



Wolfgang Böttcher
Jan Nikolas Dicke
Holger Ziegler
(Hrsg.)

Evidenzbasierte Bildung

Wirkungsevaluation in Bildungspolitik
und pädagogischer Praxis

WAXMANN

Evidenzbasierte Bildung

Wolfgang Böttcher, Jan Nikolas Dicke,
Holger Ziegler (Hrsg.)

Evidenzbasierte Bildung

Wirkungsevaluation in Bildungspolitik
und pädagogischer Praxis



Waxmann 2009
Münster / New York / München / Berlin

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-8309-2240-7

© Waxmann Verlag GmbH, Münster 2009

www.waxmann.com

info@waxmann.com

Umschlaggestaltung: Christian Averbeck, Münster

Umschlagfoto: © manu – Fotolia.com

Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier,
säurefrei gemäß ISO 9706

Alle Rechte vorbehalten

Printed in Germany

Inhalt

<i>Wolfgang Böttcher, Jan Nikolas Dicke & Holger Ziegler</i> Erziehungswissenschaft, Bildungspolitik und Bildungspraxis. Anmerkungen zu einem schwierigen Verhältnis	7
--	---

Formen, Methoden und Probleme der Wissensgenerierung

<i>David Gough</i> Qualitative, Quantitative and Mixed Methods Systematic Reviews to Support Professional Decision Making in Education	23
--	----

<i>Stefanie Albus, Heinz-Günter Micheel, Hans-Uwe Otto & Andreas Polutta</i> Wirkungsforschung in der Jugendhilfe. Erfordernisse für angemessene Indikatoren und methodisches Design in der wissenschaftlichen Wirkungsevaluation	35
---	----

<i>Inka Bormann & Thomas Krikser</i> Indikatoren als spezifische Wissensform und ihre handlungskordinierende Bedeutung.....	47
---	----

<i>Hans-Peter Blossfeld, Jörg Doll & Thorsten Schneider</i> Die Nationale Bildungspanelstudie (NEPS).....	59
--	----

Erziehungswissenschaft als Begleiter von Politik und Praxis

<i>Steve Fleischman</i> User-driven Research in Education: A Key Element Promoting Evidence-Based Education.....	69
--	----

<i>Nils Berkemeyer & Nils van Holt</i> Wissen über Vernetzung von Schulnetzwerken. Einige Überlegungen zur Aufbereitung wissenschaftlichen Wissens für die Schulpraxis.....	83
---	----

<i>Elke Grimminger</i> Interkulturelle Kompetenz von Lehrkräften als Teilaspekt professionellen Handelns. Entwicklung und Evaluation eines Fortbildungskonzepts für Sportlehrkräfte	95
--	----

<i>Ursula Fiechter, Barbara Meili, Larissa Trösch & Birgit Weinmann-Lutz</i> Was kann man sicher wissen? Epistemologische Überzeugungen von Lehramtstudierenden sowie Lehrpersonen nach mehrjähriger Berufspraxis	105
---	-----

<i>Britta Oerke & Katharina Maag Merki</i> Einfluss der Implementation zentraler Abiturprüfungen auf die leistungsbezogenen Attributionen von Schülerinnen und Schülern vor dem Abitur.....	117
---	-----

Eckhard Rohrmann

Lernbehinderte Kinder und Jugendliche oder lernbehindernde Schulen? Plädoyer für einen Perspektivenwechsel bei der Betrachtung und Behandlung von Lernschwierigkeiten und Verhaltensauffälligkeiten..... 127

Christoph Ratz & Eva Stumpf

Leistungsentwicklung und soziales Selbstkonzept von Kindern mit und ohne sonderpädagogischem Förderbedarf in bayerischen Kooperationsklassen. Eine Langzeitstudie 139

Sebastian Stark & Christine Wiezorek

Der Bildungsauftrag der Schule und fachbezogene Bildungsstandards..... 153

Franz Schott & Shahram Azizi Ghanbari

ComTrans. Ein theoretisch fundierter und praktisch erprobter Ansatz als Beitrag zur Wirkungsevaluation und evidenzbasierten Bildungspraxis 165

Josef Muth

Bildungsgutscheine in der Weiterbildungspolitik – Potenziale und Risiken eines nachfrageorientierten Steuerungsmodells. Eine empirische Untersuchung am Beispiel des Förderinstruments Bildungsscheck NRW 175

Das Verhältnis von Erziehungswissenschaft, Politik und Praxis

David Wastell & Sue White

Fetish and Folly in the IT-enabled reform of Children's Services: Technology as Magic 189

Kathrin Dederling

Schulministerien und ihr Umgang mit Leistungsvergleichsstudien..... 201

Herbert Altrichter

Datenfeedback und Unterrichtsentwicklung – Probleme eines Kernelements im „neuen Steuerungsmodell“ für das Schulwesen 211

Aline Kadach & Olga Zlatkin-Troitschanskaia

Wirksamkeit der internen Evaluation als einzelschulisches Steuerungsinstrument..... 227

Ramona Buske & Olga Zlatkin-Troitschanskaia

Kollektive Innovationsbereitschaft von Lehrkollegien als Moderator für die Wirksamkeit bildungspolitischer Steuerung..... 239

Autorinnen und Autoren..... 251

Wolfgang Böttcher, Jan Nikolas Dicke & Holger Ziegler

Erziehungswissenschaft, Bildungspolitik und Bildungspraxis

Anmerkungen zu einem schwierigen Verhältnis

Die Kritik an Bildung und Sozialer Arbeit hat eine lange Tradition. Man könnte ihre Entwicklungsgeschichte geradezu als „Kritik-Geschichte“ schreiben. Dabei käme allerdings eine kunterbunte Geschichte heraus: mal zu schnell, mal zu langsam, mal zu autoritär, mal zu lax, mal zu viel, mal zu wenig; im Endeffekt scheint es so zu sein, dass alle – und aus allen möglichen Gründen – unzufrieden sind. Keinem konnte man es recht machen. Umso verblüffender ist, dass bis in die jüngere Vergangenheit hinein die Frage nach den belegbaren Ergebnissen von Bildungs- und Sozialer Arbeit allenfalls sporadisch thematisiert wurde.

1. Die ökonomisierte Bildung

Seit gut zwei Jahrzehnten jedoch hat sich eine neue Kritik etabliert, die eng verkoppelt ist mit der Idee einer Neustrukturierung der Steuerung pädagogischer Arbeit. Wesentlich geht es um ein Konzept, das als „Dezentralisierung“ bezeichnet wird. Unmittelbar verknüpft hiermit ist die „Outputsteuerung“, die sich geradezu radikal von einer vorgeblich bislang betriebenen „Inputsteuerung“ abzusetzen sucht. Die Reformen dienen vor allem dem Zweck, Ergebnisse pädagogischer Arbeit zu verbessern, sie zielen also – zugespitzt gesagt – auf die „Produktionsfunktion“ der Systeme: eine grundlegend ökonomische Perspektive.

Knapp gesagt geht es beim Prinzip der Dezentralisierung darum, die Verantwortung für Ergebnisse pädagogischer Arbeit weitgehend auf die Ebene der jeweiligen Handlungseinheiten zu delegieren, in denen die pädagogische Interaktion vollzogen wird: den Schulen, den Kindertagesstätten, den Erziehungsberatungsstellen, den Kinderheimen etc. Diese Dezentralisierung wird durch Instrumente der Re-Zentralisierung austariert. Dies geschieht im Kern erstens mittels Aufgaben- bzw. Zieldefinitionen, die häufig Ergebnis eines Diskurses zwischen Akteuren verschiedener Anspruchs- und Interessensgruppen sind. Zweitens werden Ergebniskontrollen und Aufsichtsinstrumente verschiedener Art eingesetzt: Audits, Inspektionen, Tests, Evaluationen, Rechenschaftsberichte u. ä. Diese Dimension wird allgemein gesprochen als Outputorientierung oder gar Outputsteuerung bezeichnet. Die Implikationen dieser Neuorientierung sind gewaltig. Einige Beobachter sehen in dem zunehmend vernehmbaren „call for greater accountability and efficiency in service delivery“ gar den international „biggest trend affecting professional work in the past 30 years“ (Leicht, Walter, Sainsaulieu & Davies, 2009).

