

1

2 0 1 2

Aspekte „Neuer Steuerung“?

Heinrich Mintrop/Gail Sunderman

Zentrale Steuerung von Schulentwicklung mit Quoten und Sanktionen

Barbara Asbrand/Nina Heller/Sigrid Zeitler

Die Arbeit mit Bildungsstandards in Fachkonferenzen

Franziska Schulze

Folgen zentraler Lernstandserhebungen

Fabian Dietrich/Maike Lambrecht

Schulinspektion und die Hoffnung auf den zwanglosen Zwang der „besseren Evidenz“

Tobias Feldhoff/Luzia Durrer/Stephan G. Huber

Steuerung eines Schulsystems

Uwe Maier/Harm Kuper

Vergleichsarbeiten als Instrumente der Qualitätsentwicklung an Schulen

Bildungsforschung – disziplinäre Zugänge

Knut Schwippert

Empirische Bildungsforschung: Perspektiven der Erziehungswissenschaft

INHALT

EDITORIAL

Martin Heinrich/Fabian Dietrich

**Editorial zum Schwerpunktthema:
Aspekte „Neuer Steuerung“?**5

ASPEKTE „NEUER STEUERUNG“?

Heinrich Mintrop/Gail Sunderman

**Zentrale Steuerung von Schulentwicklung mit Quoten
und Sanktionen: Das US-amerikanische Großexperiment in
Schulen für benachteiligte Bildungsschichten**8

Barbara Asbrand/Nina Heller/Sigrid Zeitler

**Die Arbeit mit Bildungsstandards in Fachkonferenzen
Ergebnisse aus der Evaluation des KMK-Projektes *for.mat*** 31

Franziska Schulze

**Folgen zentraler Lernstandserhebungen: Unterscheiden sich Lehrkräfte
in der Auseinandersetzung mit Lernstandserhebungen?
Empirische Befunde zu Formen unterrichtsbezogener Steuerungslogik**43

Fabian Dietrich/Maike Lambrecht

**Menschen arbeiten mit Menschen
Schulinspektion und die Hoffnung auf den zwanglosen
Zwang der „besseren Evidenz“**57

Tobias Feldhoff/Luzia Durrer/Stephan G. Huber

**Steuerung eines Schulsystems
Eine empirische Analyse, wie Akteure die Steuerungskonfigurationen des
Schulsystems wahrnehmen und sich deren zukünftige Gestaltung wünschen**.....71

BERICHT

Uwe Maier/Harm Kuper

Vergleichsarbeiten als Instrumente der Qualitätsentwicklung an Schulen

Überblick zum Forschungsstand.....88

BILDUNGSFORSCHUNG – DISZIPLINÄRE ZUGÄNGE

Knut Schwippert

Empirische Bildungsforschung: Perspektiven der Erziehungswissenschaft..... 100

Die
Deutsche
Schule

Zeitschrift für
Erziehungswissenschaft
Bildungspolitik und pädagogische Praxis

DDS

Herausgegeben von der Gewerkschaft Erziehung und Wissenschaft

2

2012

Vorschau

Themenschwerpunkt: Schulwettbewerbe und Best Practice

Es gibt mittlerweile eine Reihe von Wettbewerben, mit denen die Qualität einzelner Schulen im Sinne von Best Practice ausgezeichnet wird. Den Schulen und ihren Konzepten wird damit eine Leitbildfunktion zugeschrieben, die sie wiederum zur Profilierung in der Bildungslandschaft nutzen können. Wie Wettbewerbe und Preise tatsächlich von Schulen, aber auch von anderen Akteuren, wie z.B. Stiftungen als Initiatoren entsprechender Ausschreibungen, genutzt werden und inwieweit Entwicklungsprozesse in Schulen befördert werden können, ist bislang nicht systematisch in den Blick genommen worden. Ebenso fehlt ein systematisierender Überblick über unterschiedliche Instrumente der Auszeichnung. In diesem Zusammenhang erweist sich auch der häufig bemühte Best Practice-Begriff zumindest hinterfragbar, scheint er doch für ein vereinfachtes Transfervverständnis zu stehen. Im Schwerpunkt wird diesen Fragen theoretisch sowie im Rahmen von Fallstudien nachgegangen; nicht zuletzt kommt eine Preisträgerschule selbst zu Wort.

Heft 2/2012 der Zeitschrift erscheint im Mai 2012.



Waxmann • Steinfurter Str. 555 • 48159 Münster • www.waxmann.com

CONTENTS

EDITORIAL

Martin Heinrich/Fabian Dietrich

Editorial to the Focus Topic:

Aspects of “New Governance”?5

ASPECTS OF “NEW GOVERNANCE”?

Heinrich Mintrop/Gail Sunderman

**Centralized Management of School Improvement with Quota and
Sanctions: The US-American Experiment with High-Stakes**

Accountability in Schools Serving Disadvantaged Populations.....8

Barbara Asbrand/Nina Heller/Sigrid Zeitler

Working with Educational Standards in Faculty Meetings

Results from the Evaluation of the KMK-Project *for.mat*31

Franziska Schulze

Consequences of Standardized Performance Tests:

Do Teachers Differ in Their Interpretation of Standardized Performance Tests?

Empirical Findings to Types of a Teaching-related ‘Control Logic’43

Fabian Dietrich/Maike Lambrecht

People Work with People

School Inspection and the Hope of the Unforced Force of “Better Evidence”57

Tobias Feldhoff/Luzia Durrer/Stephan G. Huber

Governance of a School System

An Empirical Analysis of How Actors Perceive the Governance Configurations

of the School System and How They Want Its Future Configuration to Be.....71

REPORT

Uwe Maier/Harm Kuper

Mandatory Testing as an Instrument for School Quality Development

A Research Survey from an International Perspective88

EDUCATIONAL RESEARCH – DISCIPLINARY APPROACHES

Knut Schwippert

Empirical Educational Research: Perspectives of Educational Sciences 100

**Die
Deutsche
Schule**

Zeitschrift für
Erziehungswissenschaft
Bildungspolitik und pädagogische Praxis
Herausgegeben von der Gewerkschaft Erziehung und Wissenschaft

DDS

2

2012

Preview

Focus Topic: School Competitions and Best Practice

By now, there are a number of competitions by which the quality of individual schools is awarded in the sense of best practice. Thus, a model function is attributed to these schools and to their concepts, which can be used by them to promote themselves in the educational landscape. Up to now, it has neither been taken into account systematically how competitions and awards are made use of by schools, but also by other actors, e.g. by the foundations that initiate the respective competitive biddings, nor to what extent development processes in schools can be conveyed thereby. Likewise a systematizing survey of different kinds of awards is missing. In this context the often used term of best practice turns out to be at least a questionable one as it seems to stand for a simplified understanding of transfer. The focus topic addresses these issues theoretically as well as in the context of case studies; last but not least the floor is given to an award-winning school.

No 2/2012 will be out in May 2012.



Waxmann • Steinfurter Str. 555 • 48159 Münster • www.waxmann.com

Editorial zum Schwerpunktthema: Aspekte „Neuer Steuerung“?

Editorial to the Focus Topic: Aspects of „New Governance“?

„Neue Steuerung“ ist in den letzten Jahren zu einem prominenten Schlagwort im bildungspolitischen Diskurs geworden. Terminologisch handelt es sich zunächst um einen Abgrenzungsbegriff gegenüber einer ‚alten‘ und damit als ‚nicht mehr zeitgemäß‘ qualifizierten Steuerung. Unklar bleibt in den theoretischen und empirisch fundierten Diskursen allerdings, worin genau das ‚Neue‘ jener Steuerung besteht bzw. bestehen soll oder inwiefern hier nicht auch ‚alte Steuerung in neuem Gewand‘ Einzug ins System hält.

Deutlich wird dabei auch, dass die Verwendung des Begriffs ‚Neue Steuerung‘ im Singular eine theoretisch-systematische Konsistenz und Kohärenz impliziert, die bislang nicht gegeben ist. Unstrittig ist, dass Steuerungsinstrumente wie etwa Bildungsstandards, Leistungstests, aber auch Schulinspektionen zu den neuen Formen der Handlungskoordination zwischen Bildungspolitik, Bildungsverwaltung und den Professionellen vor Ort zählen, die der ‚Neuen Steuerung‘ zugerechnet werden. Diese prägen die aktuellen Diskussionen über die „deutsche Schule“ folgenreich. Daher widmet sich Heft 1/2012 der DDS solchen „Aspekten ‚Neuer Steuerung““. Deren Vielgestaltigkeit unterstreicht, dass ‚Neue Steuerung‘ keineswegs einer in sich kohärenten Steuerungslogik folgt und dementsprechend auch nicht etwa mit ‚Deregulierung‘ oder ‚Dezentralisierung‘ gleichzusetzen ist. Bei aller Heterogenität und begrifflichen Unklarheit zeigt sich jedoch auch in den im Themenschwerpunkt versammelten Beiträgen als gemeinsamer Nenner jener ‚Neuen Steuerung‘ ein Bedeutungsgewinn der Frage nach den Rezeptions- bzw. Umgangsweisen mit den verschiedenen Steuerungsimpulsen auf der Ebene der professionellen Praxis. Dieses verstärkte Interesse lässt erkennen, dass sich ‚Neue Steuerung‘ in einer manifesten oder latenten Enttäuschung eines vormals handlungsleitenden Planungs- oder Steuerungsoptimismus begründet. Davon ausgehend richtet sich das Interesse im engeren evaluativen Sinne auf die Überprüfung der Wirksamkeit von Steuerungsversuchen oder allgemeiner auf deren jeweils eigensinnige Adaption seitens der verschiedenen betroffenen Akteure.

Heinrich Mintrop und Gail Sunderman untersuchen die in den USA insbesondere im Zuge des „No Child Left Behind Act“ vorangetriebene Zentralisierung der Schulentwicklung. Bezogen auf das bildungspolitische Ziel der Förderung von Schülerinnen und Schülern aus sozial benachteiligten Bildungsschichten gehen sie der Frage nach, welche intendierten und nicht-intendierten Folgen diese US-spezifische Variante ‚Neuer Steuerung‘ zeitigt. Die sich eröffnende Möglichkeit einer internationalen Kontrastierung lässt auch erkennen, dass die Auswirkungen einer solchen Zentralisierung immer auch von der konkreten Gestalt und Qualität der lokalen Steuerung und professionellen Selbststeuerung abhängen, auf die sie trifft.

Im Beitrag von Barbara Asbrand, Nina Heller und Sigrid Zeitler wird mit *for.mat* ein Instrument zur Implementation einer kompetenz- und standardbasierten Unterrichtsentwicklung evaluiert. Damit geht es um einen Steuerungsimpuls, der sich unmittelbar auf eine Veränderung der professionellen Praxis von Lehrerinnen und Lehrern richtet. Über die Erfassung expliziter Bewertungen der evaluierten Fortbildungsmaßnahme hinaus werden hier implizite handlungsleitende Werthaltungen rekonstruiert, welche die Alltagspraxis der Lehrkräfte prägen.

Auch Franziska Schulze fokussiert auf die Veränderungen, welche ein Steuerungsimpuls auf die professionelle Praxis hat: Sie untersucht die „unterrichtsbezogene Steuerungslogik“, welche sich in der Auseinandersetzung von Lehrkräften mit den Ergebnissen zentraler Lernstandserhebungen zeigt. Den Bezug zwischen Steuerung und unterrichtlicher Praxis stellt sie dadurch her, dass sie Unterrichtsplanung selbst als Steuerungspraxis konzipiert.

Fabian Dietrich und Maike Lambrecht nehmen Schulinspektionen als Instrument einer evidenzbasierten Steuerung in den Blick. Dabei kontrastieren sie die bildungspolitische Programmatik, mit der die Etablierung der Schulinspektionen begründet und legitimiert wird, mit der Adaption derselben seitens eines Schulinspektors. In deren Mittelpunkt steht statt der Idee der Evidenzbasierung eine pädagogische Adressierung der schulischen Akteure durch die Inspektion.

Tobias Feldhoff, Luzia Durrer und Stephan G. Huber wenden die Thematisierung der eigensinnigen Adaption von Steuerungsimpulsen auf den unterschiedlichen Ebenen des Bildungssystems forschungsprogrammatisch. Am Beispiel einer Studie zum Reformprozess im Schweizer Kanton Uri stellen sie dar, wie neue Steuerungsmodelle in Schulsystemen anhand von Governance-Reglern analytisch und empirisch beschrieben werden können. In diesem Zusammenhang differenzieren sie auch zwischen der formalen Regelstruktur und der materialen Steuerungskonfiguration und zeigen so auf, dass eine Steuerungslogik sich nicht gänzlich in den jeweiligen formalen Regelungen abbildet, sondern genauso in den Wahrnehmungen der Beteiligten.

Nach dem Durchgang durch die Vielgestaltigkeit der „Aspekte ‚Neuer Steuerung‘“ wird mit dem Bericht von Uwe Maier und Harm Kuper nochmals der Bogen zum Ausgangspunkt, dem kritischen Blick auf das US-amerikanische Schulsystem, gespannt. In ihrem Überblicksartikel zum internationalen und deutschsprachigen Forschungsstand zu Effekten zentraler Tests konkretisieren sie das Forschungsdesiderat einer über die Ebene der Einzelschule hinausgehenden Untersuchung der Folgen einer testbasierten Schulreform.

Wenn – wie im Rahmen dieser Beiträge – Zugänge zu den deutenden Umgangsweisen mit verschiedenen Steuerungsimpulsen auf unterschiedlichen Ebenen des Schulsystems eröffnet werden, kann dieses dazu beitragen, die Reaktionen nicht allein als Störungen, sondern als eigensinnige und damit als immanent sinnvolle Bezugnahmen auf die Steuerungsversuche wahrzunehmen und anzuerkennen. Darin könnte sich eine qualitative Neuerung begründen, die es rechtfertigen würde, von ‚Neuer Steuerung‘ zu sprechen und nicht von vermeintlich oder tatsächlich ‚optimierter‘ alter Steuerung.

Martin Heinrich/Fabian Dietrich

Heinrich Mintrop/Gail Sunderman

Zentrale Steuerung von Schulentwicklung mit Quoten und Sanktionen: Das US-amerikanische Großexperiment in Schulen für benachteiligte Bildungsschichten

Zusammenfassung

Seit mehr als zwei Jahrzehnten erproben die Regierungen der Vereinigten Staaten von Amerika Systeme zur Sicherstellung und Kontrolle von Schulqualität. Um die Schulentwicklung zentralistisch zu steuern und voranzubringen, setzt man vor allem auf die Wirkung quantitativer Wachstumsziele, verbunden mit konsequenten Sanktionierungen. Die Design-Merkmale dieses Systems haben tatsächlich Auswirkungen auf Schulen und Verwaltungen gehabt. Allerdings haben sie nicht zu den beabsichtigten Resultaten geführt, sondern stattdessen unerwartete organisationale Dysfunktionalitäten erzeugt und zu einer unerwünschten Lernkultur geführt.

Schlüsselwörter: Schulqualität, Leistungsanreize, Lehrermotivation, Schulentwicklung, Rechenschaftssystem, Zentralisierung

Centralized Management of School Improvement with Quota and Sanctions: The US-American Experiment with High-Stakes Accountability in Schools Serving Disadvantaged Populations

Abstract

In the last two decades or so, governments in the United States have experimented with systems of school quality assurance and control that heavily rely on the power of quantitative growth targets combined with stringent sanctions in guiding and incentivizing school development efforts from the center. The design features of these systems have indeed triggered pro-active responses on the part of schools and local administrative units, as intended. Yet, they have been unable to achieve their intended goals. Instead, they produced unexpected organizational dysfunktionalities and resulted in an undesirable learning culture.

Key words: accountability systems, performance incentives, school quality, work motivation, school improvement, centralization

In vielen Ländern durchleben wir derzeit eine Phase durchgreifender Zentralisierung in der Bildungspolitik, in der höhere Ebenen der Bildungsadministration zunehmend die Initiative bei der Bewältigung von Problemen der Schulentwicklung übernehmen und das traditionelle Bildungsestablishment – Lehrerschaft, Standesvertretungen, Universitäten – zu Nebenakteuren in ihren ureigensten Aufgabenkreisen machen. In den USA hat diese Entwicklung schon in den achtziger Jahren begonnen und in den letzten zwanzig Jahren enorm an Intensität gewonnen. Den Höhepunkt dieser Zentralisierung markiert die Verabschiedung eines Bundesgesetzes im amerikanischen Kongress im Jahre 2001 mit dem vielversprechenden Namen „No Child Left Behind“ (NCLB). Vor zehn Jahren von einer großen Koalition von Demokraten und Republikanern, Wirtschaftsverbänden und Bürgerrechtsorganisationen mit hohen Erwartungen aus der Taufe gehoben (vgl. Debray 2006), sind die wesentlichen Instrumente des NCLB-Bildungsreformwerks offenbar ausgereizt. Inzwischen sind Hoffnung und Überschwang in Ablehnung umgeschlagen.¹

In diesem Beitrag behandeln wir das US-amerikanische System als eine spezielle Variante zentraler Steuerung von Schulentwicklung und erklären sein besonderes Design vor dem Hintergrund des nationalen Kontextes. Wir untersuchen seine Auswirkungen auf Verwaltungen und Schulen über einen Zeitraum von zehn Jahren und diskutieren im Anschluss daran, welche Lehren für Qualitätssicherungssysteme aus den Erfahrungen gezogen werden können.

1. Der nationale Kontext

Dass es in den USA überhaupt zu einem bedeutenden System der zentralen Steuerung von Schulqualität kommen konnte, ist keineswegs eine Selbstverständlichkeit. In einem Forschungsbeitrag im Jahre 1990, der verschiedene Versuche der Einflussnahme von Politik auf das Schulleben resümierte, rief die bekannte Schulforscherin Darling-Hammond (1990) die „Macht der Basis über die (hohe) Politik“ aus. Damit brachte sie den Konsens der amerikanischen Schulforschung zum Ausdruck, dass stetig immer neue Initiativen und Maßnahmen von Seiten der Politik am Klassenraum abprallten. Großangelegte Implementationsforschungen zu zentralen Initiativen zur Unterrichtsreform konnten nur geringe Umsetzungseffekte erkennen und kamen zu dem Schluss, dass die Politik nicht vorgeben könne, was für Schulen relevant sei (vgl. McLaughlin 1990). Die Begründung für diese Machtlosigkeit zentraler Steuerung lieferte die Organisationssoziologie mit ihrer Analyse einer Struktur der losen Kopplung, die es dem System ermögliche, die Praxis vor Ort, also den Unterricht,

¹ Vgl. URL: <http://www.nytimes.com/2011/10/12/education/>.

von der administrativen und formalen Struktur der Organisation zu lösen. Die politische Zersplitterung des Systems tue ihr Übriges. Da Bildungspolitik in den USA eine Angelegenheit der Bundesstaaten sei, aber anders als in Deutschland traditionell an lokale Schulbehörden mit ihren eigenen gewählten Schulausschüssen delegiert werde, ergebe sich ein Markt der Ideen, auf dem Reformen sich eher wie ständig wechselnde oberflächliche Moden verbreiteten (vgl. Wirt/Kirst 1997; Cuban 1990; Cohen/Spillane 1992). Je mehr sich das System auf der Oberfläche verändere, desto mehr bleibe es sich im Kerne treu.

Trotzdem waren Strukturelemente einer administrativen Kontrolle vorhanden (vgl. Weiler 1990), die der Bildungshistoriker Tyack (1974) in seiner einflussreichen Studie der „*Administrative Progressives*“ des späten neunzehnten und frühen zwanzigsten Jahrhunderts herausgearbeitet hat: die Messung von Schulleistung mit Hilfe von quantitativen standardisierten Tests, die Dominanz spezialisierter (im 19. Jahrhundert männlicher) Administratoren über mehrheitlich weibliche Lehrkräfte und die herausragende Rolle empirischer Befunde zur Begründung von Reformen, die von externen wissenschaftlichen Instanzen erhoben werden. Im Zusammenspiel dieser Elemente ist Schulreform im US-Schulsystem häufig mit Bemühungen verbunden, das *scientific management* des Systems im Hinblick auf eine höhere Effizienz zu verbessern. Doch der „Kult der Effizienz“ (Callahan 1962), die Vorstellung, dass die Feinstruktur des Unterrichts generell administrativ regulierbar sei oder sein solle, ist nicht auf Manager beschränkt, die auf Effizienz bedacht sind. Progressive Pädagogen und Pädagoginnen, Wissenschaftler und Wissenschaftlerinnen, die „*Best Practice*“-Unterrichtsprogramme entwickeln (vgl. z.B. Slavin 2005), oder auch politische Akteure, die für mehr Chancengleichheit eintreten (vgl. The Education Trust 2004), setzen auf sogenannte *instructional policy* und externes *instructional management* für die Lösung unterrichtlicher Probleme.

Administrative Regulierung sollte aber nicht mit zentraler Steuerung gleichgesetzt werden. Traditionell verfügen die Bildungsministerien in den USA auf der Ebene der Bundesstaaten nur über kleine administrative Apparate, die kaum Kapazitäten jenseits der Überprüfung von formalen Regeln, Gesetzen und lokalen Finanzbudgets zur Verfügung haben. Im Staat Kalifornien z.B. laufen die Apparate der kommunalen Schulbehörden in den großen Städten bei weitem dem Bildungsministerium den Rang ab (vgl. Timar 1997). So waren die ersten Ansätze verstärkter staatlicher Steuerung in den 80er-Jahren denn auch eher traditionell (vgl. Timar/Kirp 1988; Weiler 1990): staatliche Standards, neue Lehrpläne, Stärkung der Hauptfächer usw.

