den Vorteil, dass sie Verlaufsevaluationen darstellen und ihre Ergebnisse somit korrigierend während – und nicht am Ende – der Unterrichtsprozesse nutzbar gemacht werden können. Das Verfahren der Testung zu Beginn eines Schuljahres in Frankreich erscheint aus der Außenperspektive besonders interessant, zumal damit der diagnostische Charakter herausgehoben wird. Der Bedarf der Unterstützung diagnostischer Kompetenzen wird in Deutschland spätestens seit PISA breit diskutiert: Für deutsche Lehrkräfte wurde im Rahmen einer kleineren explorativen Fallstudie diagnostiziert, dass sie die Kompetenzen ihrer Schüler gerade im unteren Leistungsbereich schlecht diagnostizieren, obgleich die diagnostische Kompetenz ein wesentlicher Kompetenzbereich professionellen Lehrerhandelns ist⁸.

In der Folge der festgestellten Defizite werden in der Bundesrepublik Maßnahmen zur Unterstützung der Einzelschule als wichtiger pädagogischer Handlungseinheit in Form der eingangs angesprochenen Vergleichs- und Orientierungsarbeiten innerhalb und zwischen den Bundesländern etabliert, um Lehrkräften eine größere Vergleichsgruppe und normorientierte Impulsgeber für Qualitätssicherung und -entwicklung zu bieten. Dass diese Maßnahmen in der jetzigen Ausgestaltung

zu den erhofften Wirkungen führen, ist zweifelhaft, wie eine zu Beginn des Jahres 2004 durchgeführte explorative Fragebogenerhebung unter Lehrkräften an Essener Schulen zeigte⁹: Für Lehrerinnen und Lehrer gehen mit dieser Form der Überprüfung insbesondere Ängste in Hinblick auf Kontrolle ihrer Arbeit einher. Zudem gibt es eine nicht zu unterschätzende Gruppe von Lehrkräften, die insbesondere der mit diesen Instrumenten einhergehenden Vergleichbarkeit über die Einzelschule hinaus ablehnend gegenüber steht. Eine Mehrheit der befragten Lehrkräfte empfindet die seit einigen Jahren in NRW durchgeführten Parallelarbeiten als zusätzliche Arbeitsbelastung; in der Wahrnehmung einer knappen Hälfte der befragten Lehrkräfte hat sich die Kommunikation im Kollegium nicht verbessert. Parallelarbeiten führen nach Einschätzung der Lehrkräfte auch nicht dazu, dass sich bisher weniger aktive Kolleginnen und Kollegen mehr als zuvor in die schulische Arbeit einbringen. Es steht zu vermuten, dass der Umfang der Überzeugungs- und Unterstützungsarbeit, die geleistet werden muss, schulförmerspezifisch, aber sicherlich auch einzelschulisch erheblich variiert.

Für die Gestaltung externer überregionaler und vor allem regelmäßiger Evaluationen erweisen sich unter anderem folgende Aspekte als zentral: – Die Funktion der Testung und Bewertung sollte deutlich formuliert werden (Systemmonitoring, integrative Sicht, Rechenschaftslegung, Schulentwicklung...), ohne dabei eine Evaluationsform für ganz unterschiedliche Zwecke einsetzen zu wollen.

– Die Entwicklung dauerhafter und regelmäßiger Evaluationen als Reformelement sollte Experten aus der Bildungspraxis einschließlich der Lehrkräfte einbeziehen, um den Bedürfnissen vor Ort zu entsprechen und zugleich die Akzeptanz für das System möglichst breit sicherzustellen.

– Die zu erreichenden Leistungsstandards und die Evaluationsrichtlinien

müssen in ihrem Zusammenspiel klar und deutlich artikuliert werden.

– Rigorose Verfahren der Qualitätssicherung sollten von vorneherein implementiert werden, mit denen Qualitätskriterien der Tests fortdauernd gesichert werden.

– Es sollte immer überlegt werden, ob Rückmeldungen Sinn machen, wie sie aussehen können und an wen sie sich richten.

– Die Entwicklung eines professionellen Trainingsprogramms unter besonderer Berücksichtigung von Test- und Prüfungsmethoden sowie Feedback- und Nutzungsmöglichkeiten kann sich als sehr sinnvoll erweisen.

Schließlich bedarf es grundlegender Forschungsarbeit zu den Wirkungsmöglichkeiten großflächiger Testformen als externe Auslöser für schulinterne Reflexion und nachfolgende Handlungsstrategien im Kontext von in Deutschland praktizierten Evaluationsformen. Hierbei kommt der Verknüpfung von Schuleffektivitätsforschung und Schulentwicklungsforschung in Verbindung mit der Vergleichenden Erziehungswissenschaft besondere Bedeutung zu.

Summary

This paper examines the ways in which the results from various types of large-scale assessments and central examinations, conducted either on a regular or a more sporadic basis, in three European countries, are fed back to the schools and other institutions within the educational systems. The article goes on to look at the current research activities in the field, where both public and private information is used to assist in the development of schools. It starts with a look at the present position in Germany, where research into this matter has only recently begun and where, because of this, there is comparatively little data available. England, France and the Netherlands seem to be particularly interesting

due to the integration of the assessment and examination processes into very different, historically developed types of administration. What are the aims within the national context of these countries and what do the results of research tell us about the realisation of these intentions in the individual schools? Finally, in Germany there is a clear necessity for a fundamental debate on the targets, methods and strategies of using examination results for the improvement of schools, in order to avoid reaching an impasse with extended test procedures.