Die Ergebnisorientierung des Modells manifestiert sich in den vier Prinzipien Effektivität, Effizienz, Evidenz und Erfolgsorientierung: den „Vier E“ der Reform (vgl. Böttcher, 2002). Effektivität fragt, bis zu welchem Grad ein Handlungssystem Ziele erreicht hat. Effizienz vergleicht Handlungsergebnisse mit Kosten und Aufwand, also dem Ressourcenverbrauch. Sie fragt nach der „Zweckmäßigkeit“ von Aktivitäten, die der Zieler-

reichung dienen sollen. Effizienz setzt Inputs und Ergebnisse ins Verhältnis. Mit Evidenz ist der Sachverhalt angesprochen, dass Aktivitäten nachweisen müssen, ihren Zweck oder ihre Zwecke erreicht zu haben. Die bloße Behauptung von Erfolgen oder Erfolgsversprechen reiche nicht mehr aus. Fehlende empirische Erfolgskontrolle unterstützt die These von der (möglichen) Verschwendung oder wenigstens mangelnden Zieltreue von Ressourcen. Das vierte Prinzip findet sich in der Idee der „Erfolgsanreize“: Den relevanten Akteuren sollen „Incentives“ oder Belohnungen für erfolgreiches Handeln angeboten werden.

Vor diesem Hintergrund lässt sich die oben behauptete neue Kritik beschreiben: Es wird bezweifelt, dass die im Rahmen der pädagogischen Arbeit geleisteten Maßnahmen ihre Ziele effektiv und effizient erreichen. Will die Pädagogik diesen Zweifel widerlegen, muss sie Evidenzen ihrer Wirksamkeit beibringen.

2. Die neue Rolle der Erziehungswissenschaft im Paradigma von Evidenz und Rechenschaftspflicht

Diese Beweisspflicht wird insbesondere an die Adresse der Praktiker gerichtet. Von ihnen wird erwartet, dass sie Leistungen dokumentieren oder evaluieren, und sie werden kontrolliert, wenn z.B. Inspektoren die Schulen besuchen, wenn zentrale Abschlussprüfungen die Kompetenzen der Schüler testen, Studiengänge akkreditiert werden. Auch in der Sozialen Arbeit findet die Gewährung von Leistungen zunehmend über Prozesse des Benchmarking und der Kostenrechnungen statt. Dabei sind die Träger von Angeboten aufgefordert, ihre Ziele von Beginn an s.m.a.r.t. (spezifisch – messbar – anspruchsvoll – realistisch – terminiert) zu formulieren und somit ihre Maßnahmen Outcomebeurteilungsverfahren und schließlich leistungsabhängigen, wirkungsorientierten Finanzierungsmodellen zugänglich zu machen (vgl. Plaßmeyer & Kohlmeyer, 2009). Die Bildungsarbeiter verspüren einen neuen und bisher unbekanntem Druck, der durch die dauerhafte Kontrolle ihrer Arbeit ausgelöst wird.

2.1 Verantwortung der Wissenschaft für den Praktiker

Es ist aber unseres Erachtens vernünftig, die Beweislast für Wirkungen gar nicht von den Pädagogen zu verlangen – zumindest solange nicht, wie Ziele und Wege der Arbeit nicht hinreichend beschrieben sind. Zwar dürfte z.B. ein Arzt dafür verantwortlich gemacht werden, wenn er eine Standardprozedur verfehlt. Aber um den Fehler zu begehen, muss dieser Standard als abgesichertes Verfahren existieren. Auch ein Anwalt dürfte zur Verantwortung gezogen werden können, wenn er einschlägige Paragraphen nicht beachtet hat. Aber weder hat der Arzt die Therapie entwickelt, noch ist der Anwalt Mitverfasser der Gesetze. Der erste folgt in der Regel gut abgesicherten Ergebnissen medizinischer Forschung, der andere kann sich entweder auf die Klarheit der Gesetze oder wenigstens auf den weitgehenden Konsens ihrer juristischen Auslegung verlassen. Aber was kann die Erziehungswissenschaft dem Lehrer, dem Sozialpädagogen, der Schulleiterin oder der Erzieherin an abgesichertem Handlungswissen anbieten?

Die im Paradigma der ergebnisorientierten Dezentralisierung geforderte Rechenschaft primär von der „pädagogischen Basis“ zu verlangen, führt in die Irre. Eine solche Perspektive unterschlägt die Rolle der Forschung. Und die kann man kritisch sehen.

Verschiedentlich wird, vor allem außerhalb der Disziplin, Kritik an der Unfähigkeit der Erziehungswissenschaft geäußert, der Praxis eine Orientierung zu geben. Die Bildungsaktivisten und Journalisten Kahl und Spiewak behaupteten in „Die Zeit“: „Dem Fach ging die pädagogische Bodenhaftung verloren“ (11.03.2005). Überhaupt sei der empirische Gehalt erziehungswissenschaftlicher Forschung mager. Eine Erziehungswissenschaft, die den üblichen Wissenschaftsstandards genüge, fände nur sehr punktuell statt, eine Informations- und Orientierungshilfe, die die Öffentlichkeit von der Erziehungswissenschaft insbesondere vor dem Hintergrund der sich entwickelnden Wissensgesellschaft erwarte, bliebe weitgehend aus. Zwei Kernbereiche des Bildungswesens, den Kindergarten und die Hochschule, habe die deutsche Erziehungswissenschaft „fast überhaupt nicht im Blick“ (Kahl & Spiewak, 2005, S. 2). Kahl und Spiewak kritisieren auch die Lehrerbildung, welche die zukünftigen Lehrer nicht auf das berufliche Handeln vorbereite.

Auch wenn sich die Erziehungswissenschaft sicherlich nicht auf Fragen der empirischen Erziehungswissenschaft reduzieren lässt (vgl. Merckens, 2006), entbehrt die Behauptung, eine an den Nutzern orientierte Forschung werde vernachlässigt, sicher nicht einer gewissen Grundlage. Während die Wissenschaftler sich häufig darüber mokieren, dass Pädagogen wissenschaftliche Publikationen nicht lesen, versäumen sie zu fragen, welche Informationen Nutzer benötigen. Wer hoffte, die eingeleitete neuere Wende zur „empirischen Bildungsforschung“ würde eine Perspektive für die praktische Pädagogik einleiten, dürfte sich ebenfalls enttäuscht sehen. Hier werden offenbar akademische Fragen bearbeitet, die dem Praktiker fremd sind. Dieses Argument ist noch ungesichert, aber es dürfte zumindest eine gewisse Plausibilität beanspruchen: Die Erziehungswissenschaft wird zunehmend dominiert durch die Pädagogische Psychologie, die ihre Kernaufgabe darin zu sehen scheint, Kompetenzen von Individuen zu vermessen. Ein Schwerpunkt ist die Entwicklung von Leistungstests. Arrondiert wird dieses Unterfangen mit dem Aufbau filigraner Kompetenzmodelle. Interessanterweise also führten die bereits mit relativ groben Instrumenten gemessenen dramatischen Defizite des Bildungssystems zu einer Verfeinerung der Messung, statt zur Entwicklung von Maßnahmen zur Reduktion der Gruppe der Risikoschüler. Dabei wäre gerade eine solche Forschungsstrategie für Lehrer und für die mit dieser Gruppe arbeitenden Sozialpädagogen tatsächlich nützlich.

Man kann empfehlen, auf einschlägigen Konferenzen der empirischen Erziehungswissenschaft die Frage nach der Nützlichkeit zu stellen. Von den meisten Referenten wird man allzu häufig nicht nur keine Antwort bekommen, vielfach löst allein die Frage Befremden aus. Es geht doch nicht um Nutzen!

Ein Symptom dürfte auch die zunehmende Dominanz quantitativer Verfahren sein, in der – häufig ohne theoretische und methodologische Reflexion – komplexe Rechenverfahren auf der Grundlage von Meinungsdaten aus vier- oder fünfstufigen Likert-Skalen exerziert werden. Auch die – den Naturwissenschaften und der Psychologie blind entnommene – These, erziehungswissenschaftliche Qualität erweise sich in internationalen Publikationen, dürfte eine Strategie unterstützen, die Probleme der hiesigen pädagogischen Praxis eher zu vernachlässigen.

Immer weniger wird, fährt man diesen Kurs weiter, die Erziehungswissenschaft einer Lehrerin sagen können, was diese tun muss, um den Risikoschüler zu einem mittleren Abschluss zu führen, oder dem Sozialpädagogen, was er tun muss, um sicher zu sein, dass es in seiner Problemfamilie zu keiner Kindeswohlgefährdung kommt. Erst von praktisch arbeitenden Pädagogen, denen die Wissenschaft bewährte Handlungen erläutern kann, lässt sich Rechenschaft verlangen. Eine nutzerorientierte Forschung kann die Beweislast für „gute“ pädagogische Arbeit vom Rücken der praktischen Pädagogen nehmen.