Die zweite Welle der Zentralisierung in den 90er-Jahren brachte etwas völlig Neues. Sie hatte drei Veränderungen, im Finanzwesen, der Arbeitsorganisation und der administrativen Kontrollkapazität, zur Voraussetzung:

- Fiskalisch ging in vielen Bundesstaaten, häufig als Folge von gerichtlich erzwungenen Angleichungen zwischen armen und reichen Kommunen, die Zuständigkeit für die Finanzierung des Schulwesens von den Kommunen auf die Bundesstaaten über. Die Folge war, dass zentrale Staatsregierungen neue fiskale Autorität erhielten (vgl. Verstegen/Knoepfel 1998).
- Beginnend in den 80er-Jahren wich die Arbeitsorganisation des „Organisationsmenschen“ (vgl. Whyte 1956) der des „Hochleistungsmenschen“ (vgl. Sennett 2006). Funktionierte der Organisationsmensch in stabilen öffentlichen und privatwirtschaftlichen Bürokratien, in denen Loyalität der Beschäftigten und das Prinzip der Seniorität vorherrschten, geht es in der heutigen Organisation von Arbeit um höhere Einsätze (*stakes*). Hochleistung wird belohnt, schwache Leistung führt zur Kündigung. Diese auf Hochleistung ausgerichtete Arbeitsorganisation ist zur vorherrschenden Leitfigur in der amerikanischen Privatwirtschaft geworden und wurde in den 90er-Jahren auf das öffentliche Schulsystem der USA übertragen.
- Eine dritte Vorbedingung sind neue Kontrollmechanismen vor allem für ehemals nur schwer kontrollierbare komplexe Arbeiten (vgl. Sennett 2006; Oldham/Hackman 2010). Mit Hilfe von neuen computergestützten Techniken der Datenerhebung und -speicherung ist es nun einer zentralen Planungsbehörde, sei sie öffentlich oder privat, möglich, klare Leistungszielvorgaben zu formulieren (*goal setting*) und mit relativ geringem Aufwand ihre Einhaltung zu überprüfen (*goal monitoring*). Die neuen Techniken erlauben es und machen es erforderlich, Ziele mit begrenzten quantitativen Leistungsindikatoren zu erfassen, zentrale Leistungserwartungen für eine große Anzahl von kleinen Leistungs- oder Arbeitseinheiten zu überprüfen und jeweils zielgenaue Sanktionen für leistungsschwache Einheiten anzuordnen. Was ehemals mittlere Vermittlungsebenen besorgten, liegt nun in der Hand des zentralen Leistungsmanagements. Im Bereich der Politik geben diese Verwaltungsstrukturen schmalen, zentral positionierten Eliten, egal ob eher effizienzorientiert oder sozial-reformerisch, neue machtvolle Steuerungsmöglichkeiten an die Hand.

Neue staatliche fiskalische Autorität, neue Arbeitsorganisation und neue Kontrollmöglichkeiten machten es nun möglich, Schulentwicklung über ein zentrales Rechenschaftssystem zu steuern.

2. Qualitätssicherung als Rechenschaftssystem für Leistung

Die gängigen Rechenschaftssysteme für Schulen in den USA bestehen von ihrer Konstruktion her aus: (a) kompetenz-, inhalts- und fertigungsorientierten Standards, (b) standardisierten Tests, die die Basis für Schul-Leistungsindikatoren sind, (c) quantitativen Zielvorgaben und Leistungsquoten (z.B. im Jahre X haben Y Prozent der Schüler und Schülerinnen Minimalstandards zu erfüllen), (d) Sanktionen, mit denen sich die höheren Politik- und Systemebenen bei den jeweils niedrigeren Gehör

verschaffen und (e) Unterstützungs- und Interventionssystemen, die neue Mittel und Ideen an die bedürftigsten Schulen verteilen sollen (vgl. Mintrop/Sunderman 2009; Fuhrman/O'Day 1996; O'Day 2002).

Das NCLB-Regelwerk fordert, dass Schulen in einem festgesetzten Zeitraum – es war der Zeitraum von 2001 bis 2014 geplant – alle Schüler und Schülerinnen auf das Leistungsniveau „*proficient*“ (fortgeschrittene Kenntnisse) zu bringen hatten. Aus der Differenz zwischen diesem Ziel und der tatsächlichen durchschnittlichen Leistung einer Schule, gemessen in jährlichen staatlichen Tests, ergibt sich die ‚angemessene jährliche Zuwachsrate‘ (*adequate yearly progress, AYP*). Für jede statistisch relevante demografische Untergruppe, differenziert nach Einkommen, Ethnie, Muttersprache und Förderungsbedarf, wird eine schulspezifische, jährlich zu erreichende Testquote kalkuliert. Dies soll die Schulen dazu zwingen, schwarze, hispanische, einkommensschwache und lernbehinderte Kinder genauso zu fördern wie weiße oder Kinder aus der Mittelschicht. Die Organisationseinheiten, die zur Rechenschaft gezogen werden, sind die Einzelschule und die kommunale Schulbehörde, nicht individuelle Lehrkräfte und auch nicht Eltern oder Schüler und Schülerinnen.

Die inhaltliche Substanz der Standards, das Anforderungsprofil der Tests, die Definition von „*proficient*“, die Bereitstellung von Inputs, die Einbindung von Schülern und Schülerinnen in das Sanktionssystem (d.h. wie Benotung und Versetzung mit den bundesstaatlichen Tests verknüpft werden sollten) und die spezielle Art und Weise der Unterstützung und Intervention in leistungsschwachen Schulen – alles das war der jeweiligen *einzelstaatlichen* Regulierung anheimgestellt.² Auf nationaler Ebene dagegen sind Sanktionsstufen vorgeschrieben, die eine Schule oder Schulbehörde durchlaufen muss, wenn sie die Zielvorgaben über einen Zeitraum von fünf Jahren nicht erfüllt. Auf der ersten Stufe steht die Veröffentlichung der schwachen Schulleistung, die die Schule oder Schulbehörde dem Druck ihres Marktes aussetzt. Auf der zweiten Stufe verliert die Schule oder Schulbehörde zunehmend ihre Autonomie und muss sich externen Interventionen und privaten Anbietern bestimmter Funktionen öffnen und dies auch bezahlen. Auf der dritten Stufe wird die Schule geschlossen, oder die kommunale Schulbehörde wird von der Staatsregierung übernommen.

Die Anhebung der Leistungsstandards eines großen nationalen Bildungssystems und insbesondere eine dramatische Verbesserung des Bildungsstandes von Gruppen mit niedrigem gesellschaftlichem Status sind enorm komplexe Angelegenheiten. Das gegenwärtig gültige amerikanische Rechenschaftssystem für Schulen hingegen besticht durch seine Simplizität. Im Kern besteht das System aus Quoten und Sanktionen. Und doch sind in Anbetracht der Bildungsmängel, mit denen zunehmend ausgegrenzte

2 Gemäß der föderalen Struktur in den USA sind die Staaten für die Bildung verantwortlich. Bundesgesetze, wie das NCLB-Gesetz, erreichen ihre Wirkung nur, wenn Staatsregierungen im Austausch gegen Bundesmittel in neue bundesweite Regeln und Kontrollen einwilligen.

Kinder aus den unteren Sozialschichten und den ethnischen Minoritäten zu kämpfen haben, Rechenschaftssysteme, die eine positive Bewegung in die Schulentwicklung bringen, grundsätzlich begrüßenswert. Entscheidend ist jedoch, dass sie, auch wenn sie nicht in allen Einzelheiten austariert sein sollten, den Bildungsauftrag der Schule im Wesen nicht verzerren. Von diesen Überlegungen lassen wir uns leiten, wenn wir im Folgenden die Auswirkungen des Systems auf die unteren Ebenen untersuchen.

3. Gemessene Leistungseffekte

Im Lichte wissenschaftlicher Studien ist es umstritten, ob die Rechenschaftssysteme die versprochene Verbesserung der messbaren Schülerleistung erbringen. Schon die Untersuchung der staatlichen Systeme der 90er-Jahre führte zu Kontroversen (vgl. Amrein/Berliner 2002; Carnoy/Loeb 2002; Haney 2000; Raymond/Hanushek 2003). Auch später ergibt sich kein klares Bild. Effektbefunde variieren von Studie zu Studie; und dort, wo Effekte gefunden werden, bleiben die kausalen Mechanismen zwischen Testergebnissen und Systemdesign im Dunkeln (vgl. Lee 2007; Nichols/Glass/Berliner 2006). Oftmals hängen gemessene Effekte von methodischen Entscheidungen der Forschenden ab (vgl. Braun 2004). Dabei geht es insbesondere um die Frage, wie mit Ausschluss von getesteten Schülern und Schülerinnen statistisch umgegangen werden soll. Eine Metaanalyse, gestützt auf eine Reihe großangelegter Studien mit nationalen Datensätzen, kommt zu dem Resultat, dass es einen mäßig positiven Effekt für die Accountability-Periode der Schulreform gegeben habe, aber keine signifikante Reduktion der Leistungsspanne zwischen oberen und unteren Statusgruppen (vgl. Lee 2008).

Gemessene Effekte und deren Interpretation hängen auch davon ab, welche Messeinheit zugrunde gelegt wird. Die meisten Rechenschaftssysteme der einzelnen Staaten erscheinen auf den ersten Blick erfolgreich zu sein. Studien, die die staatlichen Tests zur Grundlage nehmen, finden fast durchweg einen kontinuierlichen, oftmals steilen Anstieg der durchschnittlichen Testleistungen (vgl. Center for Education Policy 2008b; Neal/Schanzenbach 2008; The Education Trust 2004, 2005).

Ein bei weitem weniger positives Bild ergibt sich, wenn man die Ergebnisse des *National Assessment of Educational Progress* (NAEP) zugrunde legt, die in regelmäßigen Abständen als Zufallsstichprobe in allen Staaten erhoben werden. An das NAEP sind keine Sanktionen für Schulen oder Behörden geknüpft; NAEP ist ‚low stakes‘. Ein Vergleich von Testergebnissen der einzelnen Bundesstaaten und NAEP kann klären, wie realistisch die gemessenen Lernzuwächse wirklich sind. Um Verzerrungen (z.B. curriculare Verengungen, übertriebene Wiederholungen von Testitems) als Ursache für Testresultate auszuschließen, sollten *High-Stakes*- und *Low-Stakes*-Tests ein konsistentes Bild vermitteln (vgl. Koretz 2008; Lee 2007). Wenn man NAEP-Resultate zu-

grunde legt, fällt der Zuwachs in den Testergebnissen jedoch sehr viel geringer aus (vgl. Fuller u.a. 2006; Fuller u.a. 2007; Lee 2006, 2007). Generell sind die Ergebnisse für die Grundschulen besser als für weiterführende Schulen und in Mathematik etwas besser als im Bereich der Lesefähigkeiten. Entwickelte Steuersysteme, in denen Standards, Tests und Unterrichtsmaterialien aufeinander abgestimmt sind, scheinen positive Auswirkungen zu haben (vgl. Carnoy/Loeb 2002). Resümierend muss jedoch festgestellt werden, dass die großen Unterschiede zwischen den ethnischen und sozioökonomischen Statusgruppen weiter bestehen (vgl. Lee 2007, 2008). Zehn Jahre nach der Einführung des NCLB-Regelwerks ist klar geworden, dass das quantitative Wachstumsziel des Systems, dass die gesamte Schülerschaft das Leistungsniveau „*proficient*“ erreicht, in allen Bundesstaaten weit verfehlt worden ist.

4. Anreizfunktion und Arbeitsmotivation

In Anbetracht der weiten Diskrepanzen zwischen NAEP (Systemüberprüfung) und den jeweiligen bundesstaatlichen Tests (Rechenschaftslegung) ist es nicht geklärt, was die letzteren eigentlich messen. So viel scheint klar: Die staatlichen Rechenschaftssysteme, die mit Anreizen, Druck und Sanktionen operieren, schaffen es, die Aufmerksamkeit der Schulen auf die Tests zu richten und sie zu einer sorgfältigen Ausrichtung (*alignment*) auf die Standards und Testinhalte zu motivieren (vgl. Koretz 2008).

Die Übernahme der Verantwortung für die Leistungsentwicklung der Schülerinnen und Schüler wird von der Lehrerschaft als Idee im Allgemeinen akzeptiert, aber die konkreten Überprüfungssysteme werden eher skeptisch betrachtet (vgl. Conley/Goldman 2000; Darling-Hammond 2004; Firestone/Monfils/Schorr 2004; Hill/Lake/Celio 2002; Koretz u.a. 1996; Leithwood/Steinbach/Jantzi 2002; Loeb/Knapp/Elfers 2008; Mintrop/Trujillo 2007; Spillane u.a. 2002). Lehrkräfte halten eine begrenzte Orientierung des Unterrichts an den Tests (*teaching to the test*) für akzeptabel, lehnen aber den starken Druck und die damit verbundene Vernachlässigung von weniger leicht messbaren Lernzielen ab (vgl. Abrams/Pedulla/Madaus 2003; McNeil 2000; Valenzuela 2005).³ Jedoch greifen die Überprüfungssysteme insofern, als Schulen sie zur Orientierung für die Unterrichtsgestaltung heranziehen und energisch versuchen, den drohenden Sanktionen zu entkommen (vgl. Finnigan/Gross 2007; Mintrop 2004; Kelley/Protsik 1997).

Oftmals fühlen sich die engagiertesten Lehrkräfte im Kollegium, die am wenigsten Grund hätten, sich als Versager zu fühlen, am meisten dafür verantwortlich, die Leistungsbilanz der Schule zu verbessern. Doch wenn die erhofften Erfolge ausblei-

3 In der Tat konnte in einer Fallstudie kalifornischer Schulen ein enger Zusammenhang zwischen Testergebnissen, dem Leistungsrang einer Schule im System und der Qualität von Unterricht nicht nachgewiesen werden (vgl. Mintrop/Trujillo 2007).

ben oder sich nicht verstetigen, nachdem kurzfristige Lösungen erprobt worden sind, wie dies oft in leistungsschwachen Schulen mit geringer Entwicklungskapazität der Fall ist, dann breitet sich Demoralisierung aus (vgl. Finnigan/Gross 2007; Mintrop 2004). Einige Studien berichten von Schulen und Schulbehörden, die sich mit dem Rückenwind des Systems rapide aufwärts bewegen konnten (vgl. Skrla/Scheurich 2004; Hill/Lake/Celio 2002). Dennoch scheinen diese positiven Fälle nicht die Regel zu sein (vgl. Sunderman u.a. 2004; Abrams/Pedulla/Madaus 2003; Louis/Febey/Schroeder 2005). Vielmehr macht sich ein überwältigendes Klima der Negativität im unteren Drittel des Schulleistungsspektrums breit.

5. Praktikabilität und Lernkultur

Auf der Systemebene ist der Erfolg des Anreizmechanismus vom Zusammenspiel zwischen der Realitätsangemessenheit der Zielvorgaben, den vorhandenen Kapazitäten oder Kapazitätsreserven der unteren Ebenen, dem erzeugten Druck und der internen Wertschätzung der Leistungsforderungen oder Belohnungen abhängig (vgl. Mintrop 2004; Kelly/Protsik 1997). Das Funktionieren – und politische Überleben – der Rechenschaftssysteme in ihrer momentanen Konstruktion ist unmittelbar von der Höhe der kognitiven Anforderungen in den staatlichen Tests abhängig. Diese These kann mit Michael Fullans Konzept der zwei Zielhorizonte (vgl. Fullan 2003), die er aus der Evaluation der englischen Schulreform entwickelte, plausibilisiert werden. Innerhalb des ersten Horizontes konzentrieren sich die Rechenschaftssysteme auf Basiswissen und -fertigkeiten. Davon ausgehend erfolgen eine strikte Ausrichtung (*alignment*) der Curricula einer Schule an den staatlichen Tests, präskriptive Unterrichtsprogramme und das Festzurren von Organisationsabläufen. Außerdem kommen alle möglichen kurzfristigen Strategien, wie die besondere Aufmerksamkeit auf Schüler und Schülerinnen an den statistischen Schnittpunkten, Training von Testaufgaben, das enge Wiederholen von einzelnen Standards, die emotionale Mobilisierung der Schüler und Schülerinnen kurz vor dem Test usw., zur Anwendung. Diese scheinen, zumindest am Anfang, einen positiven Einfluss auf die Testresultate zu haben. Aber die Zuwächse scheinen nicht nachhaltig zu sein (vgl. Fullan 2003). Innerhalb eines erweiterten Zielhorizontes nehmen die Systeme kognitive Ziele jenseits von abrufbarem und wiederholbarem Wissen ins Visier, wie z.B. Begründen und komplexes Problemlösen.

Staatliche Rechenschaftssysteme, deren Tests auf relativ niedrigere kognitive Anforderungen abheben, wie z.B. im Staat Texas, der mehr als andere Staaten dem NCLB-Regelwerk Pate stand, haben innerhalb des gegenwärtigen NCLB-Regelwerks „besser“ funktioniert. Da sie auf existierende Kompetenzen von Lehrkräften und Schülern und Schülerinnen im unteren Bereich zugeschnitten sind, produzieren diese Systeme weniger Schulen, die als mangelhaft ausgewiesen sind, und brauchen gerin-

gere, weniger kostspielige Maßnahmen im Bereich der Schulentwicklung oder Lehrerfortbildung. Druck mag in diesem Fall als Steuerungsmechanismus hinreichend sein, um die Schulen auf die relativ moderaten staatlichen Ziele zu verpflichten. Im Kontrast dazu findet man staatliche Systeme, z.B. in Kalifornien, die auf höhere kognitive Leistungen abzielen und dabei völlig impraktikabel werden (vgl. Mintrop 2008; Mintrop/Trujillo 2005).

Wenn nun immer mehr Bundesstaaten es dem Staat Kalifornien gleichtäten und ihre Anforderungen für das Leistungsniveau „*Proficiency*“ den NAEP-Niveaus des Bundes angleichen, wie es auch von vielen Reformern und Reformerrinnen gefordert wird, dann hätte das den Nachweis einer enormen Anzahl von mangelhaften Schulen zur Folge, für die gewaltige Interventionskapazitäten bereitgestellt werden müssten. In der Tat zeigt Kalifornien, ein Staat mit hoher Einwanderungsquote und einer Mehrheit von nicht-weißen Schülern und Schülerinnen, wohin die Entwicklung laufen könnte. Im Jahre 2008 waren dort schon 48 Prozent aller Schulen und 61 Prozent aller kommunalen Schulbehörden von Bundessanktionen betroffen (vgl. Asimov 2008). Die Alternative zu diesem Szenario ist, die Rechenschaftssysteme im sicheren Hafen des unteren Zielhorizontes zu belassen, statt totale Dysfunktionalität zu riskieren. In der Tat haben Paterson und Hess (2008) eine Tendenz nach unten – „*race to the bottom*“ – diagnostiziert, der zufolge Staaten den starren Wachstumsanforderungen des Bundes zunehmend mit einfacheren Leistungskriterien für „*Proficiency*“ begegnen.

Ziele im Bereich des unteren Zielhorizontes sind eine lohnende Herausforderung für viele Schulen. Doch sind viele Schul- und Unterrichtsentwicklungsstrategien, die auf diesen Horizont ausgerichtet sind, wenig geeignet, Ziele im höheren Leistungsbereich zu erlangen. Kognitive Leistungsanforderungen im Basisbereich, gekoppelt mit hohem Druck, haben kommunale Schulbehörden und Schulleitungen, vor allem im eher leistungsschwachen Spektrum, dazu veranlasst, präskriptive, engmaschige Unterrichtsprogramme zu verordnen, die auf die staatlichen Tests zugeschnitten sind und die, tayloristisch, mit einer buchstäblichen Bindung an das Lehrermanual, einer so genannten „*implementation fidelity*“, zu unterrichten sind (vgl. Protheroe 2008). Das Schulwissen hält sich eng an die Testvorgaben, wird fragmentiert und stark lehrerzentriert vermittelt (vgl. Au 2007; Herman 2004). Das politische Ziel, allen Schülern und Schülerinnen die Möglichkeit zu geben, am „Lernen des 21. Jahrhunderts“ (vgl. <http://www.p21.org>; Zugriffsdatum: 13.02.2012) teilzunehmen, rückt so in weite Ferne.

6. Simplizität und Identifizierung von schwachen Schulen

Das NCLB-Regelwerk, wie übrigens auch die meisten Rechenschaftssysteme der Einzelstaaten, besticht durch seine Simplizität, die von der zentralen Administration wenig Datenkapazität verlangt, den lokalen Ebenen wenig Spielraum für Interpretationen und Zweideutigkeiten lässt und der Öffentlichkeit eine eingängige Metrik an die Hand gibt. Gemäß dem NCLB-Gesetz werden Entwicklungsfortschritte von Schulen in simplen Kompetenzstufen gemessen: „(far) below basic“, „basic“ und „proficient“. Bis dato sind aber Sanktionen nur von einem einzigen Kriterium abhängig, nämlich ob oder bis zu welchem Grade die Schulen einen jährlich zunehmenden Anteil von Schülern und Schülerinnen zur Kompetenzstufe „proficient“ geführt haben, und zwar geltend für alle ethnischen und sozio-ökonomischen Schüleruntergruppen. Dies hat zur Folge, dass Bewegungen im Bereich der Kompetenzstufe „basic“, vor allem nahe des „Proficiency“-Schnittpunkts, registriert werden, während das System positive Entwicklungen, z.B. im Bereich „far below basic“, wo viele leistungsschwache Schulen aber operieren, ignoriert (vgl. Neal/Schanzenbach 2008). Tatsächlich sind die durchschnittlichen Lernzuwächse in Schulen, die für Sanktionen infrage kommen, oftmals ähnlich denen von Schulen, die ihr AYP erreicht haben, wie eine Studie von Daten des Staates Virginia belegt (vgl. Kim/Sunderman 2005). Diese Ungenauigkeit wäre vielleicht tolerierbar, wenn AYP eines von mehreren Kriterien der Schulleistungsmessung wäre. Doch dies ist nicht der Fall. AYP als der einzige autoritative Indikator, der den Sanktionen zugrunde gelegt wird, schafft Realitäten und hat weitreichende Konsequenzen für Schulen und Schulbehörden.

In Anbetracht dessen, dass außerschulische Faktoren Schülerleistungen stark beeinflussen (vgl. Scheerens/Bosker 1997), können Schulen ihre Chancen im System beträchtlich steigern, indem sie die Zusammensetzung der Schülerschaft, die zum Test antritt, manipulieren. Eine Studie, basierend auf nationalen Daten (vgl. Balfanz/Legters 2008), hat Schulen untersucht, in denen sich die Jahrgangskohorte zwischen der 9. und 12. Klasse um mindestens 40 Prozent verringert. Nach den Kriterien des NCLB-Regelwerks wurden 41 Prozent dieser Schulen als akzeptabel eingestuft, da sie ihre AYP-Vorgaben erfüllten (vgl. Balfanz/Legters 2004, 2008), obwohl eine große Anzahl der Schüler und Schülerinnen die Schule ohne Abschluss verlässt. Ausschlussstrategien, wie z.B. zunehmendes Sitzenbleiben-Lassen vor wichtigen Jahrestests, Abschiebung von schwachen Schülern und Schülerinnen in nicht-rechenschaftspflichtige Sonderschul-Klassifikationen oder *Push-outs*, d.h. eine Tolerierung von oder Ermutigung zu Schulmüdigkeit bei schwachen Schülern und Schülerinnen (vgl. Haney 2008; Booher-Jennings 2005; Anagnostopoulos 2003), sind als Schnelllösungen in einem System naheliegend, welches Erfolg oder Misserfolg, Belohnung oder Strafe lediglich an dem Anteil der Schülerschaft bemisst, der es bis zur „Proficiency“-Kompetenzstufe geschafft hat.