Anmerkungen

- 1) vgl. zur Leistungsentwicklung Deutschlands in historischer Perspektive van Ackeren 2002
- 2) vgl. den Beitrag von Knoche/Sprütten zur aktuellen Entwicklung und Bedeutung von Bildungsstandards in diesem Heft
- 3) vgl. den Beitrag von Klemm/Meetz in diesem Heft
- 4) vgl. dazu van Ackeren/Bellenberg 2004
- 5) vgl. dazu van Ackeren 2003
- 6) vgl. zu den angestrebten Zielen auch Tabelle 1
- 7) Rolff 2002
- 8) Deutsches PISA-Konsortium 2001
- 9) vgl. van Ackeren/Bellenberg 2004

Literatur

- Ackeren, I. van: Von FIMS und FISS bis TIMSS und PISA, Schulleistungen in Deutschland im historischen und internationalen Vergleich, in: Die Deutsche Schule 2/2002, 157-175
- Ackeren, I. van: Evaluation, Rückmeldung und Schulentwicklung, Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und den Niederlanden, Münster, New York, München und Berlin 2003
- Ackeren, I. van, Bellenberg, G.: Parallelarbeiten, Vergleichsarbeiten und Zentrale Abschlussprüfungen, Bestandsaufnahme und Perspektiven, in Rolff, H.-G., Holtappels, H. G., Klemm, K., Pfeiffer, H., Schulz-Zander, R.: Jahrbuch der Schulentwicklung Band 13, Daten, Beispiele und Perspektiven, Weinheim und München 2004 [in Vorbereitung]
- Ackeren, I. van, Rolff, H.-G.: Die Bedeutung großflächiger Leistungsuntersuchungen für die Schulentwicklung, in Weegen, M., Böttcher, W., Bellenberg, G., Ackeren, I. van: Bildungsforschung und Politikberatung, Schule, Hochschule und Berufsbildung an der

- Schnittstelle von Erziehungswissenschaft und Politik, Weinheim und Basel 2002, 145-166
- Coe, R., Visscher, A.: School Improvement through Performance Feedback, Lisse 2002
- Deutsches PISA-Konsortium: PISA 2000, Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich, Opladen 2001
- Kohler, B., Schrader, F.-W.: Ergebnisrückmeldung und Rezeption, von der externen Evaluation zur Entwicklung von Schule und Unterricht, in: Empirische Pädagogik (Themenheft), 1/2004
- Rolff, H.-G.: Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit an Schulen?, in Weinert, F. E.: Leistungsmessungen in Schulen, Weinheim und Basel 2001, 337-365
- Rolff, H.-G.: Rückmeldung und Nutzung der Ergebnisse von großflächigen Leistungsuntersuchungen, Grenzen und Chancen, in Rolff, H.-G., Holtappels, H. G., Klemm, K., Pfeiffer, H., Schulz-Zander, R.: Jahrbuch der Schulentwicklung Band 12, Daten, Beispiele und Perspektiven, Weinheim und München, 75-98

Die Autorin

Isabell van Ackeren, Jahrgang 1974, studierte Biologie, Germanistik und Erziehungswissenschaft für die Sekundarstufen I und II in Essen mit Auslandsaufenthalten in England und Amsterdam. Zwischen 2000 und 2003 war sie wissenschaftliche Mitarbeiterin in der Arbeitsgruppe ‚Bildungsforschung/Bildungsplanung‘ am Lehrstuhl für Bildungsforschung an der Universität in Essen. Sie promovierte 2003 im Fach Erziehungswissenschaft und ist seitdem wissenschaftliche Assistentin in der Arbeitsgruppe. Ihre Arbeitsschwerpunkte liegen in der Schulwirksamkeits- und Schulentwicklungsforschung sowie in der Internationalen Vergleichenden Erziehungswissenschaft. Zu ihrer projektbezogenen Arbeit gehört u. a. die wissenschaftliche Begleitung des ‚Projektwettbewerb Schulen und Unternehmen‘, einer Initiative der Alfried Krupp von Bohlen und Halbach-Stiftung, das Teilprojekt ‚Schulsystem und Kultur der Einzelschule als Kontext des naturwissenschaftlichen Lernens‘ im Rahmen der DFG-Forschergruppe und des Graduiertenkollegs ‚Naturwissenschaftlicher Unterricht‘ am Campus Essen und das vom Bundesministerium für Bildung und Forschung geförderte Projekt ‚Comparison of the Educational Systems in Selected Countries: Understanding the Inter-National Variation of PISA-Results‘ sowie in Fortführung dieser Untersuchung das Projekt ‚Vertiefender Vergleich der Bildungssysteme Kanadas und Deutschlands‘ in Kooperation mit dem Deutschen Institut für Internationale Pädagogische Forschung in Frankfurt a. M. und Berlin.