Um einen weiteren Zeugen für obige These zu benennen, möchten wir auf Hans N. Weiler verweisen. Er ging ziemlich unfreundlich mit der Disziplin um – und das auch noch „in der Höhle des Löwen“, nämlich im Rahmen des 18. Kongresses der Deutschen Gesellschaft für Erziehungswissenschaft, der im März 2002 in München stattfand. Auch er kritisierte die Praxisferne der Erziehungswissenschaft, die sich eher mit den „Randgebieten des Weltgeschehens“ befasse als mit dem, „was die Menschen in den Schulen und in Klassenzimmern eigentlich bewegt“ (Weiler, 2003, S. 182). Auch in der Sozialpädagogik ist die Kritik teilweise vernichtend. „[A]mong German professors of social work“, so befindet etwa Heinz Kindler (2008, S. 322), „there are not many who have ever done and published a high-quality empirical study [...]. [E]mpirical research as it is published at the moment in Germany simply does not examine effects of social work, and [...] most of the people doing social work research may not be prepared to understand and use more advanced empirical methods“.

2.2 Die Verantwortung der Erziehungswissenschaft für die Politik

Hans Weiler bezog seine Kritik aber nicht nur auf die Art der Forschung in der Erziehungswissenschaft. Darüber hinaus monierte er, dass die deutsche Erziehungswissenschaft sich zwar „eine gewisse behäbige, hier und da an Plüsch erinnernde Ehrwürdigkeit erworben [habe], aber von wirklichem Einfluss kann eigentlich keine Rede sein“ (Weiler, 2003, S. 182). Seine Kritik fußt auf der Grundannahme, dass die Erziehungswissenschaft der Bildungspolitik zu helfen habe, und dies insbesondere in Umbruchsituationen, wie sie aktuell z. B. mit den (anstehenden) Reformen der Hochschulen und der Lehrerbildung vorlägen. Die Erziehungswissenschaft lasse die Politik „sträflich im Stich“, sie „kommt ihrer Pflicht, bildungspolitische Entscheidungen empirisch zu erhellen und kritisch zu begleiten, in einem kaum angemessenen Umfang nach“ (ebd.). Nun wisse er ganz gut, dass die Bildungspolitik durchaus dazu neige, sich eher an „Gesichtspunkten politischer Opportunität oder ideologischer Korrektheit zu orientieren“ (ebd., S. 186), aber eine Wissenschaft, die große Schwierigkeiten damit habe, „sich mit einem empirischen, sozial- und verhaltenswissenschaftlich orientierten Selbstverständnis anzufreunden“ (ebd., S. 194), mache es „der Politik all zu leicht [...], sich auf das Fehlen einschlägiger wissenschaftlicher Belege herausreden zu können“ (ebd., S. 199).

Weiler bringt mit seiner Kritik die These zum Ausdruck, dass nur eine Erziehungswissenschaft politisches Gewicht erlangen könne, die sich für die einschlägige Politik als nützlich erweise. Das kann gelingen, wenn sie Wissen erzeugt, das der Bildungs- und Sozialpolitik in wenigstens zwei Zusammenhängen hilft. Sie kann erstens Informationen liefern, die bildungs- und sozialpolitische Entscheidungen rechtfertigen. Sie kann zwei-

tens diese Politikbereiche mit wissenschaftlichen Argumenten versorgen, die sie im Verteilungskampf mit anderen Ressorts stärken.

Andererseits benötigt eine solche Ausrichtung auch die Nachfrager aus der Politik. Eine Politik, die die Erziehungswissenschaft danach fragt, was sie weiß – und was sie weiß, um bestimmte Probleme zu lösen –, eine Politik, die eine solche Forschung fördert und hilft, das dann erzeugte Wissen auch bekannt zu machen und zu implementieren, ist, freundlich gesagt, kaum erkennbar. Das gilt auch vor dem Hintergrund, dass die Politik, immer wieder modische Trends aufnehmend, aktuell die Leitbegriffe „Outputsteuerung“ und „Evidenz“ vor sich herträgt. Am Beispiel der Schulpolitik und Hochschulpolitik sieht man allerdings, dass sie hierunter offenbar eher ein multiples Rechenschaftssystem versteht und nicht den Ansatz verfolgt, mittels systematischer Prozess- und Ergebnisanalyse pädagogischer oder bildungspolitischer Maßnahmen eine an Zielen orientierte Qualitätsentwicklung voranzutreiben. Ein solches Konzept lässt sich in einer „qualitätsentwickelnden Evaluation“ finden, die andere Akzente setzt als die kontrollorientierten Verfahren der Pädagogischen Psychologie oder die grundlagentheoretischen Modellierungen von Kompetenzen (vgl. Böttcher, 2009).

3. Programm-Evaluierung und Evidenz

Wir versuchen im Folgenden ein kurzes Plädoyer für eine Evaluationsforschung, die sich das Ziel setzt, pädagogische Programme oder bildungspolitische Maßnahmen nutzerorientiert und evidenzbasiert zu verbessern (vgl. Böttcher, 2008).

Unter Evaluation versteht man – grob gesprochen – die systematische Bewertung eines Produktes oder einer (Dienst-)Leistung. Diese Minimaldefinition benennt den „harten Kern“ des Verfahrens: die methodische Sorgfalt (Systematik) und das Normative (Bewertung), das aus der Tatsache resultiert, dass Bewertungs- bzw. Qualitätsmaßstäbe (mehr oder weniger) subjektiv sind, oder anders gesagt: dass sie sich an Zielen – mithin normativen Vorgaben – orientieren. Evaluationen beanspruchen in aller Regel, Wissen zu liefern, das zur Qualitätsentwicklung eines Produktes oder einer Dienstleistung beitragen kann.

Evaluation ist angewandte Wissenschaft, denn sie bedient sich der Methoden empirischer Sozialforschung. Evaluatoren verpflichten sich darauf, die Gütekriterien der empirischen Sozialforschung, wie etwa Objektivität, Transparenz der Dokumentation, Gültigkeit (Validität) und Zuverlässigkeit der Befunde (Reliabilität), zu gewährleisten. Aber Evaluation muss nicht unbedingt einem – verkürzt gesagt – positivistischen Wissenschaftsverständnis folgen, das seine Ideale aus den Naturwissenschaften bezieht. Die Geschichte der Evaluation zeigt, dass Evaluation mehr bedeutet als „nur“ zu messen, zu beschreiben oder zu bewerten. Die sogenannte „vierte Generation“ der Evaluation lässt sich insbesondere dadurch kennzeichnen, dass sie einen eindeutigen Anwendungsbezug aufweist (vgl. Mertens, 2005). Sie ergreift Partei für Nutzer von Pädagogik und die Pädagogen, sie intendiert Qualitätsentwicklung. Evaluation meint in diesem Sinne ein Verfahren systematischer Datensammlung und -interpretation mit dem Ziel, Impulse für die Verbesserung von Maßnahmen oder Systemen zu liefern. Hiermit wird das qualitätsentwickelnde Potenzial der Evaluation funktional in den Mittelpunkt gestellt.

Evaluation ist allerdings funktional vielfältig. Einschlägige Differenzierungen schreiben ihr durchaus auch eine Kontrollfunktion zu, aber die Entwicklungsfunktion steht in aller Regel im Mittelpunkt von Definitionen. Sie will sowohl das Lernen beteiligter (vor allem auch das der evaluierten) Akteure als auch das Lernen ganzer Organisationen bewirken. Dazu gehören u.a. das Aufdecken und Beseitigen vorhandener Schwachstellen in der Umsetzung pädagogischer Maßnahmen, Steigerung der Arbeitsmotivation von Mitarbeitern oder Förderung der Kommunikation in Organisationen durch Beteiligung möglichst vieler Akteure an der Planung, Durchführung und Auswertung der Evaluation (vgl. Geißler, Laske & Orthey, 2002, S. 161).

Evaluationen haben auch nicht Personenbeurteilungen zum Ziel. Dies ist die Perspektive der Psychologie oder der pädagogischen Psychologie. Im Fokus einer Erziehungswissenschaft, die sich für pädagogische Organisationen (also zum Beispiel Schulen) oder pädagogische Programme (also zum Beispiel der Bewertung eines e-learning-Konzeptes für Jugendliche ohne Schulabschluss) interessiert, stehen System- oder Programmevaluationen.