7. Standardisierung und spezielle Lernbedürfnisse

Das Rechenschaftssystem verpflichtet alle Schulen, die Gelder vom Bund annehmen, Testergebnisse nach Schüleruntergruppen zu disaggregieren. Dies hat die Transparenz vor allem hinsichtlich der Leistungen von benachteiligten Schülern und Schülerinnen erhöht und Druck auf die Schulen erzeugt, Leistungsschwächen von Schülern und Schülerinnen, die ethnischen Minoritäten angehören, aus einkommensschwachen Familien stammen, Lernbehinderungen aufweisen oder eine nicht englische Herkunftssprache haben, ernst zu nehmen. Aber das Vorgehen hat auch neue, unerwartete Nebeneffekte erzeugt. Leistungsschwache Schulen haben häufig eine hohe Anzahl von Untergruppen, was zu vielen (maximal bis zu 37) verschiedenen Leistungsvorgaben (vgl. Marion u.a. 2002) führen kann. Je höher die demografische Diversität ist, desto höher ist die statistische Wahrscheinlichkeit, dass eine Vorgabe nicht erfüllt werden kann und Sanktionen folgen (vgl. Kim/Sunderman 2005; Linn 2005; Novak 2003).

Es hat sich herausgestellt, dass es den Schulen besondere Schwierigkeiten bereitet, Zielvorgaben für Kinder in zwei Kategorien zu erreichen, nämlich für solche mit nicht-englischer Herkunftssprache, „*limited English proficiency*“ (LEP), und solche mit Lernbehinderungen, „*learning disabilities*“ (LD). Die Folge ist, dass viele solide arbeitende Schulen unter das Sanktionsregime fallen, weil sie ihre LEP- oder LD-Untergruppen nicht nahe genug an die „*Proficiency*“-Marke herangebracht haben. Die Forschung über Rechenschaftssysteme im Bereich LEP ist relativ solide. Verschiedene Implementierungsprobleme haben sich herausgeschält, vor allem der ständige Austausch von leistungsstärkeren Abgängen mit Neuzugängen mit großen sprachlichen Defiziten und die mangelnde Diagnostik der standardisierten staatlichen Tests für Zweitsprachler und -sprachlerinnen (vgl. Abedi 2004; Batt/Kim/Sunderman 2005; Coltrane 2002; Kieffer/Lesaux/Snow 2008). Ein ähnliches Muster ergibt sich für die „*Learning-Disabilities*“-Untergruppe (vgl. Batt/Kim/Sunderman 2005; Sunderman/Kim/Orfield 2005).

Es sollte nicht überraschen, dass ein System, welches Schulleistung mit hoch standardisierten und simplen Einheiten bemisst und wenig Abweichungen von einheitlichen Zielvorgaben erlaubt, bei der Behandlung von Schülern und Schülerinnen mit besonderem Lernbedarf unter Stress gerät. Aber die Lösung dieses Problems ist nicht einfach. In der Logik von Druck und Sanktionen werden Gruppen, die von der Rechenschaftslegung ausgenommen werden, etwa weil sie nur schwerlich standardisiert werden können, leicht ignoriert. Es ist nun einmal im Selbstinteresse der Schulen und der Schulbehörden, sich in ihrer Arbeit an den Forderungen mit dem größten Rechenschaftsdruck auszurichten.

8. Radikale Interventionen und Schulentwicklung

Staatsregierungen und lokale Schulbehörden vermeiden die radikalen Maßnahmen zur Umstrukturierung, die von der Bundesregierung im NCLB-Gesetz vorgeschlagen werden, wie z.B. die Schließung und Neueröffnung einer Schule mit neuem Kollegium (*reconstitution*) oder die Umwandlung einer öffentlichen in eine privat geführte Schule (vgl. Center for Education Policy 2007, 2008a; GAO 2007), zugunsten von weicheren, eher traditionellen Schulentwicklungsmaßnahmen (vgl. Center for Education Policy 2007). In der Tat haben Studien gezeigt, dass die radikaleren Methoden nicht immer die besseren sind. Bei der Rekonstituierung einer Schule mit neuem Kollegium, dies dokumentieren Studien aus Maryland (vgl. Malen u.a. 2002), Chicago (vgl. Hess 2003) und New York (vgl. Brady 2003; NYSED 2003), profitieren in einigen Fällen die Schulen vom Neuanfang, aber in vielen anderen Fällen werden die Schulen nicht besser; einige verschlechtern sich aufgrund von zusätzlichem Chaos. Die Übernahme von Schulen in private Trägerschaft (*educational management organizations*), die häufig eine größere Anzahl von leistungsschwachen Schulen managen, hat manchmal positive, manchmal negative Auswirkungen, wie Studien aus Baltimore und Philadelphia zeigen (vgl. Blanc 2003; Bracey 2002; Molnar 2005; Travers 2003a; Useem 2005; Saltman 2005). Trotz zunehmender Forschungen zu den so genannten *Charter Schools* – Schulen mit halböffentlichem Status, die oftmals von Eltern, Lehrkräften, Nachbarschaftsorganisationen, aber auch von den großen Managementorganisationen betrieben werden – gibt es wenig Anhaltspunkte, dass die Umwandlung von öffentlichen Schulen in halböffentliche *Charter Schools* eine hilfreiche Strategie ist (vgl. The Brown Center 2003; Gill u.a. 2007). Es ist zusammenfassend festzustellen, dass keine der Umstrukturierungsstrategien, mit denen bisher in den USA experimentiert wurde, sich als generell wirksam für Schulentwicklung erwiesen hat oder robust genug ist, sich gegen lokale Bedingungen durchzusetzen (vgl. auch Mathis 2009).

Die starre Abstufung von Sanktionsstufen hat für immer mehr Schulen unproduktive Turbulenzen produziert, anstatt eine sich selbst verstärkende, geordnete Entwicklungsdynamik hervorzurufen. Schulen, die über Jahre leistungsschwach gewesen sind, werden nun nicht mit den Sanktionen sozusagen „aus dem Schulentwicklungs-Dornröschenschlaf geweckt“. Ganz im Gegenteil, diese Schulen sind in ihrer Mehrzahl nicht stabil agierende Organisationen, sondern befinden sich in ständigem ungeplantem Auf- oder Abbruch. Lehrkräfte und Schulleitungen kommen und gehen; externe Berater und Beraterinnen haben sich an ihnen zu Hauf versucht, und den Bundessanktionen sind oftmals bereits intensive lokale Interventionen vorausgegangen (vgl. Mintrop 2004; Neild/Spiridakis 2002; Neild u.a. 2003). Die Stadt Philadelphia liefert ein gutes Beispiel. Hier wurden alle leistungsschwachen Schulen schon vor dem Inkrafttreten der Bundessanktionen radikal umgestaltet (vgl. Travers 2003b), ohne dass damit das Problem mangelnder Leistung beseitigt worden wäre.

Diese Schulen durchlaufen die NCLB-Sanktionsstufen, haben aber bereits schon die schärfsten Sanktionsstufen, wie radikale Neugründung oder neue Trägerschaft, vorweggenommen. Dazu kommt die Instabilität von kommunalen Schulbehörden, die bei dem ständigen Druck wegen krisenhafter Leistungen mit vermehrtem Abgang von Verwaltungsleuten und einem häufigen Wechsel an der Spitze zu kämpfen haben (vgl. Buchanan 2006).

Schulen in den städtischen Zentren, wo die Armen und die Minoritäten in den USA oftmals konzentriert leben, beschäftigen in zunehmendem Maße Lehrkräfte und Schulleitungen mit einer Fluktuationsrate von bis zu zwei Dritteln über einen Zeitraum von fünf Jahren, da die Beschäftigten nur schwer über einen längeren Zeitraum gehalten werden können (vgl. Johnson 2007). Bei diesen Lehrerkohorten geht es also weniger darum, wie sie zur Rechenschaft gezogen werden, als wie sie zum Bleiben bewegt werden können. Keine der Sanktionen oder Interventionen, weder auf der nationalen Ebene noch auf der bundesstaatlichen Ebene, scheint auf dieses Problem zu reagieren. Im Kontext der Hochleistungs-Arbeitsorganisation ist hingegen das regelmäßige „*Hiring and Firing*“ eine integrale und gewollte Dynamik.

9. Reform-Infrastruktur und Unterstützung

Forschungen haben gezeigt, dass es Schulen und Behörden gibt, die durchaus die notwendigen Kapazitäten haben, ihr Leistungsdefizit zu kompensieren. Schulen mit genügender Kapazität können so auf Druck und Orientierung des Rechenschaftssystems positiv reagieren (vgl. Diamond/Spillane 2004; Elmore 2004; O’Day 2002; Sunderman 2001). Jedoch mangelt es in vielen leistungsschwachen Schulen an professionellem Wissen, Geldern und Selbstorganisation. Diese Schulen sind nur schwer in der Lage, die Signale des Systems positiv für die Schulentwicklung umzusetzen (vgl. Elmore 2004). Ständig verstärkter Druck führt hier zu mehr Rigidität, Fragmentierung und Verschlechterung der kollegialen Beziehungen, was gemeinsames Lernen und Erproben von neuen Ansätzen stark behindert (vgl. O’Day 2002; Mintrop 2004). Schulen mit mangelnden Kapazitäten sind dazu prädestiniert, mit kurzfristigen Strategien zu operieren, die kaum neue Kapazitäten erfordern, wie z.B. eine enge Ausrichtung auf die Tests (vgl. Sunderman u.a. 2004). Im Extremfall kann dies zu einem dezidiert anderen Curriculum führen, in welchem ein testzentrierter Förderunterricht das reguläre Curriculum an den Rand schiebt (vgl. Booher-Jennings 2005; Diamond/Spillane 2004; Sunderman 2001; Valenzuela 2005).

Im Regelwerk der amerikanischen Bundesregierung spielen die bundesstaatlichen Bildungsministerien bei der Implementierung von NCLB eine zentrale Rolle, doch sind wenig neue Gelder für die Bewältigung der erweiterten Aufgaben dieser Ministerien bewilligt worden. Die „*State Departments of Education*“, traditionell rela-

tiv schwach besetzte Organisationen, haben weder Personal und Expertise noch etablierte Kommunikationskanäle, die für die Schulentwicklung benötigt werden (vgl. McDermott 2004; Sunderman/Orfield 2006), sodass sie in der Regel diese Aufgaben nach unten delegieren – an Behörden und Schulen also, die oftmals gerade deshalb den Sanktionen ausgesetzt sind, weil sie sich in der Vergangenheit aus eigener Kraft nicht verbessern konnten.

10. Zusammenfassung und Diskussion

Im Kern geht es bei Qualitätssicherung um guten Unterricht. Guter Unterricht basiert auf einer komplexen Interaktion von Wissen, Zielen, Diagnostik, Motivation, Anstrengung, Belohnung und Entscheidungsfindung, wobei Lehrkräfte lokale Einflüsse und Präferenzen mit allgemeineren professionellen Wissensbeständen und Normen vermitteln müssen. Gleichzeitig müssen lokale Kriterien übergeordneten gesellschaftlichen Leistungs-, Wissens- und Verhaltensstandards standhalten (vgl. auch O'Day 2002; Weiler 1990). Idealtypisch könnten Systeme der Qualitätssicherung lokal, professionell oder administrativ organisiert sein und zwischen diesen Ebenen vermitteln. Eine lokale Dynamik kommt zum Tragen, wenn lokale Eliten oder Eltern gesellschaftlich erforderliches Wissen überblicken können und selbst, oft in direktem persönlichen Kontakt, die Aufsicht über eine Lehrerschaft übernehmen, die im Konsens mit der lokalen Gemeinde agiert und eine eher geringe Kontrolle über ihre Arbeit hat. Professionelle Dynamiken der Rechenschaftslegung appellieren an etabliertes Wissen und verinnerlichte Erziehungs- und Leistungsnormen der Lehrerschaft, die in der Lehrerausbildung vermittelt werden und durch internen kollegialen Austausch in den Schulen oder mit Hilfe externer Experten und Expertinnen wach gehalten und verstärkt werden. Administrative Mechanismen beruhen auf externen Leistungskriterien, Regeln, Sanktionen und Ressourcen für Lehrende und Lernende, die in wechselseitiger Interaktion zwischen der Zentrale und den lokalen, nachgeordneten Ebenen kommuniziert und hierarchisch kontrolliert werden.

Rechenschaftssysteme in modernen Staaten sind immer Mischformen, bei denen lokale, professionelle und administrative Dynamiken unterschiedliches Gewicht haben können. Eine Variante ist denkbar, in der der Staat allgemeine organisatorische und regulative Rahmenrichtlinien setzt und durch Abschlussprüfungen für Schüler und Schülerinnen das Pflichtgefühl der professionellen Akteure in den Schulen weckt, alles zu tun, um möglichst vielen Schülern und Schülerinnen zum Schulerfolg zu verhelfen. Ein solches System hätte Vor- und Nachteile. Auf der einen Seite beschränkt es den notwendigen Informationsfluss zwischen den Ebenen auf passende Prüfungsaufgaben und Beurteilungskriterien und belässt die Komplexität von pädagogischen Entscheidungen in den Händen derjenigen, die über lokales und professionelles Wissen verfügen. Auf der anderen Seite lässt sich eine verwurzelte

Leistungskultur, die sich im Zusammenspiel von Schule und Familie etabliert hat, vermutlich nur schwer in diesem System dynamisieren, und mangelnde Schulleistungen, z.B. von benachteiligten sozialen Statusgruppen, bleiben tolerierbar, solange sie in konsensuale gesellschaftliche oder professionelle Erwartungen eingebettet sind.

In weiten Teilen der USA war eine Mischung aus starken lokalen Kontrollmechanismen, geringer zentraler Kontrolle der Rahmenbedingungen und schwacher Profession bis in die 60er-Jahre hinein vorherrschend (vgl. Wirt/Kirst 1997). Mit schwindender Bedeutung der kommunalen Schulbehörden und einer anhaltenden Schwäche der Profession hat sich die administrative Zentrale, sei es auf Staats- oder auf Bundesebene, in den Vordergrund geschoben. Sie macht die ungenügende Anstrengung der Profession, die dysfunktionale Lokalpolitik in den Kommunen und die niedrige Leistungserwartung für benachteiligte Schüler und Schülerinnen⁴ für Leistungsdefizite verantwortlich. Damit wird die zentrale Steuerung mit administrativen Mitteln zum Primat, und ihre Durchschlagskraft muss die mangelnde Einbeziehung der lokalen und professionellen Kräfte kompensieren.

Es gibt Indizien in der Forschung, dass die Durchschlagskraft dieses Systems nicht unbedeutend ist. Das Schulsystem ist gewissermaßen durch eine Revolution von oben in Bewegung geraten. Die Veränderungen in den staatlichen Tests zeigen an, dass das Rechenschaftssystem ernst genommen wird und die Lehrerarbeit beeinflusst. Trotz eines verbreiteten Skeptizismus machen Schulen große Anstrengungen, ihre Wachstumsziele zu erreichen. In den positiven Fallstudien wird deutlich, dass die Leistungsdefizite vormals vernachlässigter Schülergruppen sichtbar werden und ernst genommen werden. Schulen und Schulverwaltungen suchen nach Unterrichtsprogrammen, die empirisch evaluiert sind und die Testresultate verbessern können. Ein neues Management-Modell – Schulen in flexibler halb privater Trägerschaft – wird erprobt. Das Schulsystem ist also in Bewegung geraten.

Aber wie wir dokumentiert haben, hat diese Bewegung auch gravierende nachteilige Nebeneffekte. Da die verschiedenen Steuerungsinstrumente tief in die Komplexität pädagogischen Handelns vor Ort eingreifen, haben Konstruktionsfehler des Systems unabsehbare und weitreichende Folgen. Wenn die Wachstumsziele falsch gewählt werden, hat das Konsequenzen für die Anreizfunktion. Wenn Ziele gemessen an den vorhandenen Kapazitäten zu hoch angesetzt werden, stumpft die Motivation nach einer Welle der Erregung ab oder schlägt in Demoralisierung um. Dieses Szenarium scheint sich in vielen Schulen im unteren Leistungsspektrum gegenwärtig abzuspielen. Mit immer höheren Zahlen von scheiternden Schulen, gekoppelt an die zentrale Forderung nach zunehmend radikalen Lösungen, kann von einer rationalen Allokation von Ressourcen und Reformenergien für die am meisten bedürftigen

4 „The soft bigotry of low expectations“; vgl. URL: <http://www.nrrf.org/paige3-12-03.htm>; Zugriffdatum: 13.02.2012.

Schulen kaum noch die Rede sein. Mit explodierenden Zahlen von korrekturbedürftigen Schulen wird das System dysfunktional.

Gegenwärtig kann das System mit nur sehr wenigen Indikatoren operieren (vor allem mit demografischen Daten, Anwesenheit, Testdaten und Lehrertifikaten), für die die zentralen Informationssysteme Statistiken bereitstellen. Diese Indikatoren erlauben es nicht, Manipulationen der Schulen, vor allem die Exklusion „auf kaltem Wege“, zu erkennen und entsprechend zu sanktionieren. Es ist völlig unklar, was mit schwer standardisierbaren Schülergruppen geschehen soll. Werden sie vom Sanktionsregime ausgeschlossen, weil sie nicht in die simple Metrik hineinpassen, so werden Schulen, die unter Druck stehen, sie beiseiteschieben. Werden diese Schülergruppen einbezogen, hat das negative Folgen für Anreizfunktion und Motivation. Hoher Druck wirkt hier gänzlich kontraproduktiv.

Doch das entscheidende Problem des Systems ist, dass es eine Lernkultur hervorbringt, die mit gutem Unterricht wenig zu tun hat. Zumindest im unteren Leistungsspektrum des Schulsystems vollziehen sich offensichtlich eine strikte Ausrichtung an den bundesstaatlichen standardisierten *High-Stakes*-Tests und ein Wettlauf in den kognitiven Basisbereichen; beides zahlt sich nicht in entsprechenden Steigerungen in *Low-Stakes*-Tests (NAEP) aus. Der ‚Goldstandard‘ der Systemleistung ist aber der *Low-Stakes*-Test, und dieser zeigt an, dass das System sein wichtigstes Ziel, nämlich das Schließen der Leistungslücke zwischen benachteiligten und privilegierten Schülern und Schülerinnen, das „*closing of the achievement gap*“, verfehlt hat. Das US-amerikanische System der zentralen Steuerung mit administrativen Mitteln hat sich offensichtlich übernommen.

Wie könnte es weitergehen? Eine Variante wäre, das administrative Steuerungssystem besser mit einer professionellen Dynamik von Schulreform zu verknüpfen (vgl. O’Day 2002) oder lokale Stimmen stärker einzubeziehen. Das Problem ist, dass dann Interessengruppen Einfluss gewinnen, die im momentanen Diskurs als Problemträger identifiziert sind. Die Obama-Administration hat aus den Dysfunktionalitäten des NCLB-Regelwerks andere Konsequenzen gezogen. Die Funktionslogik des Primats zentraler administrativer Steuerung von Schule und Unterricht wird beibehalten; aber in den letzten zwei Jahren sind Novellierungen in allen wichtigen Funktionsbereichen geplant oder schon unternommen worden,⁵ die den Schwerpunkt des Systems verlagern. Mit der Aufgabe fixer quantitativer Zielsetzungen wird das System flexibilisiert; mit der Neuorientierung auf komplexere Standards im Bereich „*college readiness*“ wird der kognitive Horizont des Systems erweitert.⁶ Entscheidend ist, dass nun

5 Quellen: URL: <http://www.ed.gov/esea/flexibility>; URL: <http://www2.ed.gov/programs/racetothetop/>; URL: <http://www.corestandards.org/>; URL: <http://www2.ed.gov/policy/elsec/leg/blueprint/>; URL: <http://www.districtadministration.com/news/harkins>; Zugriffsdatum: 13.02.2012.

6 46 Bundesstaaten haben bis dato die neuen Standards akzeptiert. Vgl. URL: <http://www.corestandards.org/>; Zugriffsdatum: 13.02.2012.

die Hauptlast der Rechenschaft von der Organisation (Behörde, Schule) auf die einzelnen Lehrkräfte verlagert wird. Diese, so fordert die Bundesregierung von den Staaten, sollen jetzt regelmäßig mit Hilfe von standardisierten Testresultaten und standardisierten Lehrproben evaluiert werden. Evaluationen müssen in die Laufbahnentscheidungen der Behörden mit eingehen. Überdies werden Gelder für Interventionen auf die unteren fünf Prozent der Schulen beschränkt. Allerdings sind die kommunalen Schulbehörden verpflichtet, bei der Annahme von Bundesgeldern die Schulleitungen oder mindestens die Hälfte eines jeweiligen Kollegiums auszutauschen oder die Schulen in halb öffentliche *Charter Schools* umzuwandeln.

Dies ist der Plan der Regierung. Aber es ist nicht klar, ob es zu diesen Veränderungen kommen wird, da der Kongress, in dem die Skepsis an der Mikro-Steuerung des Schulwesens durch den Bund stark zugenommen hat, eine sehr viel föderalere Version der NCLB-Novellierung verabschieden könnte. Es ist klar, die Hochphase der zentralen Steuerung ist beendet: Ein Schließen der Leistungslücke wird es mit Leistungsmessung, Wachstumsquoten, Sanktionen und zentral verordneten Lösungen in den USA nicht geben. Aber die Phase einer zentralen Kontrolle der professionellen Arbeit hat vermutlich gerade erst begonnen. Und eingedenk der strukturellen und politischen Voraussetzungen wird das Land seine Vorreiterrolle in dieser Dimension so schnell nicht verlieren.

Was kann ein deutsches Lesepublikum aus den amerikanischen Erfahrungen lernen? Die USA beenden gerade eine Etappe in einem zwanzigjährigen Großversuch, der mit vier Grundannahmen begann, die wir provisorisch *Willens-Hypothese*, *Vereinfachungs-Hypothese*, *Dynamisierungs-Hypothese* und *Systemische-Reform-Hypothese* nennen wollen.

Die *Willens-Hypothese* besagt, dass der mangelnde Wille der Profession für die Leistungsdefizite des Systems verantwortlich ist und dass Anreize das zentrale Mittel sind, um das System zu verbessern. Wir wissen nun, dass diese Hypothese so nicht stimmt. Das System hat es zwar geschafft, Schulen mit Entschlossenheit auf die staatlichen Tests auszurichten und die Testresultate zu verbessern, aber vergleichbar substantielle Lernleistungen von Schülern und Schülerinnen, die sich in *Low-Stakes*-Tests zeigen müssten, scheint dieses Anreizsystem nicht hervorzubringen.

Die *Vereinfachungs-Hypothese* wurde von Management- und Psychometrie-Experten und -Expertinnen formuliert. Sie besagt, dass wir durchaus ein komplexes System menschlichen Miteinanders, nämlich den Unterricht, mit Hilfe einer begrenzten Anzahl wirkmächtiger quantitativer Indikatoren humanisieren können. Das amerikanische Großexperiment hat uns gelehrt, wie schwer es ist, die exakten numerischen Größen mit den diffuseren professionellen Werten menschlicher Fürsorge, leidenschaftlicher Wissensvermittlung und gutem Unterricht in Verbindung zu bringen. Wir haben Fallstudien erwähnt, in denen das Rechenschaftssystem machtvoll auf die-

se nicht quantitativ messbaren Werte professionellen Arbeitens ausstrahlt, aber die Standardreaktion, so scheint es, besteht darin, den kürzesten Weg zwischen geforderten Messeinheiten und antrainiertem Lehrer- und Schülerverhalten zu wählen; damit ist der Wettlauf nach unten eröffnet.