Am Beispiel der Programmevaluation kann Komplexität und Multiperspektivität verdeutlicht werden. Pädagogische Programme sind Interventionen, also organisierte, geplante, in der Regel auf gewisse Dauer angelegte Aktivitäten, die das Ziel haben, pädagogische oder soziale Kontexte zu verändern, und die Menschen oder Gruppen helfen wollen, ihre soziale Lage zu verbessern oder Lerngewinne zu erzielen. Rossi, Lipsey & Freeman (2004) haben fünf stimmige Fragenbereiche identifiziert, die das Vorgehen, die Ergebnisse und die Nutzung einer Evaluation determinieren. Hierzu zählen die Bewertung (1) des Bedarfs (*needs assessment*), (2) der Programmtheorie (*assessment of program theory*) sowie (3) des Programmprozesses (*assessment of program process*), (4) die Wirkungsmessung (*impact assessment*) oder Ergebnisevaluation sowie (5) – als wohl anspruchsvollste Evaluation – die Bewertung der Programmeffizienz (*efficiency assessment*).

Die Evaluationsstandards, die auf breitem Konsens fußen (vgl. Joint Committee, 2006, S. 51ff.), betonen die Bedeutung der Nützlichkeit (*Utility*) von Evaluationen: Die Ergebnisse von Evaluationen sollen den Adressaten Informationen liefern, die diese zur Bewertung und gegebenenfalls Verbesserung des in Frage kommenden Gegenstandes auch nutzen können. Sie sind also an den „Evaluationsbedürfnissen der vorgesehenen Evaluationsnutzer“ ausgerichtet (ebd., S. 51). Dazu gehört es u. a., die an der Evaluation beteiligten oder von ihr betroffenen Personen (*Stakeholder*) zu identifizieren, damit sie eingebunden werden können: Die Standards umfassen auch die Idee, dass Evaluationsberichte die wesentlichen Informationen in leicht verständlicher Weise zur Verfügung stellen.

4. Probleme von Evidenzen erzeugenden erziehungswissenschaftlichen Evaluationen

Die erziehungswissenschaftliche Forschung leidet ganz offensichtlich daran, dass sie nur wenige eindeutige Ergebnisse, sondern, vergleicht man verschiedene Studien zu gleichen Themen, eher „Mixed Results“ produziert. Womöglich resultiert eine gewisse Enttäuschung daraus, dass die Handlungs- und Sozialwissenschaften an Vorstellungen von der

Potenz naturwissenschaftlicher Forschungen gemessen werden. Dieser Maßstab ist deshalb falsch, weil auch die Naturwissenschaften sich nicht durchgängig eindeutiger Ergebnisse rühmen können. Aber das soll nicht Thema sein. Wir wollen einen schnellen Blick auf die „Natur“ der sozialwissenschaftlichen, insbesondere der erziehungswissenschaftlichen Forschung werfen. Hiermit soll die Erziehungswissenschaft nicht gegen Kritik immunisiert werden, aber doch in ein anderes Licht gerückt werden.

4.1 Das Problem von Erziehung als sozialem Prozess

Einige Kritiker monieren, dass eine evidenzbasierte Neuformierung der Pädagogik, die ihre Wissensbasis in randomisierten, kontrollierten Experimentalstudien findet, ein unrealistisch technologisches Modell professioneller Erziehungs- und Bildungspraxis nahe lege (vgl. Biesta, 2007; Otto, Polutta & Ziegler, 2009). Das Problem bestehe darin, dass Maßnahmen, die auf der Basis experimenteller Designs untersucht werden, klar identifizierbar und relativ standardisiert in verschiedenen Anwendungen eingesetzt werden müssen. Diese Prämisse der Programmintegrität bereitet pädagogischen Situationen typischerweise Schwierigkeiten. Sie sei, so argumentiert etwa Martyn Hammersley,

„hard, perhaps impossible, to achieve where the treatment is primarily a social process operating in many different contexts, and where the people administering the treatment are not in total control of the situation. In education and social work, for example, any application of a particular ‘treatment’ is likely to take a variety of forms across different classrooms or cases“ (2009, S. 142).

Aus einer professionstheoretischen Perspektive hat Donald Schoen (1983, S. 42–43) in seinem Klassiker über den „Reflective Practitioner“ schon vor 25 Jahren ganz ähnlich argumentiert. Schoen stellte der Vorstellung eines „high, hard ground“, auf dem Professionelle unmittelbar und effektiv auf die Erkenntnisse forschungsbasierter Techniken zurückgreifen können, die Realität eines „swampy lowland“ gegenüber, in der technologische, an der Rigorosität naturwissenschaftlicher Forschung bemessene Maßnahmen nicht greifen, da sich Situationen hier typischerweise als „confusing ‘messes’ incapable of technical solution“ darstellen würden. Die Schwierigkeit besteht also darin, valide Wirkungsaussagen für Maßnahmen im „swampy lowland“ zu entwickeln.

4.2 Die schwierigste Wissenschaft

Vor diesem Hintergrund scheint die Perspektive David C. Berliners gewinnbringend. Berliner nimmt eine interessante Position gegenüber der üblichen Unterscheidung zwischen naturwissenschaft- und sozialwissenschaftlicher Forschung ein. Zwar sei die Unterscheidung zwischen sog. harter und weicher Wissenschaft ein Teil der Wissenschaftskultur. Die ersteren lieferten harte Fakten, die präzise Vorhersagen ermöglichten und so genau seien, dass wir Brücken und Schaltkreise bauen können oder Raketen zum Mond schicken. Die andere sei weich, unbestimmt, unverlässlich, unpräzise, auf sie könne man keine praktischen Entscheidungen aufbauen (Berliner, 2002, S. 18). Richtig hingegen müsse man unterscheiden zwischen den schwierigen (hard-to-do) sowie den einfachen

(easy-to-do science) Wissenschaften. Die einfache Wissenschaft fänden wir in der Physik, der Chemie, der Geologie und ähnlichen Feldern vor, die schwierig durchzuführenden Wissenschaften seien die, die sich innerhalb des Spektrums der Sozialwissenschaften verorten (vgl. ebd.). Und ganz besonders schwierig sei es für die Erziehungswissenschaftler: „In my estimation, we have the hardest-to-do science of them all!“ (ebd.). Der hohe Schwierigkeitsgrad resultiere daraus, dass wir es in der Erziehungswissenschaft mit sehr partikularen Problemen zu tun haben, die sich immer nur innerhalb von lokaler Bedingung ereignen und damit Generalisierung und Theoriebildung limitieren. Zwei der Berliner zufolge wesentlichen Merkmale erziehungswissenschaftlicher Forschung, die für die Schwierigkeit der Forschung verantwortlich sind, wollen wir andeuten: Die Macht des Kontextes und die Allgegenwart der Interaktion.

Kontexte: Umfassende theoretische Gebäude und Generalisierung sind deshalb in der Erziehungswissenschaft so gut wie unmöglich, weil eine enorme Anzahl von Kontextbedingungen die pädagogischen Aktivitäten mitbestimmen. Dieser Tatbestand würde z. B. erklären, dass die vielen Reformprogramme, die in den USA systematischen Evaluationen unterzogen worden sind, mal positive, mal negative, mal indifferente Ergebnisse erzeugen. Eigentlich, so Berliners Fazit, gäbe es keine pädagogischen oder sozialen Programme, die ihre Wirkungen beweisen können. Dies sei aber nicht der Schwäche der Disziplin, sondern der Eigenart ihres Gegenstandes geschuldet. Es gibt offenbar kein Programm, das unabhängig vom jeweiligen pädagogischen Einsatzort konsistente Ergebnisse geliefert hätte. „These huge context effects cause scientists great trouble in trying to understand school life“ (ebd., S. 19). Und hieraus folge, so Berliner, auch die besondere Bedeutung der qualitativen Forschung innerhalb der Erziehungswissenschaft.

Die Allgegenwart der Interaktion: Um zum Beispiel zu verstehen, was in einem Klassenzimmer geschehe, müsste in „hohen Potenzen“ gedacht werden: Es interagieren nicht nur unterschiedliche Charakteristika von Schülern (Herkunft, Interesse, Können etc.), es interagieren auch Eigenschaften der Lehrer (Ausbildung, berufliche Motivation, Ich-Stärke etc.) und Merkmale der Umwelt (Kultur des Einzugsgebietes einer Schule, Management des Schulleiters, Lehrpläne etc.) miteinander. Und mehr noch: Keiner weiß, in welche Richtung sich die Wirkungen ausbreiten: „Because of the myriads of interactions, doing educational science seems very difficult, while science in other fields seems easier“ (ebd.). Berliner veranschaulicht die Besonderheit mit einem interessanten Beispiel: Die Naturwissenschaften können relativ häufig verallgemeinbare Aussagen treffen. Aber man stelle sich vor, dass Newtons drittes Gesetz sowohl auf der südlichen wie der nördlichen Hemisphäre funktioniere – allerdings nicht in Italien oder Neuseeland (ebd.).

Berliner kritisiert die aktuelle Forschungspolitik der USA, die auf quantitative und den Naturwissenschaften angelehnte experimentelle Verfahren setze. Sie ignoriere die Besonderheiten der Erziehungswissenschaft und sei dem Gegenstand unangemessen (ebd., S. 20). Eine Vielfalt von ethnografischen Studien, Fallstudien, Surveys, Langzeitstudien, experimentellen Ansätzen und Handlungsforschung sei hingegen nötig: „A single method is not what the government should be promoting for educational researchers. It would do better by promoting argument, discourse, and discussion“ (ebd.). Für die Evaluationsforschung würde eine solche Perspektive nahe legen, einen allzu engen Fokus darauf aufzugeben, ob sich ein spezifisches Programm (certis paribus) mit Blick auf bestimmte Wirkungsparameter in experimentellen Situationen als wirksam erweist. Vielmehr beste-

he eine wesentliche Aufgabe darin, ein generatives Verständnis von Wirkung zu entwickeln und in Evaluationsstudien anzuwenden. Solchen Studien ginge es dann vor allem darum, über Verursachungsprozesse und -mechanismen aufzuklären. Weniger die Formulierung mechanischer Handlungspräskriptionen, sondern vielmehr erklärende Deutungsangebote über Wirkungszusammenhänge würden im Zentrum solcher Analysen stehen. Ein empirisch fundiertes, nichtsdestoweniger jedoch theoretisches Wissen über Prozess-Mechanismus-Wirkungs-Konfigurationen (vgl. Pawson & Tilley, 2004) kann einen wesentlichen Hintergrund für professionelle (Ermessens-)Entscheidungen darstellen, wenn es darum geht, zu begründen, welches Vorgehen mit Blick auf die je individuellen, fallspezifischen Konstellationen und soziale Einbettungen der Adressatinnen und Adressaten angemessen erscheint.

4.3 Die Distanz der Praxis und der Politik

Die Hoffnung auf eine akzeptierte, evidenzbasierte und dabei wirkungsmächtige Forschung ist möglicherweise aber auch naiv. Die Praktiker und die Politiker müssten „mitspielen“. Hier ist aber Skepsis angebracht. Selbst wenn die Forschung die Wirkung eines pädagogischen Programms ermittelt hat und erklären kann, ist noch völlig offen, wie Praktiker darauf reagieren. Ihre hohe organisatorische und professionelle – und über Generationen ausgeübte – Autonomie erlaubt ihnen, sich von noch so gut begründeten Vorgaben zu emanzipieren. Standardisierungen des Handelns dürften als unprofessionell gelten und ihre Umsetzung auf Widerstand treffen.

Auch die Politik dürfte sich von noch so überzeugenden Ergebnissen nichts vorschreiben lassen. Eher ist zu vermuten, dass angesichts der vorwiegend legitimatorischen Funktion politischer Entscheidungen das Wissen um Wirkungen von politisch initiierten Maßnahmen nur dann akzeptiert werden dürfte, wenn sie sich mit den Intentionen decken. Des Weiteren liegt es nahe zu vermuten, dass es ohnehin bei einer Anzahl von Maßnahmen gar nicht um direkte (pädagogische) Wirkungen geht, sondern um die politischen Wirkungen. Um ein Beispiel zu nennen: Bei der Einführung von Kopfnoten für Schüler dürfte es kaum darum gegangen sein, ihnen soziale Kompetenzen zu vermitteln. Allerdings kann eine solche Maßnahme einer bestimmten Wählerklientel vermitteln, dass eine Regierung „eingreift“ und eine Maßnahme einführt, die man „von früher“ kennt.

Auch kann Wissen deshalb problematisch sein, weil es Handlungen als zwingend erscheinen lassen würde, die politisch nicht durchsetzbar oder nicht bezahlbar sind. Auch hier ein Beispiel: Bei der wissenschaftlichen Analyse outputorientierter Steuerung könnte sich erweisen, dass mangelhafte oder fehlende Ressourcen der Grund für unterschrittene Ergebniserwartungen waren.

Schließlich – aber keinesfalls erschöpfend – ist noch darauf hinzuweisen, dass mit der Anzahl von Informationen Entscheidungen nicht unbedingt einfacher werden. Das Gegenteil ist häufig der Fall. Komplexität erhöht den Reflexionszwang, Ideologie erleichtert Entscheidungen (vgl. Brunsson, 1982).

Politik und Wissenschaft stehen nicht notwendigerweise in einem durch gemeinsame Rationalität geprägten Verhältnis (vgl. auch Weegen, Böttcher, Bellenberg & van Ackeren, 2002). Die aktuelle Situation ist womöglich typisch auch für ihre Zukunft: Es fällt auf, dass Politik und Verwaltung kaum Anstrengungen unternehmen, die eigenen Maß-

nahmen und Politikprogramme zu evaluieren. Für die Politik ist das womöglich eine gute Entscheidung. Der praktischen Pädagogik dürfte das nicht helfen, auch der erziehungswissenschaftlichen Evaluation nicht.

5. Zu dieser Publikation

Ziel dieser Publikation ist es, die Diskussion um Ergebnissteuerung und Evidenz sowie die Rolle der Erziehungswissenschaft im Kontext dieser Paradigmen zu befördern. Wir stehen sehr am Anfang. Die Publikation ist Resultat einer Konferenz der Kommission „Bildungsorganisation, Bildungsplanung und Bildungsrecht“ in der Sektion „Empirische Bildungsforschung“ der DGfE. Die Tagung wurde durch das Bundesministerium für Bildung und Forschung unterstützt. Durch diese Hilfe konnten wir eine beträchtliche Anzahl von Nachwuchswissenschaftlern sowie internationale Gastreferenten gewinnen. Unser Dank geht persönlich an Frau Dr. Buchhaas-Birkholz.

Ziel der Tagung war es, folgende Fragen zu klären: Auf welche Formen wissenschaftlicher Wissensproduktion ist eine evidenzbasierte Bildungspolitik und -praxis angewiesen? Kann die Agenda wissenschaftlicher Wirkungsforschung überhaupt sinnvoll in eine bildungspolitische Agenda übertragen werden? Wird wissenschaftliche Forschung vor einem solchen Hintergrund gestärkt – oder wird sie technologisch vereinseitigt und in ihrer Autonomie und Kritikfähigkeit beschnitten? Können Wirkungsevaluationen über experimentelle Wirkungsbeschreibungen von Programmen hinausgehen und verallgemeinerungsfähige Wirkungserklärungen liefern? Legt eine evidenzbasierte Steuerung eine stärkere Professionalisierung der Bildungs- und Erziehungspraxis nahe oder impliziert sie einen Prozess der Deprofessionalisierung durch Standardisierung?

Im Folgenden präsentieren wir Ihnen die Ergebnisse, die wir in Anlehnung an die eingangs skizzierten Problemfelder systematisieren.

Teil 1 dieses Sammelbandes beschäftigt sich mit den Formen und Methoden, aber auch den Problemen der Wissensgenerierung, wagt also eine Annäherung an die Erziehungswissenschaft als „schwierige Wissenschaft“.

Inwiefern Metaanalysen geeignet sind, Wissen über das Bildungswesen zu generieren und auf diese Weise eine professionelle Entscheidungsfindung zu unterstützen, diskutiert David Gough. Eine seiner Kernaussagen ist die, dass man sich auf die Ergebnisse einer einzelnen empirischen Studie nicht verlassen könne. Er führt verschiedene Begründungen für diese These an und erläutert sie an interessanten Beispielen. Er beschreibt mit „Expert Opinion“ und „Literature Reviews“ Methoden, die bei forschungsbasierter Entscheidungsfindung hilfreich, die aber wenig robust sind. Eine Alternative böten „Systematic Reviews“. Sie haben selbst den Status wissenschaftlicher Forschung und folgen hohen methodischen Standards und einem detaillierten und transparenten Ablauf. Gough beschreibt die möglichen Verfahren mit ihren Stärken und den Problemen, die sie überwinden müssen. Systematic Reviews ermöglichen, das breite erziehungs- und sozialwissenschaftliche Wissen, das einzelne empirische Forschungen generiert haben, in einer Art und Weise zusammenzubringen, dass evidenzbasierte Entscheidungen möglich werden. Die bisherige Vernachlässigung dieser Synthetisierung will Gough überwinden.