Die *Dynamisierungs-Hypothese* besagt, dass eine Revolution von oben benötigt wird, um die Starrheit der mittleren und unteren Ebenen des Systems zu verflüssigen. Das amerikanische Großexperiment hat gezeigt, dass diese Hypothese insofern stimmt, als Initiativen der Zentrale in der Tat untere Ebenen in große Turbulenzen stürzen können. Aber die Treffsicherheit dieses Impulses ist wenig überzeugend, wenn die Macht von Sanktionen auf eine wenig entwickelte Reforminfrastruktur trifft.

Die letzte Hypothese, die hier angesprochen werden soll, ist die *Systemische-Reform-Hypothese*. Sie postuliert, dass Systemelemente, wie Lehrpläne, Tests, Unterrichtsmaterialien und Lehrerfortbildungen aufeinander abgestimmt sein sollten, um die Entwicklungsanstrengungen der Lehrenden in einer Struktur von Vorhersehbarkeit, Datenevidenz und externer Unterstützung abzusichern. Wir haben Anhaltspunkte dafür, dass sich diese Hypothese, nach Lage der Daten in den USA, als haltbar erweisen könnte.

Literatur

- Abedi, J. (2004): The No Child Left Behind Act and English Language Learners: Assessment and Accountability Issues. In: Educational Researcher 33, H. 1, S. 4-14.
- Abrams, L./Pedulla, J./Madaus, G. (2003): Views from the Classroom: Teachers' Opinions of Statewide Testing Programs. In: Theory into Practice 42, H. 1.
- Amrein, A.L./Berliner, D.C. (2002): High-Stakes Testing, Uncertainty, and Student Learning. In: Education Policy Analysis Archives 10, H. 18. URL: <http://epaa.asu.edu/epaa/v10n18/>; Zugriffsdatum: 22.08.2003.
- Anagnostopoulos, D. (2003): The New Accountability, Student Failure, and Teachers' Work in Urban High Schools. In: Educational Policy 17, H. 3, S. 291-316.
- Asimov, N. (2008, September 5): State Falling Way behind No Child Left Behind. San Francisco Chronicle, S. A-1.
- Au, W. (2007): High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. In: Educational Researcher 36, H. 5, S. 258-267.
- Balfanz, R./Legters, N.E. (2004): Locating the Dropout Crisis: Which High Schools Produce the Nation's Dropouts? In: Orfield, G. (Hrsg.): Dropouts in America: Confronting the Graduation Rate Crisis. Cambridge, MA: Harvard Education Press, S. 57-84.
- Balfanz, R./Legters, N.E. (2008): NCLB and Reforming the Nation's Lowest-Performing High Schools. In: Sunderman, G.L. (Hrsg.): Holding NCLB Accountable: Achieving Accountability, Equity, and School Reform. Thousand Oaks, CA: Corwin Press, S. 191-207.
- Batt, L./Kim, J./Sunderman, G.L. (2005): Limited English Proficiency Students: Increased Accountability under NCLB. Cambridge, MA: The Civil Rights Project at Harvard University.

- Blanc, S. (2003, Summer): Supports and Challenges in the Multiple Provider Model: Principals Offer Mixed Perspectives on First Year of Takeover. Philadelphia Public School Notebook.
- Booher-Jennings, J. (2005): Below the Bubble: "Educational Triage" and the Texas Accountability System. In: *American Educational Research Journal* 42, H. 2, S. 231-268.
- Bracey, G. (2002): The Market in Theory Meets the Market in Practice: The Case of Edison Schools. Education Policy Studies Laboratory: Education Policy Research Unit (EPRU), College of Education, Arizona State University. URL: <http://www.asu.edu/educ/eps/EPRL/documents/EPRL%202002-107/EPRL-0202-107-EPRU.htm>; Zugriffsdatum: 02.11.2005.
- Brady, R. (2003): *Can Failing Schools be Fixed?* Washington, DC: Thomas B. Fordham Foundation.
- Braun, H. (2004): Reconsidering the Impact of High-Stakes Testing. In: *Education Policy Analysis Archives* 12, H. 1. URL: <http://epaa.asu.edu/epaa/v12n1/>; Zugriffsdatum: 13.02.2012.
- Buchanan, B. (2006): *Turnover at the Top: Superintendent Vacancies and the Urban School.* Lanham, PA: Rowman & Littlefield.
- Callahan, R. (1962): *Education and the Cult of Efficiency.* Chicago: University of Chicago Press.
- Carnoy, M./Loeb, S. (2002): Does External Accountability Affect Student Outcomes? In: *Education Evaluation and Policy Analysis* 24, S. 305-332.
- Center for Education Policy (2007): *Moving beyond Identification: Assisting Schools in Improvement.* Washington, DC: Author.
- Center for Education Policy (2008a): *Has Student Achievement Increased since 2002? State Test Score Trends through 2006-07.* Washington, DC: Author.
- Center for Education Policy (2008b): *Managing More than a Thousand Remodeling Projects: School Restructuring in California.* Washington, DC: Author.
- Cohen, D.K./Spillane J. (1992): Policy and Practice: The Relations Between Governance and Instruction. In: *Review of Research in Education* 18, S. 3-49.
- Coltrane, B. (2002, November): English Language Learners and High-Stakes Tests: An Overview of the Issues. URL: <http://www.cal.org/ericcl/DIGEST>; Zugriffsdatum: 13.05.2004.
- Conley, D./Goldman, P. (2000): Half Full, Half Empty? Educator Responses over Time to State-level Systematic Reforms. In: *International Journal of Educational Reform* 9, H. 3, S. 249-260.
- Cuban, L. (1990): Reforming Again, Again, and Again. In: *Educational Researcher* 19, H. 1, S. 3-13.
- Darling-Hammond, L. (1990): Instructional Policy and Practice: The Power of the Bottom over the Top. In: *Educational Evaluation and Policy Analysis* 12, H. 3, S. 339-347.
- Darling-Hammond, L. (2004): Standards, Accountability, and School Reform. In: *Teachers College Record* 106, H. 6, S. 1047-1085.
- Debray, E. (2006): *Politics, Ideology, and Education: Federal Policy during the Clinton and Bush Administrations.* New York: Teachers College Press.
- Diamond, J.B./Spillane, J. (2004): High-Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality? In: *Teachers College Record* 106, S. 1145-1176.
- Elmore, R.F. (2004): *School Reform from the Inside Out: Policy, Practice and Performance.* Cambridge, MA: Harvard Education Press.

- Finnigan, K./Gross, B. (2007): Do Accountability Policy Sanctions Influence Teacher Motivation? Lessons from Chicago's Low-Performing School. In: *American Educational Research Journal* 44, H. 3, S. 594-629.
- Firestone, W./Monfils, L./Schorr, R. (2004): *The Ambiguity of Teaching to the Test: Standards, Assessment, and Educational Reform*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
- Fuhrman, S./O'Day, J. (Hrsg.) (1996): *Rewards and Reform*. San Francisco: Jossey-Bass.
- Fullan, M. (2003): *Change Forces with a Vengeance*. London/New York: Routledge Farmer.
- Fuller, B./Gesicki, K./Kang, E./Wright, J. (2006): Is the No Child Left Behind Act Working? The Reliability of how States Track Achievement. Working Paper 06-1 Policy Analysis for California Education, PACE. 3653 Tolman Hall, Berkeley, CA 94720-1670. Tel: 510-642-7223; Fax: 510-642-9148; E-Mail: pace@berkeley.edu. URL: http://pace.berkeley.edu/pace_publications.html.
- Fuller, B./Wright, J./Gesicki, K./Kang, E. (2007): Gauging Growth: How to Judge No Child Left Behind? In: *Educational Researcher* 36, H. 5, S. 268-278.
- GAO (U.S. Government Accountability Office) (2007): *No Child Left Behind Act: Education Should Clarify Guidance and Address Potential Compliance Issues for Schools in Corrective Action and Restructuring Status*. Washington, DC: Author.
- Gill, B./Zimmer, R./Christman, J./Blanc, S. (2007): *State Takeover, School Restructuring, Private Management, and Student Achievement in Philadelphia*. Santa Monica: Rand.
- Haney, W.M. (2000): The Myth of the Texas Miracle in Education. In: *Education Policy Analysis Archives* 8, H. 41. URL: <http://epaa.asu.edu/epaa/v8n41/>; Zugriffsdatum: 22.08.2006.
- Haney, W.M. (2008): Evidence on Education under NCLB (and How Florida Boosted NAEP Scores and Reduced the Race Gap). In: Sunderman, G.L. (Hrsg.): *Holding NCLB Accountable: Achieving Accountability, Equity, and School Reform*. Thousand Oaks, CA: Corwin Press, S. 91-102.
- Herman, J.L. (2004): The Effects of Testing on Instruction. In: Fuhrman, S./Elmore, R. (Hrsg.): *Redesigning Accountability*. New York: Teachers College Press, S. 141-146.
- Hess, G.A. (2003): Reconstitution – Three Years Later: Monitoring the Effect of Sanctions on Chicago High Schools. In: *Education and Urban Society* 35, H. 3, S. 300-327.
- Hill, P./Lake, R./Celio, M. (2002): *Charter Schools and Accountability in Public Education*. Washington, DC: Brookings Institution Press.
- Johnson, S. (2007): *Finders and Keepers: Helping New Teachers Survive and Thrive in Our Schools*. Indianapolis, IN: Jossey-Bass.
- Kelley, C./Protsik, J. (1997): Risk and Reward: Perspectives on the Implementation of Kentucky's School-based Performance Award Program. In: *Educational Administration Quarterly* 33, H. 4, S. 474-505.
- Kieffer, M.J./Lesaux, N.K./Snow, C.E. (2008): Promises and Pitfalls: Implications of NCLB for Identifying, Assessing, and Educating English Language Learners. In: Sunderman, G.L. (Hrsg.): *Holding NCLB Accountable: Achieving Accountability, Equity, and School Reform*. Thousand Oaks, CA: Corwin Press, S. 57-74.
- Kim, J./Sunderman, G.L. (2004): *Does NCLB Provide Good Choices for Students in Low Performing Schools?* Cambridge, MA: The Civil Rights Project at Harvard University.
- Kim, J./Sunderman, G.L. (2005): Measuring Academic Performance under NCLB: Implications for Educational Equity. In: *Educational Researcher* 34, H. 8, S. 3-13.
- Koretz, D.M. (2008): The Pending Reauthorization of NCLB: An Opportunity to Rethink the Basic Strategy. In: Sunderman, G.L. (Hrsg.): *Holding NCLB Accountable: Achieving Accountability, Equity, and School Reform*. Thousand Oaks, CA: Corwin Press, S. 9-26.

- Koretz, D.M./Barron, S.I./Mitchell, K.J./Stecher, B.M. (1996): Perceived Effects of the Kentucky Instructional Results Information System. Santa Monica, CA: RAND.
- Lee, J. (2006): Tracking Achievement Gaps and Assessing the Impact of NCLB on the Gaps: An In-depth Look into National and State Reading and Math Outcome Trends. Cambridge, MA: The Civil Rights Project at Harvard University.
- Lee, J. (2007): The Testing Gap: Scientific Trials of Test-driven School Accountability Systems for Excellence and Equity. Charlotte, NC: Information Age.
- Lee, J. (2008): Is Test-driven External Accountability Effective? Synthesizing the Evidence from Cross-State Causal-Comparative and Correlational Studies. In: Review of Educational Research 78, S. 608-644.
- Leithwood, K./Steinbach, R./Jantzi, D. (2002): School Leadership and Teachers' Motivation To Implement Accountability Policies. In: Educational Administration Quarterly 38, H. 1, S. 94-119.
- Linn, R.L. (2005): Conflicting Demands of No Child Left Behind and State Systems: Mixed Messages about School Performance. In: Education Policy Analysis Archives 13, H. 33.
- Loeb, H./Knapp, M./Elfers, A. (2008): Teachers' Response to Standards-based Reform: Probing Reform Assumptions in Washington State. In: Education Policy Analysis Archives 16, H. 8, S. 1-29.
- Louis, K.S./Febey, K./Schroeder, R. (2005): State-Mandated Accountability in High Schools: Teachers' Interpretations of a New Era. In: Educational Evaluation and Policy Analysis 27, S. 177-204.
- Malen, B./Croninger, R./Muncey, D./Redmond-Jones, D. (2002): Reconstituting Schools: „Testing“ the „Theory of Action“. In: Educational Evaluation and Policy Analysis 24, H. 2, S. 113-132.
- Marion, S.T./White, C./Carlson, D./Erpenbach, W.J./Rabinowitz, A./Sheinker, J. (2002): Making Valid and Reliable Decisions in Determining Adequate Yearly Progress. Washington, DC: Council of Chief State School Officers.
- Mathis, W.J. (2009): NCLB's Ultimate Restructuring Alternatives: Do They Improve the Quality of Education? East Lansing, MI: The Great Lakes Center for Education Research & Practice.
- McDermott, K.A. (2004): Incentives, Capacity, and Implementation: Evidence from Massachusetts Education Reform. In: Journal of Administration Research and Theory 16, S. 45-65.
- McLaughlin, M. (1990): The Rand Change Agent Study Revisited. In: Educational Researcher 19, H. 9, S. 11-16.
- McNeil, L. (2000): Contradictions of School Reform. New York/London: Routledge.
- Mintrop, H. (2004): Schools on Probation: How Accountability Works (and Doesn't Work). New York: Teachers College.
- Mintrop, H. (2008): Low-Performing Schools' Programs and State Capacity Requirements: Meeting the NCLB Educational Goals. In: Sunderman, G.L. (Hrsg.): Holding NCLB Accountable: Achieving Accountability, Equity, & School Reform. Thousand Oaks, CA: Corwin Press, S. 137-151.
- Mintrop, H./Sunderman, G. (2009): Predictable Failure of Sanctions-driven Accountability for School Improvement – and Why We May Retain It Anyway. In: Educational Researcher 38, H. 5, S. 353-364.
- Mintrop, H./Trujillo, T. (2005): Corrective Action in Low Performing Schools: Lessons for NCLB Implementation from First-Generation Accountability Systems. In: Education Policy Analysis Archives 13, H. 48.
- Mintrop, H./Trujillo, T. (2007): The Practical Relevance of Accountability Systems for School Improvement: A Descriptive Analysis of California Schools. In: Educational Evaluation and Policy Analysis 29, H. 4, S. 319-352.

- Neal, D./Schanzenbach, D.W. (2008): *Left Behind by Design: Proficiency Counts and Test-based Accountability*. Chicago: University of Chicago.
- Neild, R.C./Spiridakis, K. (2002): *Preliminary Results: Philadelphia New Teacher Survey*. Philadelphia, PA: Research for Action.
- Neild, R.C./Useem, B./Travers, E./Lesnick, J. (2003): *Once & for All: Placing a Highly Qualified Teacher in Every Philadelphia Classroom*. Philadelphia, PA: Research for Action.
- Nichols, S./Glass, G./Berliner, D. (2006): *High-Stakes Testing and Student Achievement: Does Accountability Pressure Increase Student Learning?* In: *Educational Policy Analysis Archives* 17, H. 1, S. 1-72.
- Novak, J.F.B. (2003, December): *Penalizing Diverse Schools? Similar Test Scores but Different Students Bring Federal Sanctions*. Berkeley: Policy Analysis for California Education.
- NYSED (New York State Education Department), Office of New York City School Improvement and Community Services (2003): *2002-03 Registration Review Initiative: A Summary of the Registration Review Report Findings*. New York City: Author.
- O'Day, J. (2002): *Complexity, Accountability, and School Improvement*. In: *Harvard Educational Review* 72, H. 3, S. 293-329.
- Oldham, G./Hackman, J.R. (2010): *Not What It Was and not What It Will Be: the Future of Job Design Research*. In: *Journal of Organizational Behavior* 31, H. 2-3, S. 463-479.
- Owens, A./Sunderman, G.L. (2006): *School Accountability Under NCLB: Aid or Obstacle for Measuring Racial Equity?* Cambridge, MA: The Civil Rights Project at Harvard University.
- Protheroe, N. (2008): *The Impact of Fidelity of Implementation in Effective Standards-based Instruction*. In: *Principal* 87, H. 6, S. 38-41.
- Raymond, M.E./Hanushek, E.A. (2003): *High-stakes Research: Accountability Works after All*. In: *Education Next* 3, H. 3, S. 48-55.
- Saltman, K.J. (2005): *The Edison Schools*. New York: Routledge.
- Scheerens, J./Bosker, R. (1997): *The Foundations of Educational Effectiveness*. Oxford, GB: Pergamon.
- Sennett, R. (2006): *The Culture of New Capitalism*. New Haven: Yale University Press.
- Skrla, L./Scheurich, J. (2004): *Displacing Deficit Thinking in School District Leadership*. In: Skrla, L./Scheurich, J. (Hrsg.): *Educational Equity and Accountability*. New York: Routledge Farmer, S. 109-132.
- Slavin, R. (2005): *Evidence-based Reform: Advancing the Education of Students at Risk*. Report Prepared for: *Renewing Our Schools, Securing Our Future; a National Task Force on Education*. o.V.
- Sunderman, G.L. (2001): *Accountability Mandates and the Implementation of Title I Schoolwide Programs: A Comparison of Three Urban Districts*. In: *Educational Administration Quarterly* 37, S. 503-532.
- Sunderman, G.L./Kim, J.S./Orfield, G. (2005): *NCLB Meets School Realities: Lessons from the Field*. Thousand Oaks, CA: Corwin Press.
- Sunderman, G.L./Orfield, G. (2006): *Domesticating a Revolution: No Child Left Behind and State Administrative Response*. In: *Harvard Educational Review* 76, H. 4, S. 526-556.
- Sunderman, G.L./Tracey, C.A./Kim, J./Orfield, G. (2004): *Listening to Teachers: Classroom Realities and No Child Left Behind*. Cambridge, MA: The Civil Rights Project at Harvard University.
- The Brown Center (2003): *Charter Schools: Achievement, Accountability, and the Role of Expertise*. Washington, DC: Brookings Institution.
- The Education Trust (2004): *Measured Progress: Achievement Rises and Gaps Narrow, but Slowly*. Washington, DC: Author.

- The Education Trust (2005): Stalled in Secondary: A Look at Student Achievement since the No Child Left Behind Act. Washington, DC: Author.
- Timar, T. (1997): The Institutional Role of State Education Departments. In: American Journal of Education 105, H. 3, S. 231-260.
- Timar, T./Kirp, D. (1988): State Efforts to Reform Schools: Treading Between a Regulatory Swamp and an English Garden. In: Educational Evaluation and Policy Analysis 10, H. 2, S. 75-88.
- Travers, E. (2003a): Characteristics of Schools under Diverse Providers 2002-2003. Philadelphia, PA: Research for Action.
- Travers, E. (2003b): Philadelphia School Reform: Historical Roots and Reflections on the 2002-2003 School Year under State Takeover: A Talk Initially Presented at The Friends Association of Higher Education and Friends Council on Education Conference at Swarthmore College and Pendle Hill. Penn GSE Perspectives on Urban Education 2, H. 1. URL: <http://www.urbanedjournal.org/commentaries/c0007.pdf>; Zugriffsdatum: 02.11.2005.
- Tyack, D. (1974): The One Best System: A History of Urban Education. Cambridge, MA: Harvard University Press.
- Useem, E. (2005): Learning from Philadelphia's School Reform: What Do the Research Findings Show so Far? Philadelphia, PA: Research for Action.
- Valenzuela, A. (2005): Leaving Children Behind: How "Texas-style" Accountability Fails Latino Youth. Albany: State University of New York Press.
- Verstegen, D./Knoeppel, R. (1998): Equal Education under the Law: School Finance Reform and the Courts. In: Journal of Law and Politics 14, H. 3, S. 555-592.
- Weiler, H. (1990): Comparative Perspectives in Decentralization in Education: An Exercise in Contradiction? In: Educational Evaluation and Policy Analysis 12, H. 4, S. 433-448.
- Wirt, F./Kirst, M. (1997): The Political Dynamics of American Education. Berkeley, CA: McCutchan.

Heinrich Mintrop, Associate Professor, Educational Policy and Leadership, Director of the Doctoral Program in Leadership for Educational Equity, Graduate School of Education, University of California, Berkeley, USA.

Anschrift: 3647 Tolman Hall, University of California, Berkeley, CA. 94720, USA
E-Mail: mintrop@berkeley.edu

Gail Sunderman, Ph.D., Senior Research Scientist at the George Washington University Center for Equity and Excellence in Education, Director of the Mid-Atlantic Equity Center.

Anschrift: MAEC at The George Washington University Center for Equity and Excellence in Education, 1555 Wilson Blvd., Suite 515, Arlington, Virginia 22209, USA
E-Mail: glsunderman@yahoo.com

Barbara Asbrand/Nina Heller/Sigrid Zeitler

Die Arbeit mit Bildungsstandards in Fachkonferenzen

Ergebnisse aus der Evaluation des KMK-Projektes *for.mat*

Zusammenfassung

Im Rahmen des KMK-Projekts for.mat wurden Materialien und Fortbildungskonzepte für die Fachkonferenzarbeit entwickelt, die Lehrkräfte im Prozess der Implementation der Bildungsstandards unterstützen sollen. Für die Evaluation wurden die Materialien an rheinland-pfälzischen Schulen erprobt. Die in dem Beitrag vorgestellten ausgewählten Ergebnisse der Evaluation beziehen sich auf die Relevanz impliziter Muster der Unterrichtsplanung und die Bedeutung der Heterogenität der Fachkonferenzen.

Schlüsselwörter: Bildungsstandards, kompetenzorientierter Unterricht, Unterrichts-entwicklung, Lehrerfortbildung, Evaluationsforschung

Working with Educational Standards in Faculty Meetings

Results from the Evaluation of the KMK-Project *for.mat*

Abstract

Within the KMK-project for.mat (KMK: The Standing Conference of the Ministers of Education and Cultural Affairs of the Laender in the Federal Republic of Germany), learning material and concepts for professional learning were developed in order to support teachers in the process of implementing educational standards. These materials were tested for evaluation purposes in different schools in Rhineland-Palatinate. The selected results of the evaluation study presented in this article refer to the relevance of implicit patterns in lesson planning and the relevance of heterogeneity in faculties.

Keywords: national educational standards, competence orientation, school improvement, teacher education and training, evaluation research

1. Einführung

Im KMK-Projekt *for.mat*¹ wurden Materialien entwickelt, mit Hilfe derer Fachkonferenzen im Prozess der Implementation der Bildungsstandards dabei unterstützt werden sollen, die Unterrichtspraxis in Richtung eines kompetenzorientierten Unterrichts weiterzuentwickeln. Das zentrale Ziel war die Unterstützung des Perspektivenwechsels von der Input- zur Outputsteuerung. Im Folgenden werden zunächst die mit dem Projekt *for.mat* verfolgten Konzepte eines kompetenzorientierten Unterrichts und der intendierten Unterrichtsentwicklung in Fachkonferenzen vorgestellt. Im zweiten Teil des Beitrags werden wesentliche Befunde der Evaluationsstudie dargestellt, in der nach dem Nutzen der Fortbildungs- und Unterrichtsmaterialien für Lehrkräfte im Prozess der Implementation der Bildungsstandards gefragt wurde. Gegenstand der Evaluation, die als dokumentarische Evaluationsforschung (vgl. Bohnsack 2007) durchgeführt wurde, war die Erprobung der in *for.mat* entwickelten Materialien an rheinland-pfälzischen Schulen.

2. Das KMK-Projekt *for.mat*

2.1 Bildungsstandards und kompetenzorientierter Unterricht

Mit der Einführung von Bildungsstandards sind für Lehrkräfte weitreichende Konsequenzen für die Planung und Durchführung ihres Unterrichts verbunden. In Bezug auf die Unterrichtsplanung und -reflexion stellt sich nun aus der Perspektive der Lehrkräfte nicht mehr die Frage, welche Inhalte bereits gelehrt wurden, sondern welche Kompetenzen die Schülerinnen und Schüler bereits erworben haben und wie sie individuell im weiteren Kompetenzerwerb gefördert werden können. „Kompetenzorientierter Unterricht“ (vgl. z.B. Blum u.a. 2006; Ziener 2006; Klinger 2009; Lersch 2010) ist gekennzeichnet von offenen Aufgaben, Lebensweltbezug sowie kompetenzorientierter Diagnose von Schülerleistungen als Ausgangspunkt für individuelle Förderung. Ziel ist der langfristige Kompetenzaufbau, der es den Lernenden ermöglicht, den eigenen Kompetenzzuwachs wahrzunehmen und zu reflektieren. Die Fokussierung auf die Förderung individueller Kompetenzentwicklung impliziert die verstärkte Integration mitlaufender Diagnose in die Unterrichtspraxis sowie differenzierende und individualisierende Lehr-Lernarrangements. Somit bietet eine kompetenzorientierte Unterrichtsgestaltung auch Potenziale im Umgang mit heterogenen Lerngruppen. Mit der Einführung von Kerncurricula (in einigen Bundesländern), die sich an den Bildungsstandards orientieren und im Gegensatz zu Lehrplänen Unterrichtsinhalte nicht im Detail definieren, gewinnen schuleige-

1 *for.mat* = „Bereitstellung von Fortbildungskonzeptionen und -materialien zur kompetenz- bzw. standardbasierten Unterrichtsentwicklung, vor allem Lesen, Geometrie, Stochastik“ (URL: www.kmk-format.de; Zugriffsdatum: 05.02.2012); federführend bei der Durchführung des KMK-Projekts war das Land Rheinland-Pfalz.

ne Curricula für die Entwicklung und Förderung fachlicher und fachübergreifender Kompetenzen an Bedeutung. Auch die Erstellung schuleigener Curricula, die den kumulativen Kompetenzerwerb über die Jahrgangsstufen hinweg fokussieren, an Stelle von vorrangig an Inhalten ausgerichteten Stoffverteilungsplänen, erfordert einen Perspektivenwechsel der Lehrkräfte von der Input- zur Outputorientierung (vgl. Cohnen/Klinger/Schaedler 2009).

2.2 Unterrichtsentwicklung in Fachkonferenzen

Auf Grund der Domänenspezifität der Bildungsstandards ist davon auszugehen, dass sich auf der Ebene der Einzelschule vornehmlich Fachkonferenzen um eine Umsetzung der Reform bemühen. Mit dem KMK-Projekt *for.mat* versucht man, an das für die Fächer Mathematik und Naturwissenschaften ausgelegte Modellprogramm SINUS („Programm zur Steigerung des Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“; vgl. Prenzel/Friedrich/Stadler 2009) anzuschließen und das Konzept der Unterrichtsentwicklung auch für alle Bildungsstandardfächer weiterzuentwickeln.² Explizites Ziel von *for.mat* ist es, die unterrichtsbezogene Kooperation der Lehrkräfte in den Fachkonferenzen und die Weiterentwicklung des eigenen Unterrichts anzuregen (vgl. Bethge/Priebe 2009; Cohnen/Schaedler 2009). Die Fachkonferenzen werden im KMK-Projekt *for.mat* als „professionelle Lerngemeinschaften“ angesehen (vgl. DuFour/Eaker 1998; Hord 2004; Bonsen 2005); die Materialien³ sollen dazu dienen, auf der Ebene der Fachkonferenz unterrichtsbezogene Diskussionen und Reflexionen anzuregen, den Perspektivenwechsel der Lehrkräfte zu fördern und den Blick auf die Entwicklung und Analyse von Schülerkompetenzen zu richten (vgl. Cohnen/Klinger/Schaedler 2009, S. 101f.).

2.3 Fortbildungskonzeptionen und -materialien aus *for.mat*

Mit der Einführung der Bildungsstandards stehen Lehrkräfte vor der Aufgabe, geeignete Lernumgebungen für den Kompetenzerwerb zu entwickeln. Im KMK-Projekt *for.mat* wurden Arbeitshilfen für Fachkonferenzen entwickelt, die sie bei der Planung und Analyse von kompetenzorientiertem Unterricht unterstützen sollen: Zum Beispiel wurden Analyseschemata für die Analyse von Aufgabenschwierigkeiten entwickelt sowie für jedes Fach eine Kompetenzmatrix erstellt. Mit dieser kann in den Fachkonferenzen eine Kompetenzexegese geleistet werden, indem die allge-

2 In SINUS bzw. dem Folgeprojekt SINUS-Transfer arbeiteten Fachkonferenzen über einen längeren Zeitraum gemeinsam an der Konzeption und Analyse von offenen, anwendungsbezogenen Aufgabenstellungen für den Unterricht. Die Fachkonferenzen wurden in ihrer Entwicklung von Set-Koordinatoren und -Koordinatorinnen kontinuierlich begleitet und arbeiteten in Schul-Sets auch schulübergreifend zusammen (vgl. Prenzel/Friedrich/Stadler 2009).

3 Die Materialien sind zugänglich in der Projektpublikation (Klinger 2009) und auf der Website des Projekts: www.kmk-format.de.

mein gehaltenen Kompetenzbeschreibungen der Bildungsstandards bezogen auf einen Unterrichtsgegenstand konkretisiert und unterschiedliche Anforderungsniveaus einzelner Aufgabenstellungen differenziert werden (vgl. Ziener 2006, S. 46). Für Fachkonferenzen bietet die Arbeit mit der Kompetenzmatrix die Möglichkeit, gemeinsam Schwerpunkte für den Unterricht zu setzen oder das Lehrwerk auf das Angebot an kompetenzorientierten Aufgabenstellungen hin zu analysieren (vgl. Klinger 2005; Klinger/Bünder 2006; Cohnen/Klinger/Schaedler 2009, S. 112ff.). Im Bereich der Naturwissenschaften wurden in *for.mat* Planungslandkarten entwickelt zu den Aspekten naturwissenschaftlichen Unterrichts „Aufgaben“, „Diagnose“, „Experiment“, „Planung“ und „Kontexte“, mit Hilfe derer über die Analyse des Ist-Zustandes, die Entwicklungsperspektive und die Bezugnahme auf die Bildungsstandards Möglichkeiten der Kompetenzentwicklung der Schülerinnen und Schüler in den Fokus gebracht werden können (vgl. Klinger 2009). Anders als die Kompetenzmatrices dienen die Planungslandkarten dem induktiven Vorgehen aus einer konkreten Fragestellung heraus, z.B.: „Welche Kriterien sind bei der Auswahl von Unterrichtsthemen besonders wichtig?“ Die Materialien aus *for.mat* sind in erster Linie für den Einsatz in moderierten Fachkonferenzen und schulinternen Lehrerfortbildungen, also für die Hand von Moderatorinnen und Moderatoren bzw. Lehrerfortbildnerinnen und -fortbildnern gedacht, können aber auch von Lehrkräften selbst in der Fachkonferenzarbeit eingesetzt werden.

3. Methodisches Vorgehen der Evaluationsstudie

Mit der Evaluation wurde die Fragestellung verfolgt, inwiefern die in dem KMK-Projekt *for.mat* entwickelten Materialien und Konzepte in der Praxis der Fachkonferenzarbeit einsetzbar sind. Für die Evaluation wurden die Materialien an acht rheinland-pfälzischen Schulen, darunter Gymnasien, Gesamtschulen und Regionale Schulen, in Fachkonferenzen unterschiedlicher Bildungsstandardfächer eingesetzt und erprobt.⁴ In diesen Schulen wurde jeweils eine Fachkonferenzsitzung, bei der die Arbeitshilfen aus *for.mat* durch erfahrene Fachmoderatorinnen bzw. -moderatoren vorgestellt wurden und mit ihnen gearbeitet wurde, teilnehmend beobachtet. Etwa ein halbes Jahr danach wurde eine Gruppendiskussion mit den beteiligten Lehrkräften

4 Längerfristige Entwicklungsprozesse auf Schul- bzw. Fachkonferenzebene zu initiieren und zu unterstützen ist das Ziel der Materialien und Konzepte des Projekts *for.mat*; gleichwohl war die Begleitung von Fachkonferenzen durch qualifizierte Fachmoderatorinnen und -moderatoren innerhalb des Projekts *for.mat* nicht vorgesehen und wurde an den genannten rheinland-pfälzischen Schulen nur zum Zweck der Erprobung der Materialien und Konzepte durchgeführt. Dabei ist zu beachten, dass im Rahmen der Evaluation lediglich eine einmalige moderierte Fachkonferenz in den Schulen durchgeführt wurde, während die Materialien und Konzepte aus *for.mat* tatsächlich für die Unterstützung längerfristiger Unterrichtsentwicklungsprozesse gedacht sind. Dies gilt es bei der Einschätzung der Befunde aus der Evaluation zu berücksichtigen.

geführt. Dieser Zeitraum sollte den Lehrkräften erlauben, die Materialien in der Fachkonferenzarbeit zu erproben.⁵

Im Mittelpunkt der Evaluationsstudie stand die Auswertung der Gruppendiskussionen der Lehrkräfte mit Hilfe der dokumentarischen Methode (vgl. Bohnsack 2007). Für die dokumentarische Evaluationsforschung ist die Unterscheidung zwischen Bewertungen und Werthaltungen entscheidend (vgl. Bohnsack 2006, S. 142ff.). Bei den Bewertungen handelt es sich um diejenigen evaluativen Äußerungen, die die Befragten explizit äußern, ihre auf der Ebene von Alltagstheorien reflektierten Einschätzungen über die Wirksamkeit von Maßnahmen oder Konzepten. Würde sich eine Evaluationsmaßnahme darauf beschränken, diese Einschätzungen auf der Ebene des explizit verfügbaren Wissens abzufragen, würden die Evaluationsergebnisse nicht über jene Bewertungen hinausgehen, die die Befragten vermutlich auch ohne die Evaluation vornehmen würden (vgl. ebd.). Wenn die Wirkung einer Maßnahme im Hinblick auf die Veränderung routinierter Alltagspraxen evaluiert werden soll, ist ein empirischer Zugriff auf die Werthaltungen notwendig, die den Befragten nicht reflexiv verfügbar sind, aber ihre Handlungspraxis bestimmen und in der dokumentarischen Interpretation zu den explizit geäußerten Bewertungen der Befragten in Beziehung gesetzt werden (vgl. ebd.). Diese für die dokumentarische Evaluationsforschung zentrale Unterscheidung zwischen Bewertung und Werthaltungen entspricht der Unterscheidung zwischen kommunikativem und konjunktivem Wissen (vgl. Mannheim 1980, S. 155ff.). Das Potenzial der dokumentarischen Methode besteht darin, empirischen Zugriff auf beide Wissensformen zu ermöglichen.⁶ Im Anschluss an Mannheim wird davon ausgegangen, dass sich in einer durch die Forscherinnen und Forscher nicht beeinflussten Kommunikation der Erforschten deren konjunktiver Erfahrungsraum aktualisiert und das konjunktive Wissen rekonstruiert werden kann. Dabei interessiert sich die dokumentarische Interpretation nicht für den Einzelfall – weder für die konkrete Gruppe (die Teilnehmenden der Gruppendiskussion oder die jeweilige Fachkonferenz) noch für die Individuen, die an einer Gruppendiskussion teilnehmen.⁷ Vielmehr wird im Zuge der komparativen Analyse von den Fällen ab-

5 Des Weiteren wurden für die Evaluation die in *for.mat* entwickelten Materialien analysiert und mit den Fachmoderatorinnen und -moderatoren, die mit den rheinland-pfälzischen Lehrkräften im Rahmen der Erprobung gearbeitet hatten, eine bilanzierende Gruppendiskussion durchgeführt.

6 Für die Forschungspraxis hat Bohnsack dies begrifflich systematisiert, indem er Orientierungsrahmen (konjunktives Wissen, implizite, handlungsleitende Orientierungen, Werthaltungen) und Orientierungsschemata (kommunikatives Wissen, Bewertungen) unterscheidet. Beide zusammen bilden die Orientierungsmuster der Erforschten, die Gegenstand der Rekonstruktion sind (Bohnsack 2003).

7 Zum Verhältnis von Individuum und Gruppe im Kontext des Gruppendiskussionsverfahrens ist des Weiteren anzumerken, dass die dokumentarische Interpretation von Gruppendiskussionen vor allem darauf abzielt, die von der jeweiligen Gruppe geteilten Orientierungsrahmen herauszuarbeiten (vgl. Przyborski 2004). Nur diese werden in der komparativen Analyse als typisch für den jeweiligen konjunktiven Erfahrungsraum der Gruppe angesehen und entsprechend als Ergebnis der empirischen Rekonstruktion berichtet.

strahiert und in der Typenbildung herausgearbeitet, inwiefern Orientierungsrahmen für bestimmte konjunktive Erfahrungsräume typisch sind (vgl. Bohnsack 2007, S. 141ff.). Die Daten aus der Evaluation des KMK-Projekts *for.mat* konnten in den größeren Zusammenhang der qualitativ-empirischen Begleitforschung zur Implementation der KMK-Bildungsstandards (vgl. Zeitler/Heller/Asbrand 2012) eingeordnet und auch für die Evaluationsstudie auf ein breites Sample und vielfältige empirische Vergleichshorizonte zurückgegriffen werden.⁸

4. Ergebnisse

Im Folgenden werden zentrale Befunde der Evaluationsstudie ausgehend von einem Fallbeispiel dargestellt. Zunächst wird exemplarisch gezeigt, wie eine Gruppe von Lehrkräften, die sich durch eine heteronome Orientierung und ein instruktivistisches Unterrichtsverständnis auszeichnet,⁹ mit kompetenzorientierten Aufgaben umgeht, die sie in der Fortbildung kennengelernt hat. Dies konnte in der Evaluationsstudie im kontrastierenden Fallvergleich mit Gruppen herausgearbeitet werden, die über Erfahrungen in der Unterrichtsentwicklung verfügen und andere Erwartungen an Fortbildungen formulieren. Der Fallvergleich kann hier aus Platzgründen allerdings nur angedeutet werden.

4.1 Fallbeispiel: Kompetenzorientierte Aufgaben als Anregung zur Reflexion über Unterricht?

Die Gruppe *Garonne* besteht aus drei Englischlehrerinnen und einem Englischlehrer einer Regionalen Schule in Rheinland-Pfalz. In der Gruppendiskussion werden die neuen Aufgabenformate und die Frage fokussiert, ob die angebotenen Materialien im Unterricht verwertbar sind. In der im Folgenden wiedergegebenen ersten Reaktion der Gruppe *Garonne* auf die Eingangsfrage der Gruppendiskussion zeigen sich die Defizitperspektive auf die Fähigkeiten der Schülerinnen und Schüler, die Ausrichtung der Unterrichtsvorbereitung dieser Gruppe von Englischlehrkräften am Lehrwerk und

8 Sowohl in dem größeren Forschungskontext als auch in der Evaluation des KMK-Projekts *for.mat* erwiesen sich – entgegen der anfänglichen Suchstrategie beim Sampling – unterschiedliche Fachkulturen empirisch als nicht bedeutsam für den Umgang mit den Bildungsstandards bzw. dem Konzept der Kompetenzorientierung. Die rekonstruierten Orientierungsmuster fanden sich gleichermaßen bei Lehrkräften aus mathematisch-naturwissenschaftlichen Fächern wie bei Lehrerinnen und Lehrern, die sprachliche Fächer unterrichten. In der komparativen Analyse stellten sich vielmehr die Erfahrungen der Lehrkräfte bei der Bewältigung schulischer Entwicklungs Herausforderungen in Schul- und Unterrichtsentwicklungsprozessen sowie die schulformspezifischen professionellen Bezugssysteme (Pädagogik vs. Fachwissenschaft) als für den Umgang mit Bildungsstandards relevante konjunktive Erfahrungsräume heraus (vgl. ausführlich Zeitler/Heller/Asbrand 2012).

9 Vgl. zu den rekonstruierten Rahmenorientierungen Autonomie vs. Heteronomie sowie Lehr-Lernverständnis ausführlich Zeitler/Heller/Asbrand 2012.

das Muster der Unterrichtsvorbereitung auf der Ebene der habitualisierten Routinen: die Suche nach fertigen, didaktisierten Unterrichtsmaterialien, deren Anforderungen zu den Schülerfähigkeiten passen.

- Bf [...] wir haben uns aufgrund der neuen Bildungsstandards auch überlegt jetzt ein anderes Englischbuch einzuführen;
 [...]

 Cm ham aber dann festgestellt dass unser neues Buch da doch sehr viel Möglichkeiten bietet; und äh woll'n jetzt einfach mal sehn wie wir uns da einarbeiten;(4)

 Bf Ja und ich denke irgendwo wenn jetzt irgendwelches neue Material kommt es gibt ja mittlerweile auch schon einiges [...]; es es gibt ja auch von Verlagen jetzt schon so einige Sachen die also speziell auf diese Bildungsstandards hinarbeiten und so'n bisschen Material ham wir davon auch und da guckt man dann auch schon mal was kann man verwerten und ähm [...]denk ich also zumindest wir hier als–als–als Stufenlehrer ham immer schon mal geguckt was kann man da nehmen, was–was passt da jetzt zu, was kann man einfach mal probieren jetzt in grade in den siebten Klassen

 Cm Problem dabei is ganz klar äh unsre Grundkurse sind gerade bei Leseverstehen ganz schön gefordert um net zu sagen

 ?f |°überfordert°

 Cm überfordert

 Af | überfordert(.)wollt ich grad sagen

Am Ende dieser Sequenz wird die Defizitperspektive dieser Lehrkräfte auf das Lernen der Schülerinnen und Schüler deutlich. Deren als unzureichend wahrgenommene Fähigkeiten verhindern aus Sicht der Lehrkräfte die Umsetzung der neuen Konzepte im Unterricht. Im Unterschied zu den Diskursen anderer Lehrergruppen diskutieren diese Lehrkräfte nicht, wie sie die Kompetenzen der Lernenden in ihrem Unterricht fördern könnten, sondern das niedrige Kompetenzniveau wird als gegeben hingenommen, als eine (unveränderliche) Rahmenbedingung von Unterricht, mit der es umzugehen gilt. Bf teilt zwar zu Beginn der Gruppendiskussion mit, die Gruppe habe sich mit den Bildungsstandards auseinandergesetzt; in dem obigen Ausschnitt zeigt sich allerdings die Art und Weise der Auseinandersetzung mit der Reform: Die Umsetzung der Bildungsstandards für den Unterricht wird von den Verlagen erwartet. Die Aktivität der Lehrkräfte beschränkt sich auf die Auswahl eines geeigneten Lehrwerks oder anderer didaktisch aufbereiteter Materialien, die „speziell auf diese Bildungsstandards hinarbeiten“. Deutlich wird auch der Modus der Unterrichtsvorbereitung, nämlich die Suche nach Passung zwischen didaktisch aufbereiteten Materialien und den Anforderungen des eigenen Unterrichts.

In der im Folgenden wiedergegebenen Sequenz, in der die Lehrkräfte direkt nach ihrer Arbeit mit den Materialien aus *for.mat* gefragt werden, werden die in der Fachkonferenz vorgestellten Planungs- und Analyseinstrumente nicht thematisiert; sie sind für die Gruppe ohne Relevanz. Vielmehr diskutieren die Lehrkräfte die im Unterricht unmittelbar verwertbaren Aufgaben:

- Bf So geht's uns ja auch mit diesem einen Heft da sind wirklich richtig gute Texte drin was wir auch schon n paar Mal jetzt
- Df Ja ja
- Bf paar von eingesetzt haben;
- [...]
- Df Allein schon dieses Gedicht was ich mit denen mal gemacht hatte dacht ich auch so n schönes Herbstgedicht jetzt ge (.) da war fast jedes dritte Wort für die was die nicht wussten ge, also war- ging erstmal wirklich so'ne Wortarbeit im Ganzen vorher und [...]; und da verschätz ich mich auch oft gell,(3)
- Bf Aber diese Hellow- diese Pyramide diese Herbstpyramide mit diesen Wörtern die denen ja unheimlich Spass gemacht; ne, und da saß nachher auch; das saß richtig gut, also (5)
- [...]
- Cm und ähm war jetzt mit'm Halloween das hat halt thematisch mal schön gepasst und da sind se eigentlich auch ganz interessiert rangegangen; das geht gar net so schlecht (3) ich denk jetzt vor Weihnachten ham wa auch wieder Material da könn ma au wieder gucken (.) zum Thema Weihnachten [...]
- Df Ja jede Menge
- ?f Ja ja viel schönes schönes

In der Auseinandersetzung mit den Unterrichtsmaterialien geht es um die Passung der Aufgaben zu den Schülerfähigkeiten, die allerdings nur undifferenziert als unveränderliche Rahmenbedingung des eigenen Unterrichtens wahrgenommen und nicht im Sinne einer Kompetenzdiagnostik systematisch analysiert werden, wie im Fallvergleich mit Gruppen deutlich wird, die in den Gruppendiskussionen über die Lernfähigkeit und die Förderung auch der lernschwächeren Schülerinnen und Schüler reflektieren (vgl. Zeitler/Heller/Asbrand 2012). Das Projekt *for.mat* reduziert sich in der Wahrnehmung dieser Lehrkräfte auf eine Sammlung von mehr oder weniger passenden Unterrichtsmaterialien. Dabei ist nicht von Bedeutung, inwiefern diese geeignet sind, den Kompetenzerwerb der Schülerinnen und Schüler zu fördern, sondern für die Lehrkräfte spielen vorrangig die Inhalte, hier die jahreszeitlichen Themen, bei der Einschätzung der Eignung der Unterrichtsmaterialien eine Rolle. Die in der moderierten Fachkonferenz präsentierten kompetenzorientierten Aufgaben werden genauso verwendet wie andere Unterrichtsmaterialien auch.