Mit der Schwierigkeit, Indikatoren als Form objektivierten Wissens zu gewinnen, beschäftigten sich gleich zwei Beiträge: Stefanie Albus, Heinz-Günter Micheel, Hans-Uwe Otto und Andreas Polutta widmen sich den methodologischen Herausforderungen, mit denen sich die Jugendhilfeforschung durch die Forderung nach empirisch fundierter Wirkungsforschung konfrontiert sieht. Die Autoren gehen der Frage nach, welches Kriterium und welche Indikatoren für die Bewertung des Erfolges bzw. Misserfolges einer jugendhelferischen Intervention geeignet wären. Ihr Beitrag ist zugleich ein deutliches Plädoyer, sich intensiv mit der Wahl von Wirkungsindikatoren in der Jugendhilfe auseinanderzusetzen.

Inka Bormann und Thomas Krikser hingegen verfolgen die Frage, welche unterschiedlichen Anforderungen unterschiedliche Akteure des Bildungswesens an Indikatoren stellen, und welche Konsequenzen dies für die Form der Handlungskoordination bedeutet. Auf Seiten von Politik, Wissenschaft sowie den Vertretern zivilgesellschaftlicher Organisationen verweisen sie zwar durchaus auf eine grundsätzliche Einigkeit über die Funktion von Indikatoren; zugleich arbeiten sie jedoch auch divergierende Vorstellungen hinsichtlich ihres Einsatzes und ihrer Entwicklung heraus, die sie auf differente Rationalitäten und Handlungslogiken zurückführen.

Hans-Peter Blossfeld, Jörg Doll und Thorsten Schneider schließlich wenden sich der bislang weitgehend unbeantworteten Frage zu, welche Faktoren den Bildungserwerb in verschiedenen Lebensumwelten und im Laufe eines Lebens beeinflussen, und wie die kumulativen Prozesse des Bildungserwerbs verlaufen. Mit der Nationalen Bildungspanelstudie (NEPS) skizzieren sie ein Projekt, das bereits in wenigen Jahren umfangreiches Analysepotential zur Klärung dieser – und anderer – Fragen erhoffen lässt. Ausführlich schildern sie die Zielstellung, die Notwendigkeit, die Rahmenkonzeption sowie die zentralen Forschungsfragen, die NEPS zugrunde liegen.

Teil 2 dieses Bandes lenkt die Aufmerksamkeit auf die Rolle, die der Erziehungswissenschaft als Begleiter von Politik und Praxis zukommt.

Steve Fleischman widerspricht der Annahme, evidenzbasierte Erziehungswissenschaft würde die Pädagogen auf den Status von Ausführenden reduzieren. Im Gegenteil behauptet er, dass die Advokaten der Evidenzbasierung den Fachkräften und ihrer Professionalität eine besondere Rolle zusprechen. Auffällig jedoch sei, dass Evidenzen in der Praxis zu selten handlungsleitend sind. Dafür gebe es Gründe. Die Nutzer, so Fleischman, seien meist nicht involviert in der Entwicklung einer Forschungsfrage, und die Wissenschaft sei kaum darum interessiert, dass ein Erzieher oder eine Lehrerin tatsächlich mit den Ergebnissen der Forschung etwas Nutzbringendes anfangen kann. Falls es doch – selten genug – so sei, wird nicht mehr gefragt, was er oder sie damit anfängt: „Evidence is not enough“. Es müsse den Nutzern nicht nur vermittelt werden, welche pädagogische Intervention funktioniert, sondern auch, wie man es erreicht, dass sie funktioniert. Fleischman entwickelt Unterschiede und Gemeinsamkeiten von „user-driven“ und „evidence-driven research“. Er skizziert einige in dieser Hinsicht viel versprechende Ansätze in den USA. Er zeigt, dass nutzergetriebene Forschung durch hohen Respekt vor der Arbeit der Praktiker geprägt ist, ohne dass die Idee der Evidenzbasierung aufgegeben werden müsse.

Mit einem in Deutschland noch vergleichsweise innovativen Reformansatz beschäftigen sich Nils Berkemeyer und Nils van Holt. Sie gehen der Frage nach, welchen Beitrag Netzwerkarbeit an Schulen im Zuge der Begleitung bildungspolitischer Innovationspro-

zesse leisten kann. In Ihrem Beitrag untersuchen sie, wie schulische Netzwerke von wissenschaftlich elaboriertem Wissen über Netzwerke profitieren können. Ergebnisrückmeldungen müssten demnach nicht nur wahr und für den Praktiker „hoch brauchbar“ sein: Vielmehr müsse eine Umsetzung unter den gegebenen Bedingungen auch möglich sein.

Auch Elke Grimminger stellt die Frage nach Gelingensbedingungen bildungspolitischer Interventionen, wendet sich aber dem Bereich der Lehrerfortbildung zu. In ihren Beitrag diskutiert sie die Frage, wie es im Zuge einer zunehmend multikulturellen Gesellschaft gelingen kann, Lehrkräften interkulturelle Kompetenz zu vermitteln. Dabei gelangt sie zu Ergebnissen, die auch aus governancetheoretischer Perspektive Aufmerksamkeit verdienen: Entscheidenden Einfluss auf die Umsetzung der Fortbildungsinhalte übten weniger fehlende Kompetenz- oder Performanzentwicklungen als vielmehr strukturelle, schulorganisatorische und persönliche Probleme aus.

Ursula Fiechter, Barbara Meili, Larissa Trösch und Birgit Weinmann-Lutz hingegen spüren der Frage nach, ob – und wenn ja: wie – sich epistemologische Überzeugungen im Verlauf der Lehrerausbildung sowie der Berufspraxis verändern. Auf der Basis qualitativer wie quantitativer Untersuchungen kommen sie zu dem Schluss, dass die Befragten postrelativistische epistemologische Überzeugungen insgesamt bevorzugten, zumindest ein Teil von ihnen aber in absolutistische Positionen zurückgefallen ist. Die Autorinnen fordern deshalb eine Lehrerbildung, in der alle Ausbildungsteile gleichermaßen von wissenschaftlich fundierter Reflexion geprägt sind.

Britta Oerke und Katharina Maag Merki lenken die Aufmerksamkeit auf die Schülebene. Sie untersuchen, welchen Einfluss der Wechsel von dezentralen zu zentralen Abiturprüfungen auf die Leistungsattributionen der Schüler nimmt. Demnach neigen Schüler in zentral geprüften Leistungskursen eher als bei dezentral geprüften Leistungskursen dazu, einen Erfolg im Abitur auf die Lehrperson sowie die Vorbereitung im Unterricht zurückzuführen. Zudem nähmen Teilnehmer an zentralen Prüfungen den erteilten Unterricht als besser wahr. Hinweise auf einen Kontrollverlust hingegen ließen sich im Zuge der Einführung des Zentralabiturs auf Seiten der Schüler nicht finden.

Dass sich (erziehungswissenschaftliches) Monitoring nicht nur als begleitende und unterstützende, sondern zugleich auch als kritische Wissenschaft verstehen sollte, zeigen die folgenden fünf Beiträge:

So lenkt Eckhard Rohrman die Aufmerksamkeit auf die Frage, welche Folgen das dreigliedrige Schulsystem für Schülerinnen und Schüler hat. In seinem Beitrag konzentriert er sich auf Kinder und Jugendliche mit sonderpädagogischem Förderbedarf, deren Scheitern im etablierten Regelschulsystem nach Ansicht des Verfassers nicht allein auf individuelle Probleme der Betroffenen, sondern auch auf das staatliche Schulwesen zurückzuführen sei. Vor dem Hintergrund internationaler Menschenrechte und auf der Basis aktueller empirischer Daten votiert Rohrman für einen flächendeckenden Ausbau sowie eine konzeptionelle Weiterentwicklung der Integrativen Gesamtschule.

Einen wesentlichen Schritt hin zu einer gemeinsamen Unterrichtung von Kindern mit und ohne sonderpädagogischen Förderbedarf hat der Freistaat Bayern im Jahr 2000 mit der Einführung sogenannter „Kooperationsklassen“ getan. Christoph Ratz und Eva Stumpf untersuchen, welche Effekte die Beschulung von Kindern mit sonderpädagogischem Förderbedarf in bayerischen Kooperationsklassen auf deren Entwicklung nimmt, und inwiefern sich die integrierte Beschulung möglicherweise nachteilig auf die Grund-

schüler der Kooperationsklassen auswirkt. Dabei kommen sie zu dem Ergebnis, dass die gemeinsame Unterrichtung von Kindern mit und ohne sonderpädagogischen Förderungsbedarf durchaus Potential habe; allerdings gelänge es den bayrischen Kooperationsklassen nicht, dieses Potential auszuschöpfen.