Zusammenfassend kann festgehalten werden:

(1) Die Lehrkräfte bewerten die kennengelernten Materialien positiv; es zeigt sich allerdings auf der Ebene der Werthaltungen, dass die Unterrichtsgestaltung weiterhin ausschließlich inputorientiert gedacht wird; die Kompetenzen der Schülerinnen und Schüler werden außerhalb des Einflussbereichs des unterrichtlichen Handelns der Lehrkräfte gesehen. Vielmehr steht die Vermittlung von Inhalten im Vordergrund der Unterrichtspraxis.

(2) Mit diesem instruktivistischen Verständnis von Unterricht korrespondiert das implizite Muster der Unterrichtsvorbereitung. Unterricht zu planen bedeutet für die Mitglieder der Gruppe *Garonne* nicht, Schülerfähigkeiten zu diagnostizieren und darauf aufbauend lernförderliche Unterrichtsarrangements zu konzipieren, sondern die

Unterrichtsvorbereitung reduziert sich auf die Wahl eines geeigneten Inputs (vgl. auch Bromme 1986; Vollstädt 1996; Seel 1997).

Die Bewertung der moderierten Fachkonferenzen durch die Lehrkräfte ist von der Erwartung an Fortbildungen dominiert, unterrichtspraktisch Verwertbares kennenlernen zu wollen (vgl. auch Lipowsky 2004). Zwar ist es fortbildungsdidaktisch sinnvoll, sich in einer einmaligen Veranstaltung den motivierenden Charakter der kompetenzorientierten Aufgaben und ihre unterrichtspraktische Relevanz für den Einstieg in die Thematik zu Nutze zu machen. Allerdings führte dies hier dazu, dass den Lehrkräften der innovative Charakter der in *for.mat* entwickelten Materialien trotz der in der moderierten Fachkonferenz ebenfalls eingesetzten Kompetenzmatrices nicht deutlich wurde.

Wie bereits angedeutet, wurde das Orientierungsmuster der Gruppe *Garonne* im kontrastierenden Fallvergleich mit Lehrergruppen rekonstruiert, die über umfangreiche Erfahrungen in der fachdidaktischen Unterrichtsentwicklung verfügen. Im Unterschied zu den Mitgliedern der Gruppe *Garonne* zeigen sich die in der Unterrichtsentwicklung erfahrenen Lehrerinnen und Lehrer unzufrieden mit moderierten Fachkonferenzsitzungen, die ihnen lediglich unterrichtspraktisch verwertbare Aufgaben präsentieren. Sie äußern vielmehr Interesse an fachlichen bzw. fachdidaktischen Zusammenhängen; in ihrer Vorstellung von Unterrichtsplanung machen schöne Aufgaben allein noch keinen guten Unterricht. Im Fall der in der Unterrichtsentwicklung erfahrenen Lehrkräfte wäre die Anschlussfähigkeit des Projektes *for.mat* möglicherweise größer gewesen, wenn in der moderierten Fachkonferenz mit den Planungs- und Analyseinstrumenten gearbeitet worden wäre.

4.2 Schlussfolgerungen und Ausblick

(1) Es zeigt sich, dass für die Weiterentwicklung von Unterricht nicht allein die Vorstellungen von Unterricht relevant sind, sondern auch die impliziten Muster der Unterrichtsplanung. Wie die Evaluationsergebnisse zeigen, bewirkt das Angebot von kompetenzorientierten Aufgaben allein keinen Perspektivenwechsel bei den Lehrkräften. Vielmehr besteht bei einer Konzentration auf Aufgaben die Gefahr, dass diese – da sie den Charakter von Unterrichtsmaterial haben – von den Lehrkräften entsprechend der vorherrschenden Muster der Unterrichtsvorbereitung (vgl. auch Bromme 1986; Vollstädt 1996) weiterhin als Unterrichtsinput wahrgenommen werden. Erst die deutliche Fokussierung auf die Analyse von Prozessen des Kompetenzerwerbs sowie die Reflexion der bisherigen Unterrichtspraxis würden aber zu einer veränderten Lernkultur im Sinne kompetenzorientierten Unterrichts führen. Die Erprobung der strukturierten Arbeitshilfen aus *for.mat* in den moderierten Fachkonferenzen zeigte, dass ein einmaliger Kontakt mit diesen Analyse- und Planungsinstrumenten vor allem für Lehrkräfte, die noch keine Erfahrung in der Unterrichtsentwicklung haben, nicht ausreichend ist. Es ist davon auszugehen, dass die Arbeitshilfen zur kom-

petenzorientierten Unterrichtsplanung erst in längerfristigen begleiteten Entwicklungsprozessen, wie sie mit dem Projekt *for.mat* beabsichtigt sind, ihr volles Potenzial entfalten können. Hier zeigt sich der erwartbare Befund, dass die Wirkung von punktuellen Lehrerfortbildungen begrenzt ist (vgl. Lipowsky 2004). An den Fachkonferenzen mit Erfahrungen in der Unterrichtsentwicklung kann beobachtet werden, dass eine Veränderung der Unterrichtspraxis in einem längerfristigen Entwicklungsprozess erreicht werden kann, der insbesondere ein Reflexionsprozess über Unterricht ist (vgl. auch Altrichter 2000). Denn nicht-reflektierte, in Routinen verfügbare Handlungsmuster können nur verändert werden, wenn sie der Reflexion zugänglich gemacht werden. Eine Professionalisierung der Fachkonferenzen und eine Veränderung grundlegender Planungsmuster werden nicht erreicht, indem andere, jetzt kompetenzorientierte Unterrichtsmaterialien zur Verfügung gestellt werden, sondern indem ein Ort der Reflexion über die unterrichtliche Praxis geschaffen wird. Hierfür bedarf es entsprechender zeitlicher und personeller Ressourcen und insbesondere der Begleitung durch qualifizierte und unterstützende Beraterinnen und Berater (vgl. auch Ostermeier 2004; Friedrich 2009; Fußangel/Gräsel 2009).

(2) Aus den Befunden der Evaluation kann gefolgert werden, dass die *Unterschiedlichkeit der Erfahrungen* von Lehrkräften bei der Planung von Fortbildungsmaßnahmen noch viel stärker berücksichtigt und die Fortbildungskonzepte und -materialien entsprechend differenziert werden müssen. Eine gelingende Implementation der Bildungsstandards hängt demnach auch von der Passung der Implementationsmaßnahme mit den Vorerfahrungen und Orientierungsmustern der Lehrerinnen und Lehrer ab. Fachkonferenzen bzw. deren Mitglieder unterscheiden sich im Umgang mit den Bildungsstandards je nach ihren Erfahrungen in Prozessen der Schul- und Unterrichtsentwicklung. Für die einen sind Planungs- und Analyseinstrumente zu anspruchsvoll, und es werden Hinweise auf unterrichtspraktisch verwertbare Materialien erwartet; für die anderen stellen Aufgaben und unterrichtspraktische Materialien keinen Impuls für weitere Entwicklungsprozesse dar. Aus diesem Grund erreichen einheitliche Maßnahmen zur Innovationsübernahme nicht unbedingt alle beteiligten Lehrkräfte. Altrichter und Heinrich (2006) sprechen in diesem Zusammenhang von „dünnen Implementationsstrategien“. Für erfolgreiche Programme der Unterrichtsentwicklung müsste vielmehr eine Differenzierung der Fortbildungsszenarien stattfinden (vgl. Goldstein 2009; Lipowsky 2004; Oelkers/Reusser 2008, S. 268ff.). Dies verweist erneut auf die Notwendigkeit qualifizierter Beraterinnen und Berater. Denn die Frage, ob eine Fortbildungsveranstaltung oder eine moderierte Fachkonferenz für eine konkrete Fachkonferenz angemessen ist, entscheidet sich weniger an den Materialien, die die Fortbildnerinnen und Fortbildner zur Verfügung haben. Von sehr viel größerer Bedeutung sind die Entscheidungen bezüglich der inhaltlichen und methodischen Gestaltung der Veranstaltungen. Ebenso wie der für Schülerinnen und Schüler geplante kompetenzorientierte Unterricht sollte auch das Lernangebot für Lehrerinnen und Lehrer adaptiv, d.h. entsprechend der je spezifischen Bedürfnisse und Voraussetzungen konzipiert sein. Die Angemessenheit

dieser fortbildungsdidaktischen Entscheidungen kann wiederum durch die Qualifizierung der Beraterinnen und Berater – Gegenstand des zweiten Teilprojekts von *for.mat* (vgl. Goldstein 2009) – beeinflusst werden.

Literatur

- Altrichter, H. (2000): Schulentwicklung und Professionalität. Bildungspolitische Entwicklungen und neue Anforderungen an Lehrer/innen. In: Bastian, J./Helsper, W./Reh, S./Schelle, C. (Hrsg.): Professionalisierung im Lehrerberuf. Von der Kritik der Lehrerrolle zur pädagogischen Professionalität. Opladen: Leske + Budrich, S. 145-163.
- Altrichter, H./Heinrich, M. (2006): Evaluation als Steuerungsinstrument im Rahmen eines „neuen Steuerungsmodells“ im Schulwesen. In: Böttcher, W./Brohm, M./Holtappels, H.G. (Hrsg.): Evaluation im Bildungswesen. Weinheim u.a.: Juventa, S. 51-64.
- Bethge, T./Priebe, B. (2009): Unterrichtsentwicklung und Schulentwicklung. In: Klinger, U. (Hrsg.): Mit Kompetenz Unterricht entwickeln. Fortbildungskonzepte und -materialien. Troisdorf: Bildungsverlag EINS, S. 69-99.
- Blum, W./Drüke-Noe, C./Hartung, R./Köller, O. (Hrsg.) (2006): Bildungsstandards Mathematik: konkret. Sekundarstufe I. Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen. Berlin: Cornelsen Scriptor.
- Bohnsack, R. (2003): Orientierungsmuster. In: Bohnsack, R./Marotzki, W./Meuser, M. (Hrsg.): Hauptbegriffe Qualitativer Sozialforschung. Opladen: Barbara Budrich, S. 132-133.
- Bohnsack, R. (2006): Qualitative Evaluation und Handlungspraxis – Grundlagen dokumentarischer Evaluationsforschung. In: Flick, U. (Hrsg.): Qualitative Evaluationsforschung. Reinbek bei Hamburg: Rowohlt, S. 135-155.
- Bohnsack, R. (2007): Rekonstruktive Sozialforschung. Einführung in Methodologie und Praxis qualitativer Forschung. Opladen: Barbara Budrich.
- Bonsen, M. (2005): Professionelle Lerngemeinschaften in der Schule. In: Holtappels, H.G./Höhmann, K. (Hrsg.): Schulentwicklung und Schulwirksamkeit. Systemsteuerung, Bildungschancen und Entwicklung der Schule. Weinheim u.a.: Juventa, S. 180-195.
- Bromme, R. (1986): Die alltaegliche Unterrichtsvorbereitung des (Mathematik-)Lehrers im Spiegel empirischer Untersuchungen. In: Journal für Mathematik-Didaktik 7, H. 1, S. 3-22.
- Cohnen, T./Klinger, U./Schaedler, K. (2009): Die Fachkonferenz und Fachgruppe als Ort der Unterrichtsentwicklung. In: Klinger, U. (Hrsg.): Mit Kompetenz Unterricht entwickeln. Fortbildungskonzepte und -materialien. Troisdorf: Bildungsverlag EINS, S. 101-126.
- DuFour, R./Eaker, R. (1998): Professional Learning Communities at Work. Best Practices for Enhancing Student Achievement. Bloomington, Indiana: National Educational Service.
- Friedrich, A. (2009): Die Bedeutung der lokalen Koordination – Was berichten die Lehrkräfte. In: Prenzel, M./Friedrich, A./Stadler, M. (Hrsg.): Von SINUS lernen – Wie Unterrichtsentwicklung gelingt. Stuttgart/Seelze-Velber: Klett/Kallmeyer, S. 164-167.
- Fußangel, K./Gräsel, C. (2009): Die Kooperation in schulübergreifenden Lerngemeinschaften. Die Arbeit der Sets im Projekt „Chemie im Kontext“. In: Maag Merki, K. (Hrsg.): Kooperation und Netzwerkbildung. Seelze-Velber: Kallmeyer, S. 120-131.
- Goldstein, K. (Hrsg.) (2009): Kompetenzprofil und Qualifizierungskonzept für Beraterinnen und Berater für Unterrichtsentwicklung. In: Klinger, U. (Hrsg.): Mit Kompetenz Unterricht entwickeln. Fortbildungskonzepte und -materialien. Troisdorf: Bildungsverlag EINS, S. 133-147.
- Hord, Shirley M. (Hrsg.) (2004): Learning Together, Leading Together. Changing Schools Through Professional Learning Communities. New York/London/Oxford, Ohio: College Press.

- Klinger, U. (2005): „Mit Bildungsstandards Unterrichts- und Schulqualität entwickeln.“ In: Friedrich Jahresheft 2005: Standards, S. 130-143.
- Klinger, U. (Hrsg.) (2009): Mit Kompetenz Unterricht entwickeln – Fortbildungskonzepte und -materialien. Troisdorf: Bildungsverlag EINS.
- Klinger, U./Bünder, W. (2006): Kompetenzorientierte Unterrichtsplanung. Die Entwicklung einer Kompetenzmatrix auf der Grundlage von Bildungsstandards. In: Unterricht Chemie 17, S. 14-18.
- Lersch, R. (2010): Didaktik und Praxis kompetenzfördernden Unterrichts. URL: <http://www.schulpaedagogik-heute.de/index.php/artikel-112.html>; Zugriffsdatum 18.01.2012.
- Lipowsky, F. (2004): Was macht Fortbildungen für Lehrkräfte erfolgreich? Befunde der Forschung und mögliche Konsequenzen für die Praxis. In: Die Deutsche Schule 96, H. 4, S. 462-479.
- Mannheim, K. (1980): Strukturen des Denkens. Frankfurt a.M.: Suhrkamp.
- Oelkers, J./Reusser, K. (2008): Qualität entwickeln – Standards sichern – mit Differenzen umgehen. Hrsg. vom Bundesministerium für Bildung und Forschung. Bonn/Berlin: BMBF.
- Ostermeier, C. (2004): Kooperative Qualitätsentwicklung in Schulnetzwerken. Münster u.a.: Waxmann.
- Prenzel, M./Friedrich, A./Stadler, M. (Hrsg.) (2009): Von SINUS lernen – Wie Unterrichtsentwicklung gelingt. Stuttgart/Seelze-Velber: Klett/Kallmeyer.
- Przyborski, A. (2004): Gesprächsanalyse und dokumentarische Methode. Qualitative Auswertung von Gesprächen, Gruppendiskussionen und anderen Diskursen. Wiesbaden: VS.
- Seel, A. (1997): Von der Unterrichtsplanung zum konkreten Lehrerhandeln – Eine Untersuchung zum Zusammenhang von Planung und Durchführung von Unterricht bei Hauptschullehrerstudentinnen. In: Unterrichtswissenschaft 25, H. 3, S. 257-273.
- Vollstädt, W. (1996): Unterrichtsplanung im Schulalltag. Ergebnisse einer empirischen Untersuchung. In: Pädagogik 48, H. 4, S. 17-22.
- Zeitler, S./Heller, N./Asbrand, B. (2012): Bildungsstandards in der Schule. Eine rekonstruktive Studie zur Implementation der Bildungsstandards. Münster u.a.: Waxmann.
- Ziener, G. (2006): Bildungsstandards in der Praxis. Kompetenzorientiert unterrichten. Stuttgart/Seelze-Velber: Klett/Kallmeyer.

Barbara Asbrand, Prof. Dr., geb. 1967, Universitätsprofessorin im Fachbereich Erziehungswissenschaften der Goethe-Universität Frankfurt am Main.

Anschrift: Senckenberganlage 15, 60054 Frankfurt a.M.
E-Mail: b.asbrand@em.uni-frankfurt.de

Nina Heller, Dr., geb. 1981, Lehrerin an der Dr. Ernst Schmidt-Realschule Ebern.

Anschrift: Franz-Ludwig-Straße 20, 96047 Bamberg
E-Mail: ninaheller@gmx.de

Sigrid Zeitler, Dr., geb. 1980, wissenschaftliche Mitarbeiterin am Lehrstuhl für Schulpädagogik der Friedrich-Alexander-Universität Erlangen-Nürnberg.

Anschrift: Regensburger Straße 160, 90478 Nürnberg
E-Mail: Sigrid.Zeitler@ewf.uni-erlangen.de

Franziska Schulze

Folgen zentraler Lernstandserhebungen: Unterscheiden sich Lehrkräfte in der Auseinandersetzung mit Lernstandserhebungen? Empirische Befunde zu Formen unterrichtsbezogener Steuerungslogik

Zusammenfassung

Mittels zentraler Lernstandserhebungen erhalten Schulleitungen und Lehrkräfte Informationen über die Leistungen ihrer Schülerinnen und Schüler. Auf der Grundlage einer Erhebung mit Thüringer Lehrkräften (n=43) wurde der bisher ungeklärten Frage nachgegangen, ob sich individuelle Unterschiede in der Auseinandersetzung mit den Ergebnissen des Thüringer Kompetenztests zeigen. Im Ergebnis können verschiedene Formen der unterrichtsbezogenen Steuerungslogik identifiziert und Unterschiede hinsichtlich der Auseinandersetzung von Lehrkräften mit unterschiedlichen Unterrichtsfächern aufgezeigt werden.

Schlüsselwörter: Neue Steuerung, Lernstandserhebungen, Ergebnismeldung, Lehrkräfte

Consequences of Standardized Performance Tests: Do Teachers Differ in Their Interpretation of Standardized Performance Tests?

Empirical Findings to Types of a Teaching-related 'Control Logic'

Abstract

As a result of standardized performance tests, school administration and teachers receive information concerning student achievement. This study examines how the feedback information, resulting from the Thüringer Kompetenztest, is interpreted and handled by teachers, based on a survey with Thuringian teachers (n=43). Different types of 'control logic' in schools are identified. Additionally, differences in the interpretation of evaluation results can be found between teachers of different subjects.

Keywords: new governance, performance tests, evaluation feedback, teachers

1. Neue Steuerung: Lernstandserhebungen als Ausgangspunkt für die Schul- und Unterrichtsentwicklung

Die Schulsysteme der Länder der Bundesrepublik Deutschland erfahren seit einigen Jahren eine Reihe tiefgreifender Veränderungen. Intendiert ist dabei im Allgemeinen eine Umorientierung von einer input- hin zu einer outputorientierten Form der Steuerung von schulischen Arbeits- und Entwicklungsprozessen (vgl. Oelkers/Reusser 2008). Mit dieser Strategie, auch *Neue Steuerung* genannt, wird vielfach der Anspruch verbunden, Schulleitungen und Lehrkräften ein höheres Maß an Verantwortung für die erfolgreiche Gestaltung der Schul- und Unterrichtsorganisation zu übertragen. Neben der Formulierung bundesweiter Bildungsstandards stellt die Einführung von Lernstandserhebungen¹ ein wesentliches Element der Umorientierung dar. „Lernstandserhebungen beziehen sich auf umfassendere Lern- bzw. Kompetenzentwicklungen (bisher erreichte Fertigkeiten, Fähigkeiten, Kenntnisse) und nicht wie Klassenarbeiten auf das unmittelbar vorangegangene Unterrichtsgeschehen“ (Peek/Dobbelstein 2006, S. 41). Durch die Rückmeldung der Ergebnisse aus diesen Erhebungen erhalten Lehrkräfte und Schulleitungen wissenschaftlich erhobene Informationen über den fachbezogenen Lernstand ihrer Schülerinnen und Schüler sowie indirekt damit auch über das Ergebnis ihrer Arbeit. Diese fachleistungsbezogenen Informationen über die Wirksamkeit des Geleisteten sollen den schulischen Akteurinnen und Akteuren bei der Planung und Gestaltung der Qualitätsentwicklungsarbeit auf der Ebene des Unterrichts helfen (vgl. Altrichter 2010; Netzwerk Empiriegestützte Schulentwicklung 2006). In den letzten Jahren wurde eine Vielzahl von Lernstandserhebungen durchgeführt; mittlerweile gibt es in allen Bundesländern Lernstandserhebungen, die zum Teil vom Institut zur Qualitätsentwicklung im Bildungswesen zentral entwickelt und dezentral in den Ländern eingesetzt und deren Ergebnisse zurückgemeldet werden (**Vergleichsarbeiten** in den Fächern Deutsch und Mathematik in der dritten Klasse der Grundschule [VERA 3], **Vergleichsarbeiten** in den Fächern Deutsch, Englisch und Mathematik in der sechsten Klasse [VERA 6] und **Vergleichsarbeiten** in den Fächern Deutsch, erste Fremdsprache und Mathematik in der achten Klasse [VERA 8]). Darüber hinaus gibt es zusätzlich landesspezifische Verfahren, zu denen z.B. **MARKUS** (**M**athematik-**G**esamterhebung **R**heinland-**P**falz: **K**ompetenzen, **U**nterricht, **S**chulkontext; vgl. Schrader/Helmke 2004), **Lernstand 9** bzw. **Lernstand 8** in Nordrhein-Westfalen (vgl. Kühle/Peek 2007) und die **Kompetenztests** in Thüringen (vgl. Nachtigall/Jantowski 2007) zählen.

1 Der Begriff der Lernstandserhebungen wird im weiteren Textverlauf als Synonym für ebenfalls gebräuchliche Begriffe wie Schulleistungstests, Vergleichsarbeiten, Kompetenztests, Gesamterhebungen oder Qualitätsuntersuchungen verwendet. Gemeint sind regelmäßig durchgeführte bundeslandweite Erhebungen zu Lern- bzw. Kompetenzentwicklungsständen der Schülerinnen und Schüler in bestimmten Kernfächern (meist Deutsch, Mathematik und Englisch).

1.1 Nutzung der Ergebnisrückmeldungen aus Lernstandserhebungen

Es ist davon auszugehen, dass die Nutzung von Ergebnisrückmeldungen durch verschiedene Aspekte bedingt wird. Groß Ophoff u.a. (2006) unterscheiden als wesentliche Aspekte erstens die Form und den Inhalt der dargestellten Informationen, zweitens die individuellen Bedingungen der Rezipierenden und drittens schulische und externe Bedingungen (vgl. ebd., S. 32). Die Daten, welche den Adressatinnen und Adressaten der Rückmeldung zur Verfügung gestellt werden, können nach dem Aggregationsniveau in Informationen über das Bundesland, die Schule, die Klasse und über den Lernstand jedes einzelnen Schülers und jeder einzelnen Schülerin unterschieden werden (vgl. Schneewind/Kuper 2009). Außerdem spielt nach Altrichter (2010) die Auswahl der Bezugsnorm (soziale, ipsative/individuelle und sachliche/kriteriale) für die Rückmeldung der Ergebnisse eine Rolle. Dieser erste Aspekt der Informationsqualität ist insofern von Bedeutung, als dass er die folgende Nutzung gewissermaßen präkonfiguriert. Er ist nach Rolff (2002) jedoch nicht hinreichend dafür, dass tatsächlich Veränderungen in Schulen vollzogen werden. Auch wenn davon ausgegangen werden kann, dass die Daten rezipiert werden, stellt sich die Frage, in welcher Weise die Daten in Abhängigkeit der Aspekte individueller und schulischer Bedingungen genutzt werden.

1.2 Modelle zur Ergebnisnutzung

Helmke (2004) hat ein Rahmenmodell vorgelegt, das diesen möglichen Nutzungsprozess in Abhängigkeit von individuellen und schulischen Bedingungen in Form von vier aufeinander bezogenen Phasen beschreibt. Die (1.) *Rezeption* von leistungsbezogenen Daten aus den Lernstandserhebungen beinhaltet sowohl die erfolgreiche Übermittlung und Wahrnehmung als auch „die kognitive Verarbeitung des semantischen Gehalts der zurückgemeldeten Informationen (Ergebnisse)“ (Hartung-Beck/Diemer 2009, S. 216). Dabei sollten die zurückgemeldeten Informationen so aufbereitet und kommentiert sein, dass Fehlinterpretationen und Missverständnisse seitens der Lehrkräfte auf ein Minimum reduziert werden. Bei der (2.) *Reflektion* wird nach den Gründen und Ursachen für gute oder schlechte bzw. erwartete oder erwartungswidrige Ergebnisinformationen gesucht. Dies ist der eigentliche Schwerpunkt für pädagogische Impulse, denn je „intensiver die [...] Ergebnisse innerschulisch für daran anknüpfende Diskussionen genutzt werden [...], desto höher sind die Chancen, dass daraus fundierte und realistische Maßnahmen abgeleitet werden“ (Helmke 2004, S. 102). Die Rezeption und Reflektion kann zu verschiedenen (3.) *Aktionen* in den Bereichen der Schul- und Unterrichtsentwicklung führen. Die Wirkung der eingeleiteten Handlungen und Aktionen wird dem Modell zufolge schließlich durch eine (4.) *Evaluation* überprüft. Die tatsächliche Ausprägung dieser vier Schritte hängt nach dem Modell Helmkes einerseits von individuellen und andererseits von schulischen und externen Bedingungen ab. Bensen und von der Gathen (2004) haben

ein alternatives Rahmenmodell zur Rückmeldungsnutzung entwickelt, welches die Einzelschule fokussiert. Dabei werden neben verschiedenen Arten des Nutzens (instrumenteller, konzeptioneller, symbolischer und strategischer Nutzen) sowie beabsichtigten und unbeabsichtigten Wirkungen von ErgebnISRückmeldungen auch unterschiedliche Faktoren dargestellt, welche die Art und Intensität der Nutzung von ErgebnISRückmeldungen aus Lernstandserhebungen bedingen. „Das ‚Was‘ als Input eines solchen Systems, also Anlage und Inhalt einer Daten-Rückmeldung, beeinflusst in diesem Modell das ‚Wie‘ auf der Prozessebene“ (ebd., S. 247). Beide Modelle stimmen in der Annahme überein, dass die Nutzung von Ergebnissen aus Lernstandserhebungen sowohl von der Gestaltung der Rückmeldungen, von der Reflexionsbereitschaft der Lehrkräfte, vom Schul- und Kooperationsklima als auch von externen Unterstützungssystemen abhängt.

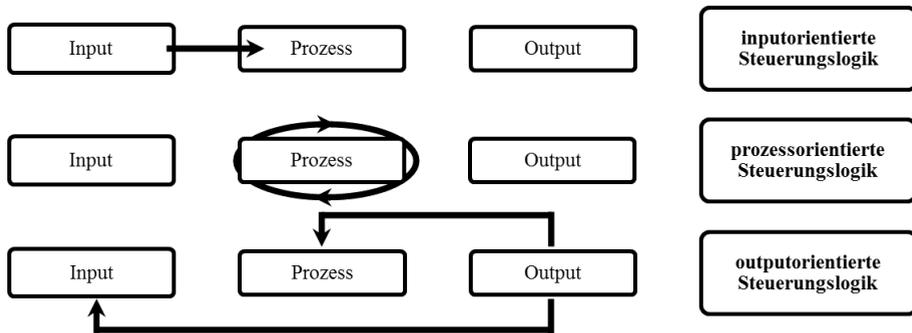
1.3 Rezeptionsstudien

Um die Nutzung von ErgebnISRückmeldungen in Schulen zu untersuchen und um zu prüfen, ob die Ergebnisse aus den Rückmeldungen in den Schulen zu Veränderungen führen, sind zwischenzeitlich bereits verschiedene Rezeptionsstudien durchgeführt worden. Bei diesen Studien steht zunächst die Frage im Vordergrund, welche Bereiche des unterrichtlichen Handelns durch Rückmeldungen beeinflusst werden und welche Kontextfaktoren die Akzeptanz und Nutzung von Lernstandserhebungen bedingen (vgl. Maier 2008). Die Ergebnisse der Rezeptionsstudien machen deutlich, dass die Nutzung von ErgebnISRückmeldungen sehr unterschiedlich verläuft. Während einige Lehrkräfte die Ergebnisse aus den Erhebungen für die Unterrichtsanalyse und -weiterentwicklung nutzen, werden die Ergebnisse in anderen Schulen kaum beachtet (vgl. Kühle/Peek 2007; Nachtigall/Jantowski 2007; Schrader/Helmke 2003). Zudem stellen die Studien heraus, dass die Nutzung von Rückmeldungen aus Lernstandserhebungen durch eine offene Haltung der Lehrkräfte gegenüber externen Evaluationen positiv beeinflusst wird (vgl. Kühle/Peek 2007; Schrader/Helmke 2004). Maier (2008) verweist in diesem Zusammenhang auf den Aspekt der Aufgeschlossenheit der Thüringer Lehrkräfte gegenüber zentralen Tests und der Auseinandersetzung mit den daraus resultierenden Ergebnissen: „Die Kompetenztests in Thüringen werden eher akzeptiert, weniger als Belastung angesehen und haben für die Lehrkräfte einen höheren lerndiagnostischen Hinweischarakter“ (S. 470). In einigen Studien wurden unterschiedliche Rezeptions- und Nutzungstypen beschrieben (vgl. Groß Ophoff/Hosenfeld/Koch 2007; Hartung-Beck 2009; Rossi/Freeman 1993; Stamm 2003). Groß Ophoff, Hosenfeld und Koch (2007) identifizierten beispielsweise aufgrund der Daten aus der Befragung zu VERA 2004 drei Typen von Lehrkräften, die sich unterschiedlich intensiv mit den Daten auseinandersetzten. Der Rezeptionstyp 1 zeichnet sich durch eine intensive positive Auseinandersetzung aus und schätzt die Verständlichkeit und Nützlichkeit der ErgebnISRückmeldungen als hoch ein. Der Rezeptionstyp 2 hin-

gegen ist gekennzeichnet durch eine weniger intensive Auseinandersetzung mit den Rückmeldungen. Außerdem wird hier die Verständlichkeit niedriger als bei Typ 1 eingeschätzt. Der Rezeptionstyp 3 setzt sich intensiv, aber kritisch mit den Ergebnismrückmeldungen auseinander und schätzt die Nützlichkeit als niedrig ein. Diese Rezeptionsprofile unterschieden sich auch signifikant hinsichtlich der durchgeführten Aktionen zur Unterrichtsentwicklung: Der Rezeptionstyp 1 zeichnet sich durch eine aktive Maßnahmenableitung aus, wohingegen der Typ 3 seltener Maßnahmen zur Unterrichtsentwicklung einführt. Der Rezeptionstyp 2 lässt sich hinsichtlich der Maßnahmenableitung zwischen den anderen beiden Typen einordnen.

Erkenntnisse über unterschiedliche Formen der Auseinandersetzung der Lehrkräfte mit den Lernstandserhebungen im Sinne unterschiedlicher auf den Unterricht bezogener Steuerungslogiken liegen bislang in den Studien kaum vor. Dederling (2011) verweist in ihrer systematischen Analyse zur innerschulischen Auseinandersetzung mit Ergebnismrückmeldungen neben unterschiedlichen Rezeptions- und Nutzungstypen auch auf „unterschiedliche Motive der Datenverarbeitung“ (S. 77), welche sie jedoch nicht weiter ausführt. Im Rahmen der Interviewstudie *Folgen zentraler Lernstandserhebungen* (vgl. Diemer/Kuper 2011; Diemer u.a. 2009) wurde erhoben und untersucht, welche Arten von Aktivitäten der Rezeption, der Reflektion, der Maßnahmenableitung und der Evaluation in Schulen im Zusammenhang mit der Nutzung zentraler Lernstandserhebungen realisiert werden. Angelehnt an das Nutzungsmodell nach Helmke (2004) werden diese Tätigkeiten auch in dieser Studie als Phasen dargestellt. Die Ergebnisse aus dem Interviewmaterial zeigen jedoch, dass die Phasen *Rezeption*, *Reflektion*, *Aktion* und *Evaluation* zum Teil unabhängig voneinander ablaufen. So leiten beispielsweise Lehrkräfte Maßnahmen ab, die einerseits zu ihrer Rezeption passen, andererseits jedoch nicht zu den reflektierten Gründen für das Zustandekommen der Ergebnisse. Auch wenn innerhalb der Phase *Reflektion* die Gründe bezüglich des Zustandekommens der Ergebnisse input- und prozessorientiert identifiziert werden, erfolgt die Maßnahmenableitung tendenziell prozessorientiert und ohne Reflexionsbezug. Diese Diskrepanz deutet darauf hin, dass die Ergebnisse aus Lernstandserhebungen nicht ausschließlich als Informationen über den Output wahrgenommen werden: „Die durch Lernstandserhebungen erzeugten empirischen Informationen beziehen sich prima facie nicht zwangsläufig auf die Outputdimension des jeweiligen Systems, sondern können prinzipiell auch als Zustandsbeschreibungen der Input- oder der Prozessdimension verstanden werden“ (Diemer/Kuper 2011, S. 557). Vor diesem Hintergrund lassen sich bezüglich einer steuerungstheoretischen Analyse der Nutzung zentraler Lernstandserhebungen neben einer outputorientierten auch input- und prozessorientierte Steuerungslogiken unterscheiden, welche von den Lehrkräften – in Auseinandersetzung mit den zur Verfügung gestellten Informationen und bezogen auf den Unterricht – realisiert werden (siehe Abbildung 1):

Abbildung 1: Modell zu den unterrichtsbezogenen Steuerungslogiken von Lernstandserhebungen²



Quelle: eigene Darstellung

Eine inputorientierte Steuerungslogik zeichnet sich dadurch aus, dass die Informationen aus Lernstandserhebungen als Lernausgangslage gesehen werden und als solche zu Anfang des Schuljahres in die Unterrichtsplanung eingehen. Lernstandserhebungen stellen hierbei die Grundlage für Lernzielerwartungen dar. Prozessorientierte Steuerungslogik meint, dass die Rückmeldungen als Teil des Unterrichtsprozesses wahrgenommen und die Informationen für die Maßnahmenableitungen weitgehend für die Gestaltung dieses Prozesses genutzt werden. Lernstandserhebungen geben also Orientierungen für die Weiterarbeit im Unterricht. Bei der outputorientierten Steuerungslogik werden die Ergebnisse als abschließende Bewertungen der Leistungen der Schülerinnen und Schüler (Informationen über den Output) mit reflektiertem Rückgriff auf die Eingangsgrößen und die Prozesse betrachtet. Lernstandserhebungen sind dabei ausschließlich Erhebungsinstrumente zur Leistungsüberprüfung. Altrichter (2009) betont in diesem Kontext das Wissen um die unterschiedliche Wahrnehmung der Lernstandserhebungen seitens der Lehrkräfte als Nutzungsvoraussetzung: „Das dahinter liegende Problem besteht darin, dass die *Passung der angebotenen Informationen (und ihre Lokalisierung im Arbeitsprozess)* auf Arbeits- und Denkstile der Lehrpersonen eine Voraussetzung für ihre Nutzung ist“ (S. 219).

2. Fragestellung

Die angeführten Ergebnisse der Interviewstudie *Folgen zentraler Lernstandserhebungen* verdeutlichen, dass es neben der angestrebten und zahlreich diskutierten outputorientierten Steuerungslogik der Rückmeldung von Ergebnissen aus Lernstandserhebungen auch andere Formen der Auseinandersetzung mit Lernstandserhebungen gibt, die

² Dieses Modell wurde im Rahmen der Auswertungen im DFG-Projekt *Folgen zentraler Lernstandserhebungen* (Leitung: Prof. Dr. H. Kuper) von T. Diemer und F. Schulze 2009 entwickelt und in der Qualifikationsarbeit von N. Richter und F. Schulze 2009 erstmals verwendet.

von den Lehrkräften realisiert werden. Bislang gibt es Forschungsarbeiten, in denen verschiedene gegenstandsnahe Nutzungstypologien entwickelt wurden. Dabei spielten die unterschiedlichen Steuerungslogiken kaum eine Rolle. Die vorliegende Untersuchung widmete sich der Frage, in welcher Art und Weise sich Lehrkräfte mit den Ergebnisrückmeldungen aus den Thüringer Kompetenztests auseinandersetzen und welche Steuerungslogiken von den Lehrkräften hierbei im Unterricht realisiert werden. Es sollte überprüft werden, ob sich die in der Interviewstudie *Folgen zentraler Lernstandserhebungen* gefundenen Argumentationslinien auch in der Ergebnisrezeption und -nutzung von Lehrkräften und Schulleitungen zeigen lassen, die mithilfe eines quantitativen Erhebungsinstruments befragt wurden. Diese Untersuchung stellte eine erste Pilotierung eines Fragebogens dar, mit dem unterschiedliche Argumentationsweisen hinsichtlich der unterrichtsbezogenen Steuerungslogiken (input-, prozess- und outputorientiert) abgebildet werden. Ziel der Erhebung war es, weiterführende Informationen darüber zu generieren, wie die Ergebnisrückmeldungen von Lernstandserhebungen in Schulen Verwendung finden.

3. Methode

3.1 Design

Ein wesentlicher Teil der Untersuchung bestand zunächst in der Entwicklung eines standardisierten Fragebogens, der die auf den Unterricht bezogene Steuerungslogik der Lehrkräfte untersucht und der erfasst, als welche Informationsart (input-, prozess- oder outputorientiert) die Lernstandserhebungen und ihre Rückmeldungen wahrgenommen werden. Die Studie diente dazu, im Sinne eines Pretests die Eignung des Untersuchungsinstruments und den reibungslosen Untersuchungsablauf zu testen (vgl. Bortz/Döring 2006). Darüber hinaus konnten gleichzeitig Informationen über die Auseinandersetzung der Lehrkräfte mit Ergebnisrückmeldungen aus zentralen Lernstandserhebungen erfasst und beschrieben werden.

3.2 Stichprobe

Die Erhebung wurde in Thüringen im Schuljahr 2008/2009 durchgeführt. Dieses Bundesland ist aufgrund der dort gegebenen langjährigen Erfahrungen mit Lernstandserhebungen ausgewählt worden. Die *Kompetenztests* werden seit dem Schuljahr 2002/2003 in der dritten Klasse in Deutsch und Mathematik, in der sechsten Klasse in Deutsch, Mathematik und Englisch und in der achten Klasse in Mathematik geschrieben. Diese Tests werden als Diagnoseinstrument eingesetzt, ohne dass eine Verpflichtung zur Notenvergabe besteht (vgl. Diemer/Kuper 2011).

An der Untersuchung nahmen 43 Lehrkräfte teil;³ davon lehrten 13 an Regelschulen, acht an Gesamtschulen und 21 an Gymnasien. In der vorliegenden Stichprobe sind die Lehrkräfte von Gymnasien und Gesamtschulen gemessen an der Gesamtverteilung in Thüringen überrepräsentiert, die von Regelschulen dagegen unterrepräsentiert (vgl. Thüringer Ministerium für Bildung, Wissenschaft und Kultur/Statistikstelle 2010). Das mittlere Alter der Lehrkräfte in der Studie⁴ (MW 49,9 Jahre bei einer Altersspanne zwischen 38 und 61 Jahren) entspricht dem durchschnittlichen Alter der Lehrkräfte in den genannten Thüringer Schultypen von 50,1 Jahren (vgl. ebd.). Von den 43 Lehrkräften sind in dieser Stichprobe 34 Lehrkräfte weiblich und sieben Lehrkräfte männlich.⁵ In der vorliegenden Stichprobe ist der Anteil an weiblichen Lehrkräften etwas höher, spiegelt jedoch die Verteilung der Geschlechter in den Schulen tendenziell wider (vgl. ebd.). Die Unterrichtsfächer der befragten Lehrkräfte sind heterogen.

3.3 Erhebungsinstrument

Der Fragebogen umfasst eine Skala zu den von den Lehrkräften realisierten Steuerungslogiken der Lernstandserhebungen mit drei Subskalen, deren Items mit einer vierstufigen Likert-Skala (1 = stimme voll zu, 2 = stimme eher zu, 3 = stimme eher nicht zu, 4 = stimme gar nicht zu) gemessen wurden. Die unterrichtsbezogene Steuerungslogik kann als Konstrukt bezeichnet werden, das unterschiedliche Orientierungen der Wahrnehmung von Lernstandserhebungen aufweist (input-, prozess- und outputorientierte Steuerungslogik). Die Items beziehen sich an dieser Stelle auf die Wahrnehmung der Lernstandserhebungen und nicht auf die direkte Maßnahmenableitung. Die Formulierung der Items basierte auf dem kategorial ausgewerteten Datenmaterial des Forschungsprojekts *Folgen zentraler Lernstandserhebungen* (vgl. Diemer u.a. 2009). Somit kann von einer inhaltlichen Validität der Items ausgegangen werden. Jedoch zeigt sich aufgrund der mittleren Korrelationen der Subskalen untereinander eine eher geringe diskriminante Validität.

Die in Tabelle 1 aufgeführten Reliabilitäten der Subskalen variieren zwischen $\alpha = .89$ und $\alpha = .92$. Zusammengefasst zeigen die Ergebnisse der Reliabilitätsanalyse gute bis sehr gute Werte für die interne Konsistenz und damit den Grad der Zuverlässigkeit der einzelnen Subskalen. Die ihnen zugrunde liegenden einzelnen Items weisen des Weiteren eine hohe Eigentrennschärfe auf ($r_{ii} > .60$). Sie können demzufolge als prototypisch für die jeweiligen Skalen gelten. Die Überprüfung der Binnenstruktur der Subskalen zu den Formen der Steuerungslogik mithilfe einer Korrelationsanalyse

3 Zur Erfassung der Nutzung von Rückmeldeergebnissen wurden insgesamt 126 Fragebögen verschickt. Davon kamen 43 Fragebögen zurück; die Rücklaufquote beträgt demzufolge 34 Prozent.

4 Acht Lehrkräfte haben dabei keine Angaben zu ihrem Alter gemacht.

5 Zwei Personen haben diese Frage nicht beantwortet.

Tabelle 1: Skalendokumentation zu den unterrichtsbezogenen Steuerungslogiken

Subskalen	Anzahl der Items ¹	Interne Konsistenz der Subskalen (Cronbachs Alpha)	N	Beispielitem
Inputorientierte Steuerungslogik	8	.89	39	„Lernstandserhebungen stellen mir Informationen über die Lernausgangslage in meiner Klasse bereit.“
Prozessorientierte Steuerungslogik	6	.92	42	„Lernstandserhebungen geben mir Rückmeldungen über meinen Unterrichtsprozess.“
Outputorientierte Steuerungslogik	7	.92	38	„Lernstandserhebungen dienen der Zertifizierung von Schülerleistungen.“

¹ Aufgrund mangelnder Trennschärfe und inhaltlicher Differenzen wurden einige Items des Originalfragebogens ausgeschlossen.

Quelle: eigene Berechnungen

zeigt, dass die in den drei verwendeten Subskalen erfassten Steuerungslogiken in einem 2-seitig signifikanten interkorrelativen Zusammenhang stehen (in allen Fällen $p < .01$).

4. Ergebnisse

4.1 Formen der Steuerungslogik von Lernstandserhebungen im Vergleich

Insgesamt ist festzustellen, dass die Items der Subskalen zur input- und outputorientierten Steuerungslogik weder prinzipiell positiv noch negativ beantwortet wurden. Nur die Subskala zur prozessorientierten Steuerungslogik wurde durchweg positiv beantwortet. Es sind Tendenzen sichtbar, welche Aussagen über die Wahrnehmung der Informationen aus Lernstandserhebungen seitens der Lehrkräfte zulassen: Mithilfe der Varianzanalyse mit dem allgemeinen linearen Modell lässt sich in den Ergebnissen zu den Formen der Steuerungslogik zeigen, dass die Lehrkräfte sich mit den Ergebnissen aus Lernstandserhebungen unterschiedlich auseinandersetzen und dabei der prozessorientierten Steuerungslogik (MW 2,37, SD 0,63) tendenziell eher zustimmen.⁶ Auch die Tests der Innersubjekteffekte belegen das Vorhandensein von signifikanten Unterschieden zwischen den Subskalen ($p < .01$). Die Tests der Innersubjektkontraste zeigen, dass es zwischen den Subskalen signifikante Unterschiede auf .01 Niveau gibt: Die inputorientierte Steuerungslogik unterscheidet sich signifikant von der prozessorientierten Steuerungslogik, die prozessorientierte von der outputorientierten

⁶ Die Subskala zur inputorientierten Steuerungslogik der Lernstandserhebungen weist einen Mittelwert von 2,54 auf (SD 0,54); die Subskala zur outputorientierten Steuerungslogik von Informationen aus Ergebnismeldungen hat einen Mittelwert von 2,61 (SD 0,59).

Steuerungslogik. Die Berechnungen im allgemeinen linearen Modell ergaben, dass 21 Prozent der Varianz durch die Formen der Steuerungslogik auf der Stichprobenebene aufgeklärt wurden.⁷ Nach Bortz und Döring (2006) entspricht dies einer mittleren bis großen Effektgröße.