Sebastian Stark und Christine Wiezorek beschäftigen sich mit einem der zentralsten Steuerungsinstrumente der bundesdeutschen Bildungspolitik, den nationalen Bildungsstandards. In ihrem Beitrag verweisen sie nicht nur auf deutliche Umsetzungs- und Durchsetzungsschwierigkeiten der Reform, sondern diskutieren auch die Frage, inwiefern die verabschiedeten Standards grundsätzlich zur Erfüllung des staatlichen Bildungs- und Erziehungsauftrages geeignet seien. Eine Fokussierung schulischer Bildungsarbeit auf die Umsetzung der KMK-Vorgaben, so die Autoren, drohe das Spannungsfeld von individualisierter Leistungsförderung und Gemeinschaftsorientierung einseitig und zu Lasten eines auf Konkurrenz und Selektion reduzierten „Bildungs“-Verständnisses aufzulösen.

Auch Franz Schott und Shahram Azizi Ghanbari setzen sich kritisch mit der Konstruktion von Bildungsstandards auseinander. Sie vertreten einen Kompetenzbegriff, der sich in wesentlichen Punkten von jenem unterscheidet, den führende Vertreter der aktuellen Bildungsforschung nutzen – und der sich auch in der bildungspolitischen Praxis weitgehend etabliert zu haben scheint. Schott und Azizi Ghanbari votieren für ein Kompetenzverständnis, dem zufolge der Rekurs auf Aufgabenmengen und Kompetenzgrade unerlässlich sei, das neben dem kognitiven auch den sozial-emotionalen und den psychomotorischen Bereich berücksichtige und schließlich die Eigenschaft einer Kompetenz als Präskription betone.

Mit dem Paradigmenwechsel von einer Angebot- zu einer Nachfrageorientierung beschäftigt sich schließlich Josef Muth. Er analysiert die Effekte und Wirkungen der Bildungsschecks, die das nordrhein-westfälische Arbeitsministerium im Jahre 2006 zur Stärkung der Beschäftigungsfähigkeit von Arbeitnehmern sowie die Wettbewerbsfähigkeit kleiner und mittelständischer Unternehmen einführte. Die bislang vorliegenden Daten zur Programmumsetzung, den erreichten Teilnehmergruppen sowie den Mitnahme- und Anschubeffekten ließen insgesamt positive Wirkungen des neuen Steuerungsinstrumentes vermuten. Nichtsdestotrotz könnten und sollten nachfrageorientierte Ansätze eine angebotsorientierte Förderung nicht vollständig ersetzen.

Teil 3 des Bandes schließlich thematisiert das bisweilen schwierige, oftmals auch distanzierte Verhältnis zwischen Erziehungswissenschaft, Politik und Praxis.

Im Feld des Kinderschutzes in England und Wales setzen sich David Wastell und Sue White mit der zunehmende Standardisierung und Technologisierung der Kinder- und Jugendhilfe auseinander. Die „IT-enabled Initiatives“ werden von White und Wastell als eine spezifische, managerielle Form der Modernisierung und der Implementation einer wirkungsorientierten Programmatik identifiziert. Bezüglich der Effektivität, aber auch der feldspezifischen Angemessenheit der IT basierten Reformen kommen sie zu einer ernüchternden Einschätzung. Die Qualität des Kinderschutzes hat sich durch den systematischen Einsatz von Informationstechnologien zu Steuerung professioneller Praxis nicht verbessert.

Kathrin Dederling verweist auf die Schwierigkeiten, die die Interpretation wissenschaftlicher Ergebnisse auf ministerieller Ebene bereitet. Sie präsentiert Ergebnisse des

Projektes MiSteL, das den Umgang von vier Landesschulministerien mit den Ergebnissen der internationalen Vergleichsstudie PISA 2000 untersucht hat. In den Bereichen von Bildungspolitik und Bildungsadministration konstatiert sie für den gewählten Untersuchungszeitraum eine allenfalls begrenzt evidenzbasierte Entscheidungsfindung. Ursächlich hierfür seien unter anderem ein erheblicher Legitimationsdruck sowie Kompetenz- und Kapazitätsdefizite auf Seiten der Ministerien gewesen.

Doch auch auf der Ebene der Einzelschule scheint die Nutzung empirisch gewonnener Daten als schwierig. Herbert Altrichter beschäftigt sich in seinem Beitrag mit der Rolle, die Datenfeedbacks im Rahmen der Unterrichtsentwicklung zukommt. Trotz aller Hoffnungen und Erwartungen sei die Wirksamkeit von Datenfeedbacks empirisch bislang kaum belegt; vielmehr deuteten aktuelle empirische Studien auf tendenziell eher unbefriedigende Steuerungswirkungen hin. Altrichter liefert Erklärungsansätze für diese bildungspolitisch unbefriedigende Situation – und gibt damit zugleich erste Hinweise auf ihre Überwindung.

Aline Kadach und Olga Zlatkin-Troitschanskaia stellen die Frage, welche Funktionen der internen Evaluation im Rahmen der Neuen Steuerung des Bildungswesens zukommen. Sie verweisen nicht nur auf tendenziell positive Grundhaltungen der Lehrerschaft gegenüber der Schulentwicklung, sondern konstatieren auf einzelschulischer Ebene Zusammenhänge und positive Assoziationen zwischen den positiven Einschätzungen der Schulautonomie, der Zieltransparenz und der erfolgten Unterrichtsentwicklung mit der positiven Wahrnehmung der Evaluation. Insofern komme der internen Evaluation eine besondere Schlüsselrolle im Rahmen erfolgreicher Schul- und Unterrichtsentwicklung zu.

Das Verhältnis von individueller und kollektiver Innovationsbereitschaft auf einzelschulischer Ebene betrachten Ramona Buske und Olga Zlatkin-Troitschanskaia. Sie entwickeln zunächst ein eigenständiges, theoretisch fundiertes Modell kollektiver Innovationsbereitschaft, das sie im Folgenden auf der Basis empirisch gewonnener Daten überprüfen. Dabei können sie einen kausalen Zusammenhang zwischen kollektiver und individueller Innovationsbereitschaft nicht bestätigen, jedoch einen schwachen Zusammenhang zwischen der kollektiven Innovationsbereitschaft und der Reformumsetzung an der Schule ausmachen.

Literatur

- Berliner, D. C. (2002). Educational Research: The Hardest Science of All. *Educational Researcher*, 31 (8), 18–20.
- Biesta, G. (2007). Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research. *Education Theory*, 57 (1), 1–22.
- Böttcher, W. (2002). *Kann eine ökonomische Schule auch eine pädagogische sein?* Weinheim und München: Juventa.
- Böttcher, W. (2008). Evaluation von Schule und Jugendhilfe. In Coelen, T. & Otto, H.-U. (Hrsg.), *Grundbegriffe Ganztagsbildung. Das Handbuch* (S. 889–898). Wiesbaden: VS.

- Böttcher, W. (2009). Was leisten Evaluationen für die Qualitätsentwicklung? In Bohl, T. & Kiper, H. (Hrsg.), *Lernen aus Evaluationsergebnissen. Verbesserungen planen und implementieren* (S. 207–217). Bad Heilbrunn: Julius Klinkhardt.
- Brunsson, N. (1982). The Irregularity of Action and Action Rationality: Decisions, Ideologies and Organizational Actions. *Journal of Management Studies*, 19 (1), 29–44.
- Geißler, K. A., Laske, S. & Orthey, A. (Hrsg.) (2002). *Handbuch Personalentwicklung. Beraten, Trainieren, Qualifizieren*. Köln: Deutscher Fachverlag.
- Hammersley, M. (2009). What is Evidence for evidence-based Practice? In Otto, H.-U., Polutta, A. & Ziegler, H. (Hrsg.), *Evidence-based Practice – Modernising the Knowledge Base of Social Work?* Opladen, Farmington Hills: Barbara Budrich.
- Joint Committee on Standards for Educational Evaluation / Sanders, J.R. (Hrsg.) (2006). *Handbuch der Evaluationsstandards*. Wiesbaden: VS Verlag.
- Kahl, R. & Spiewak, M. (2005). Nur bedingt wissenschaftlich. Die Erziehungswissenschaften haben in der Forschung und der Lehrerbildung versagt. Eine Polemik. *Die Zeit*, Heft 13, 11. März.
- Kindler, H. (2008). Developing Evidence-Based Child Protection Practice: A View From Germany. *Research on Social Work Practice*, 18, 319–324.
- Leicht, K., Walter, T., Sainsaulieu, I. & Davies, S. (2009). New Public Management and New Professionalism across Nations and Contexts. *Current Sociology*, 57, 581–605.
- Merkens, H. (2006). Bildungsforschung und Erziehungswissenschaft. In Merckens, H. (Hrsg.), *Erziehungswissenschaft und Bildungsforschung* (S. 9–20). Wiesbaden: VS.
- Mertens, D. M. (2005). *Research and Evaluation in Education and Psychology*. Thousand Oaks CA: Sage.
- Otto, H.-U., Polutta, A. & Ziegler, H. (2009). A Second Generation of Evidence-Based Practice: Reflexive Professionalism and Causal Impact in Social Work. In Otto, H.-U., Polutta, A. & Ziegler, H. (Hrsg.), *Evidence-based Practice – Modernising the Knowledge Base of Social Work?* Opladen, Farmington Hills: Barbara Budrich.
- Pawson, R. & Tilley, N. (2004). *Realistic Evaluation*. London u. a.: Sage Publications.
- Platzmeyer, F. & Kohlmeyer, M. (2009). *Finanzierungsmodelle im Kontext von wirkungsorientierter Steuerung der Hilfen zur Erziehung*. Münster: Institut für Soziale Arbeit.
- Rossi, P., Lipsey, M. & Freeman, H. (2004). *Evaluation. A Systematic Approach*. Thousand Oaks, London, New Delhi: Sage Publications.
- Schoen, D. A. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Weegen, M., Böttcher, W., Bellenberg, G. & van Ackeren, I. (Hrsg.) (2002). *Bildungsforschung und Politikberatung. Schule, Hochschule und Berufsbildung an der Schnittstelle von Erziehungswissenschaft und Politik*. Weinheim/München: Juventa.
- Weiler, H. N. (2003). Bildungsforschung und Bildungsreform – Von den Defiziten der deutschen Erziehungswissenschaft. In Gogolin, I. & Tippelt, R. (Hrsg.), *Innovation durch Bildung* (S. 181–203). Opladen: Leske und Budrich.

David Gough

Qualitative, Quantitative and Mixed Methods Systematic Reviews to Support Professional Decision Making in Education

1. What do we know from research?

Education is a complex subject that can be understood and studied in many different ways. Research to assist in our understanding of education will thus mirror the complexity of the subject of education. There are very many different questions that we may ask for help from research. These different questions can be from many different ideological and theoretical perspectives and relevant research to answer these questions may use a variety of different methods based on a variety of different research paradigms. In sum, research in education is not just one thing. It is a mosaic of different things. It is research because it uses rigorous and explicit methods to develop insight, understanding and data. The findings of the research will be determined by the perspectives and methods that generated them.

This diversity of research provides a rich tapestry to understand the nature, practice and outcomes of education. In addition, the amount of research produced each year is so great that not even full time researchers can keep up to date with more than a small proportion of research produced in the very many different sub areas of educational research. This is problematic. If all the research that is being produced is not being read then this is a waste of all that creativity, hard work and resources.

The problem of accessing research is even greater for non researchers who should benefit from the results of research such as teachers, managers, policy makers, school students and their parents, and other members of the community. They do not have time to identify and access and read the research that is often published in many different journals, reports and books. Also, they do not necessarily have the skills to interpret the quality and relevance of the findings of the studies or to integrate them into a form that will assist them to make decisions informed by those findings.

The traditional solutions to these problems of access, appraisal and integration of research findings has been to rely on just a few individual studies rather than to consider all relevant studies, to undertake reviews of the literature and to use expert summaries of what is known. Unfortunately, all three of these solutions have weaknesses.

1.1 Reliance on a few research studies

First, any individual study may provide misleading results. One reason is because the research method may have been inappropriate. Another reason is that the research method, even if appropriate, may not have been executed appropriately.

Even if an appropriate method is used and executed properly, there may be chance variation in the results. If you are undertaking a study using a sample to make conclu-

sions about a population then the results may be based on probabilistic assumptions that may not happen to pertain. For example, a study may conclude that there is a significant difference between two groups. It is considered significant as such a difference would only happen 5 times out of a 100 by chance (95% significance level). However, this study might be one of those 5% that had such a difference by chance. Furthermore, the sample and the context in which it was studied may be very different from the situation in which you want to apply the findings and there may be contextual issues which would make the results of the study misleading. Even if you are not attempting to make statistical judgements about a population, any one study may only provide a limited perspective about an issue.

There are many examples where serious decisions have been made on a limited number of studies and it is only when all of the relevant research has been considered that it has become clear that the findings of the few studies considered had been misleading and the decisions based on them has had been misinformed. One example is the use of Albumin as a drug to treat shock after burns. For many years Albumin was the internationally accepted treatment based on some theoretical reasoning and a few well know studies seemingly showing a positive effect of the treatment. However, when all of the research studies were examined, it was found that Albumin treatment increased rather than decreased the chance of death compared to other treatments (Alderson, Bunn, Li Wan Po, Li, Roberts & Schierhout, 2004). Individual doctors had with the best of intentions being prescribing Albumin and each doctor did not see enough patients to notice the lack of positive impact. If their patients recovered, they probably thought that it was an effective drug. If the patient died, they probably assumed that the drug was effective but the patient was too sick to recover.

In some cases people may feel that the positive impact of some educational approach is obvious and so research is not really necessary. The difficulty is that such common sense is not always obvious. For example, there are many well intentioned people advocating scared straight strategies where those guilty of crimes explain to school students the error of their ways and not to follow their bad example and end up in prison. A review of all research shows that, contrary to the expectations of these well intentioned people, such programmes increase not decrease law breaking by young people (Petrosino, Turpin-Petrosino & Buehler, 2002).

1.2 Expert opinion

One way to make research findings more accessible is for academics and other researchers to provide expert opinions about the state of research knowledge. Experts have specialist skills knowledge and experience that can obviously provide useful information. This knowledge could be one or more of the following:

1. Expert in research synthesis knowledge: knowledge of their own and other researchers' studies including the methods and results of the studies and the overall research knowledge arising from this.

2. Expert research practice knowledge: practice knowledge as a researcher about the nature of the phenomena being inquired about.
3. Expert practice knowledge: knowledge as a professional practitioner in the discipline under study, rather than a researcher in that discipline. For example, a teacher rather than an educational researcher.

All three of these types of knowledge can have great value but there are two types of danger, firstly confusion between and within the types of knowledge, secondly confusion within the types of knowledge. Confusion between the types of knowledge is well illustrated by cases where experts provide expert witness testimony in courts where they may have high credibility in one type of knowledge but provide expert testimony in another. For example, where they are a highly respected medical clinician and give evidence on research knowledge which they may have some experience in but are less of an expert. There have been several court cases in the United Kingdom where there have been these concerns about medical expert witness testimony in cases of child deaths.

Confusion within a type of knowledge can occur because there may not be explicit boundaries to an expert's knowledge. These boundaries could be in terms of areas of specialism with the expert knowing more about the research in some sub areas than others. For example: an expert in research language learning for a particular age of student or for a particular theoretical or practical perspective. If these boundaries or perspectives are not clear then the meaning of the expert knowledge is also not clear. This can be very clearly seen in the debates about the nature of psychiatry and critical psychiatry in the 1950s or the organisation of maternity services in the 1970's.

1.3 Literature reviews

Another way to enable research findings to be more accessible is to undertake reviews of the literature. Research specialists identify, appraise and synthesize the findings of research so that all others interested in research findings can access more easily than themselves trying to keep up to date with all of the relevant research. Like expert opinion, these can be very useful but they can be subject to the same problems of lack of clarity about the boundaries of the research being considered. More fundamentally they do not always specify how the review was undertaken making the findings of the reviews difficult to interpret. This has led many to advocate systematic reviews; that is reviews that have explicit methods so that the basis of how the results of the reviews have been produced are made clear and their relevance understood.

2. Systematic reviews

With primary research studies we expect rigorous methods of research to be used and for the report of a study to include a clear explanation of the aims of the research, the methods used and the basis of the results and conclusions. In this way the method of production of the results is explicit and accountable. Reviews of research bring together the findings of primary research and in doing so they are themselves pieces of research; a