4.2 Gruppenspezifische Unterschiede in der Steuerungslogik von Lernstandserhebungen

Mithilfe des T-Tests bei unabhängigen Stichproben und der Varianzanalyse wurden Unterschiede in den Gruppen Geschlecht, Fächergruppe und Schulart (Gymnasium, Regelschule- und Gesamtschule) hinsichtlich der einzelnen Formen der unterrichtsbezogenen Steuerungslogik untersucht. In den Ergebnissen der Auseinandersetzung mit den Lernstandserhebungen seitens der Lehrkräfte zeigt sich, dass keine Unterschiede hinsichtlich der Steuerungslogik von männlichen und weiblichen Lehrkräften vorhanden sind. Auch gemessen an der Schulart gibt es keine Unterschiede. Dagegen lassen sich Unterschiede bezüglich der inputorientierten Steuerungslogik ($t_{33,88} = -2.54$, $p < .05$, $d = .84$) von Lehrkräften der naturwissenschaftlichen und der sprachlichen Fächer feststellen. Lehrkräfte mit naturwissenschaftlichen Fächern (MW 2,27, SD 0,45) favorisieren im Vergleich zu denen mit sprachlichen Fächern (MW 2,69, SD 0,54) eher die inputorientierte Steuerungslogik. Die Effektgröße von $d = .84$ ist nach Bortz und Döring (2002) als groß zu interpretieren. Lehrkräfte mit diesen Fächern unterscheiden sich auch hinsichtlich der prozessorientierten Steuerungslogik ($t_{33,41} = -2.13$, $p < .001$, $d = .71$). Hierbei bevorzugen Lehrkräfte mit naturwissenschaftlichen Unterrichtsfächern (MW 2,16, SD 0,54) eher die prozessorientierte Steuerungslogik.⁸ Diese Effektgröße ist als mittel einzuschätzen (vgl. Bortz/Döring 2002).

5. Zusammenfassung und Diskussion

Das Ziel dieser Untersuchung war es, die unterrichtsbezogene Steuerungslogik von Lehrkräften dahingehend zu analysieren, ob die Informationen aus Lernstandserhebungen input-, prozess- oder outputorientiert rezipiert werden. Die Einstellungen der Lehrkräfte gegenüber den Lernstandserhebungen werden durch die argumentativ unterschiedliche Auseinandersetzung mit den Erhebungen und ihrer Ergebnismeldung verdeutlicht. Hinsichtlich der Subskalen zu den Formen der Steuerungslogik konnten Unterschiede im Antwortverhalten festgestellt werden: Lehrkräfte tendieren weniger zu den Items der input- und outputorientierten Steuerungslogik und eher zu den prozessorientierten Items und bewerten diese po-

7 Die Voraussetzungen für die Anova im allgemeinen linearen Modell sind erfüllt, da die Sphärizität gegeben ist.

8 Lehrkräfte mit sprachlichen Unterrichtsfächern haben einen Mittelwert von 2,57 (SD 0,6).

sitiv. Die an der Untersuchung beteiligten Lehrkräfte sehen die Informationen aus den Lernstandserhebungen folglich eher als prozessorientierte Information, welche der Zustandsbeschreibung der aktuellen Unterrichtsprozesse dient und welche die Lehrkräfte für die Weiterarbeit im Unterricht nutzen können, und weniger als leistungszertifizierende Information. Auch die Tests der Innersubjekteffekte und der Innersubjektkontraste belegen diesen signifikanten Unterschied. Die gezeigten Unterschiede in der input- und prozessorientierten Steuerungslogik der Ergebnisse aus Lernstandserhebungen von Lehrkräften aus den naturwissenschaftlichen und sprachlichen Fächern weisen daraufhin, dass es Unterschiede in der Auseinandersetzung abhängig von den unterrichteten Fächern gibt. Zu erklären sind diese Unterschiede eventuell mit der erfahrenen naturwissenschaftlichen oder geisteswissenschaftlichen Prägung. Die Gemeinsamkeiten der Personen dieser Fächergruppen hinsichtlich der outputorientierten Steuerungslogik lassen sich durch die von Bildungsadministration und -forschung proklamierte Funktion von Lernstandserhebungen erklären. Diese Unterschiede in den Formen der Steuerungslogik lassen sich jedoch nicht hinsichtlich des Geschlechts und der Schulart zeigen, wobei dies auch auf die unterschiedliche Anzahl der Gruppen in dieser Stichprobe zurückzuführen ist. Die Studie verweist so insgesamt auf Variationen in der auf den Unterricht bezogenen Steuerungslogik, die über die diskutierte outputorientierte Steuerung hinausgeht. Dabei stellt die prozessorientierte und nicht die outputorientierte Steuerungslogik die bei den befragten Lehrkräften verbreitetste realisierte Form der Auseinandersetzung mit den Ergebnismeldungen dar. Dies steht im Einklang mit den ersten qualitativen Ergebnissen der Interviewstudie *Folgen zentraler Lernstandserhebungen* (vgl. Diemer/Kuper 2011). Gründe für diesen Befund könnten zum einen darin bestehen, dass die Lernstandserhebungen in Thüringen nicht bewertet werden und somit kein innerschulisches Instrument der Leistungsüberprüfung darstellen, sondern eher ein Diagnoseinstrument sind. Zum anderen könnte die vergleichsweise geringe Ausprägung einer outputorientierten Steuerungslogik dadurch erklärt werden, dass das Prinzip der Outputorientierung auf Grund der Neuheit der Innovation bislang erst ansatzweise implementiert ist. Inwiefern die unterschiedliche Wahrnehmung der Lernstandserhebungen auch Auswirkungen auf unterschiedliche Maßnahmenplanung und tatsächlich umgesetzte Maßnahmen hat, wurde nicht untersucht, kann jedoch für weiterführende Forschungen interessant sein.

Des Weiteren zielte diese Untersuchung darauf ab, die Datenqualität des entwickelten Fragebogens zu überprüfen, um eine erneute Durchführung bei einer repräsentativen Stichprobe vorzubereiten. Die Subskalen zur input-, prozess- und outputorientierten Steuerungslogik der Ergebnismeldungen weisen eine hohe interne Konsistenz auf. Jedoch korrelieren die Subskalen untereinander stark, wodurch die angestrebte Differenzierung der unterschiedlichen Formen der Steuerungslogik nicht erreicht wurde. Die Gründe hierfür können in der starken Ähnlichkeit der Itemformulierungen liegen. So wurden Items eingesetzt, die zwar ähnlich klingen, jedoch

unterschiedliche Dimensionen der Steuerung und Nutzung abfragen. Voraussetzung für die korrekte Beantwortung dieser Items ist das differenzierte Lesen, Wahrnehmen und Beantworten dieser Fragen durch die Lehrkräfte. Davon kann jedoch nicht ausgegangen werden, sodass die einzelnen Nuancen in der Itemformulierung von den befragten Lehrkräften wahrscheinlich nicht erkannt wurden. Optimal wäre es, den Fragebogen mit den Lernstandserhebungen auszugeben, sodass Lehrkräfte während der Bearbeitungszeit der Schülerinnen und Schüler den Fragebogen beantworten können.

Als Desiderat bleibt die weiterführende Untersuchung der Fragestellung anhand einer repräsentativen Stichprobe. Ziel sollte dabei die Normierung des Tests und die Identifikation von unterrichtsbezogenen Steuerungslogiken sein. Denn nur wenn bekannt ist, wie sich Lehrkräfte mit den Ergebnissen aus Lernstandserhebungen argumentativ auseinandersetzen, kann auch Einfluss auf deren Nutzung genommen werden.

Literatur

- Altrichter, H. (2009): Datenfeedback und Unterrichtsentwicklung – Probleme eines Kernelements im „neuen Steuerungsmodell“ für das Schulwesen. In: Dicke, J.N./Ziegler, H./Böttcher, W. (Hrsg.): Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis. Münster u.a.: Waxmann, S. 211-226.
- Altrichter, H. (2010): Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In: Altrichter, H./Maag Merki, K. (Hrsg.): Handbuch neue Steuerung im Schulsystem. Wiesbaden: VS, S. 219-254.
- Bortz, J./Döring, N. (³2002): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer.
- Bortz, J./Döring, N. (⁴2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin: Springer.
- Bonsen, M./von der Gathen, J. (2004): Schulentwicklung und Testdaten. Die innerschulische Verarbeitung von Leistungsrückmeldungen. In: Holtappels, H.G./Klemm, K./Pfeiffer, H./Rolff H.-G./Schulz-Zander, R. (Hrsg.): Jahrbuch der Schulentwicklung, Band 13. Weinheim u.a.: Juventa, S. 225-252.
- Dedering, K. (2011): Hat Feedback eine positive Wirkung? Zur Verarbeitung extern erhobener Leistungsdaten in Schulen. In: Unterrichtswissenschaft 39, H. 1, S. 63-83.
- Diemer, T./Kuper, H. (2011): Formen innerschulischer Steuerung mittels zentraler Lernstandserhebungen. In: Zeitschrift für Pädagogik 56, H. 4, S. 554-571.
- Diemer, T./Rucht, S./Schulze, F./Kuper, H. (2009): Zur innerschulischen Nutzung zentraler Lernstandserhebungen – Deskriptive kategoriale Grundauswertungen problemzentrierter Interviews mit Lehrer/innen und Schulleitungen in vier Sekundarschulen in Berlin und Thüringen. URL: http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/weiterbildung_bildungsmanagement/projekte/folgen_zentraler_lehr/index.html; Zugriffsdatum: 05.04.2011.
- Groß Ophoff, J./Hosenfeld, I./Koch, U. (2007): Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. In: Hosenfeld, I./Groß Ophoff, J. (Hrsg.): Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht. Landau: Verlag Empirische Pädagogik, S. 411-427.

- Groß Ophoff, J./Koch, U./Hosenfeld, I./Helmke, A. (2006): Ergebnissrückmeldungen und ihre Rezeption im Projekt VERA. In: Kuper, H./Schneewind, J. (Hrsg.): Rückmeldung und Rezeption von Forschungsergebnissen – Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster u.a.: Waxmann, S. 19-40.
- Hartung-Beck, V. (2009): Schulische Organisationsentwicklung und Professionalisierung. Folgen von Lernstandserhebungen an Gesamtschulen. Wiesbaden: VS.
- Hartung-Beck, V./Diemer, T. (2009): Sensemaking durch Outputorientierung. Erfahrungen mit der Nutzung von Lernstandserhebungen in Schulen. In: Göhlich, M./Weber, S.M./Wolff, S. (Hrsg.): Organisation und Erfahrung. Wiesbaden: VS, S. 215-225.
- Helmke, A. (2004): Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule. In: Das Seminar, H. 2, S. 90-112.
- Kühle, B./Peek, R. (2007): Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnissrückmeldungen in Schulen. In: Empirische Pädagogik 21, H. 4, S. 428-447.
- Maier, U. (2008): Vergleichsarbeiten im Vergleich – Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. In: Zeitschrift für Erziehungswissenschaft 11, H. 3, S. 453-474.
- Nachtigall, C./Jantowski, A. (2007): Die Thüringer Kompetenztests unter besonderer Berücksichtigung der Evaluationsergebnisse zum Rezeptionsverhalten. In: Empirische Pädagogik 21, H. 4, S. 401-410.
- Netzwerk Empiriegestützte Schulentwicklung (2006): Zentrale standardisierte Lernstandserhebungen. Positionspapier zur 5. EMSE-Tagung, Berlin 08.12.2006. URL: http://www.isq-bb.de/uploads/media/emse-positionspapier-07-07_04.pdf; Zugriffsdatum: 01.02.2012.
- Oelkers, J./Reusser, K. (2008): Qualität entwickeln, Standards sichern – mit Differenz umgehen. Bonn: BMBF.
- Peek, R./Dobbelstein, P. (2006): Benchmarks als Input für die Schulentwicklung – Das Beispiel der Lernstandserhebungen in Nordrhein-Westfalen. In: Kuper, H./Schneewind, J. (Hrsg.): Rückmeldung und Rezeption von Forschungsergebnissen – Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster u.a.: Waxmann, S. 41-58.
- Richter, N./Schulze, F. (2009): Folgen zentraler Lernstandserhebungen – Entwicklung eines standardisierten Fragebogens zur Erfassung der Nutzung von Ergebnissrückmeldungen. Unveröffentlichte Qualifikationsarbeit. Freie Universität Berlin.
- Rolff, H.-G. (2002): Rückmeldung und Nutzung der Ergebnisse von großflächigen Leistungsuntersuchungen. In: Rolff, H.-G./Holtappels, H.G./Klemm, K./Pfeiffer, H./Schulz-Zander, R. (Hrsg.): Jahrbuch der Schulentwicklung, Band 12. Weinheim u.a.: Juventa, S. 75-98.
- Rossi, P.H./Freeman, H.E. (1993): Evaluation: A Systematic Approach. Newbury Park, CA: Sage.
- Schneewind, J./Kuper, H. (2009): Rückmeldeformate und Verwendungsmöglichkeiten der Ergebnisse aus zentralen Lernstandserhebungen. In: Bohl, T./Kiper, H. (Hrsg.): Lernen aus Evaluationsergebnissen. Verbesserungen planen und implementieren. Bad Heilbrunn: Klinkhardt, S. 113-129.
- Schrader, F.-W./Helmke, A. (2003): Evaluation und was danach? Ergebnisse der Schulleiterbefragung im Rahmen der Rezeptionsstudie WALZER. In: Schweizerische Zeitschrift für Bildungswissenschaft 25, H. 1, S. 79-116.
- Schrader, F.-W./Helmke, A. (2004): Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER: Ergebnisse der Lehrerbefragung. In: Empirische Pädagogik 18, H. 1, S. 140-161.
- Stamm, M. (2003): Evaluation und ihre Folgen für die Bildung: Eine unterschätzte pädagogische Herausforderung. Münster u.a.: Waxmann.

Thüringer Ministerium für Bildung, Wissenschaft und Kultur/Statistikstelle (2010): Personal.
URL: <http://www.schulstatistik-thueringen.de>; Zugriffsdatum: 08.12.2010.

Franziska Schulze, M.A. Erziehungswissenschaft, geb. 1984, wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaft, Abteilung Schulpädagogik/Schul- und Unterrichtsforschung der Westfälischen Wilhelms-Universität Münster.

Anschrift: Bispinghof 5/6, 48143 Münster
E-Mail: Franziska.Schulze@uni-muenster.de

WAXMANN

Wirksame Schulleitung braucht Leadership. Sie liegt in der personalen und mentalen Kompetenz einer Führungsperson, der Entwicklung im eigenen Bereich Richtung zu geben, ein starkes Engagement der Einzelnen auszulösen, die organisationale Leistungsfähigkeit zu steigern und Charakterstärke zu zeigen. Die Autoren der Studie legen den Fokus auf diese vier bedeutsamen Schlüsselemente von Führungseigenschaften, um herauszufinden, ob sie dazu beitragen, die Bildungsprozesse der Schülerinnen und Schüler zu verbessern.



Michael Schratz
Martin Hartmann
Wilfried Schley

Schule wirksam leiten

Analyse innovativer
Führung in der Praxis

2010, 208 S., br. 34,90 €,
ISBN 978-3-8309-2313-8

Waxmann
order@waxmann.com
www.waxmann.com

Fabian Dietrich/Maike Lambrecht

Menschen arbeiten mit Menschen

Schulinspektion und die Hoffnung auf den zwanglosen Zwang der „besseren Evidenz“

Zusammenfassung

Im Konzept der evidenzbasierten Steuerung drückt sich einmal mehr die Hoffnung auf eine intentionale Gestaltbarkeit des Schulsystems aus. Die dezentrale Steuerung über empirische Evidenzen verspricht eine „Schulentwicklung durch Einsicht“ und damit eine Rationalisierung des Steuerungsprozesses: Inspektoren und Inspektorinnen sowie schulische Akteure sollen sich „auf Augenhöhe“ begegnen, und Schulentwicklung soll durch die rationale Auseinandersetzung mit „objektiven Fakten“ vorangetrieben werden. Fraglich ist jedoch, welcher Logik die evidenzbasierte Steuerung in der Praxis tatsächlich folgt. Am Beispiel der Schulinspektion wird gezeigt, dass der Evidenzbegriff auf der Ebene der Steuerungsprogrammatisierung sehr unterschiedlich ausgestaltet wird. Aus rekonstruktiver Perspektive zeigt sich, dass die evidenzbasierte Steuerung maßgeblich davon abhängt, inwieweit die durch die Schulinspektion erzeugten Evidenzen glaubhaft gemacht werden können. Damit ist nicht die Generierung, sondern die Vermittlung von Evidenzen die zentrale Strukturproblematik der Schulinspektionspraxis.

Schlüsselwörter: Evidenz, Steuerung, Schulinspektion, Rekonstruktion

People Work with People

School Inspection and the Hope of the Unforced Force of “Better Evidence”

Abstract

The concept of evidence-based steering expresses once more the hope of an intentional structuring of the school system. Decentralised steering via empirical evidence promises “school development through reason” and thereby a rationalisation of the process of steering. Inspectors and actors at school shall meet each other “at eye level”, and school development shall be promoted by a rational dispute of the “objective facts”. However, the question is which logical structure this conception follows. With school inspection as example, it is demonstrated that the notion of evidence is formulated very differently at the level of steering objectives. From a reconstructive perspective it becomes appar-

ent that evidence-based steering crucially depends on the extent to which evidence, generated via school inspection, can be made credible. Consequently, not the generation of evidence but rather the conveyance of evidence is the central structural problem of the practice of school inspection.

Keywords: evidence, steering, school inspection, reconstruction

1. Steuerungshoffnungen

In der Programmatik einer *evidenzbasierten Steuerung* (vgl. van Ackeren u.a. 2011, S. 171f.) reproduziert sich in neuer Gestalt die Hoffnung auf eine gezielte Gestaltbarkeit des schulischen Bildungssystems. Allerdings handelt es sich dabei nicht einfach um eine Wiederbelebung eines alten Steuerungsoptimismus. Vielmehr kann die Konjunktur der evidenzbasierten Steuerung als Reaktion auf die in den letzten Jahrzehnten regelmäßig wiederkehrende Enttäuschung des sich in den 1960er- und 1970er-Jahren um den Begriff der Planung und später um den Begriff der Steuerung zentrierenden Machbarkeitsoptimismus (vgl. Kussau/Brüsemeister 2007, S. 16ff.) gelesen werden: Nicht mehr mittels einer zentralistischen Planung oder einer hierarchischen Steuerung soll das schulische Bildungssystem reformiert bzw. „optimiert“ werden. Derartige Vorstellungen scheinen empirisch wie theoretisch desavouiert. Stattdessen wird auf eine *dezentrale Steuerungswirkung empirischer Evidenzen* gesetzt. An die Stelle von Verordnungen und Erlassen treten Evidenzen, welche als „empirisch hinreichend belegte Erkenntnisse als ‚Steuerungswissen‘ für Politik und Praxis“ (van Ackeren u.a. 2011, S. 170) zum zentralen Medium für die Entwicklung des Schulsystems werden sollen.

Vor diesem Hintergrund wird erklärlich, dass sich die Reformbemühungen der letzten Jahre vor allem auf ein empirisches Monitoring des Schulsystems richten (vgl. KMK 2006). Ein wichtiges Element dieser Bestrebung stellen die in den letzten Jahren zumindest zeitweilig in allen Bundesländern etablierten Schulinspektionen¹ dar. Sie dienen der Erzeugung jener Evidenzen, welche eine „*Schulentwicklung durch Einsicht*“ (Böttger-Beer/Koch 2008, S. 254) bewirken und damit eine rationale Gestaltung des Schulsystems insgesamt ermöglichen sollen. Diese Hoffnung, über die sich die aufwändige Implementation der Schulinspektion erst legitimiert, erscheint vordergründig ausgesprochen plausibel – wer kann schon gegen eine „Schulentwicklung durch Einsicht“ sein? Fraglich ist jedoch, wie die im Rahmen der Schulinspektion generierten Daten im Sinne einer Schulentwicklung durch Einsicht steuerndes Potenzial entfalten sollen.

1 In den Bundesländern wird das Instrument teilweise als „Schulinspektion“, zum Teil aber auch als „externe Evaluation“ bezeichnet. Im Folgenden wird um der Kürze willen nur die Bezeichnung „Schulinspektion“ verwendet.

Die Forschung zur Schulinspektion gibt hier bisher wenig Aufschluss, weil in der überwiegenden Zahl der Untersuchungsansätze die Vorannahmen der bildungspolitischen Steuerungsprogrammatisik übernommen werden (vgl. Lambrecht/Rürup, im Druck). Studien zur Schulinspektion bewegen sich in einem engen Themenspektrum: Untersucht wird in der Regel, a) ob das Instrument von unterschiedlichen Akteuren des Schulsystems akzeptiert wird bzw. b) welche Wirkungen das Instrument in Bezug auf Schulentwicklung bzw. Schülerleistungen zeigt (vgl. Böttcher/Keune 2010, S. 153; Husfeldt 2011, S. 272-277). In diesen Forschungsfoki – Akzeptanz einerseits, Wirkungen andererseits – bildet sich deutlich die bildungspolitische Steuerungslogik einer Schulentwicklung über Einsicht durch Evidenzen ab. Die Studien übernehmen somit auch die Vorannahme einer evidenzbasierten Steuerung, gehen also davon aus, dass der postulierte Wirkmechanismus auch der tatsächlich wirksame ist. Die Frage nach der Evidenzgenerierung innerhalb einer evidenzbasierten Schulsystemsteuerung wird in erster Linie in Bezug auf eine adäquate Verarbeitung von vielfältig erzeugten Evidenzen seitens schulischer oder auch administrativer Akteure erörtert (vgl. z.B. Bos/Voss 2008; Maier 2008; Dederling 2009; Posch 2009). Nicht gefragt wird dagegen, inwieweit die aktuelle Schulsystemsteuerung tatsächlich in erster Linie über die Erzeugung von Evidenzen und deren rationale Verarbeitung funktioniert. Die Rekonstruktion der impliziten Steuerungslogik bzw. der Steuerungspraxis jenseits programmatischer Verlautbarungen stellt bisher eine Ausnahme dar (vgl. z.B. Baltruschat 2010).

Eine solche empirische Rekonstruktion von Steuerung soll im folgenden Beitrag versucht werden. Dazu wird in einem ersten Schritt das bildungspolitische Programm einer evidenzbasierten Schulsystemsteuerung genauer beleuchtet – und damit auch Vorstellungen darüber, wie Schulinspektionen als Steuerungsinstrumente funktionieren *sollen*. Diesen programmatischen Steuerungsvorstellungen wird dann in einem zweiten Schritt die Analyse eines Interviews mit einem Schulinspektor gegenübergestellt. Es wird also die Darstellung der Funktionsweise der Schulinspektion auf Ebene der Steuerungsprogrammatisik mit der Darstellung der Steuerungspraxis verglichen. Spätestens hier erweist sich, inwieweit der neuerliche, mit der Programmatisik einer evidenzbasierten Steuerung einhergehende Gestaltungsoptimismus berechtigt erscheint. Die Interviewanalyse entstand im Rahmen des BMBF-Verbundprojekts „Schulinspektion als Steuerungsimpuls zur Schulentwicklung und seine Realisierungsbedingungen auf einzelschulischer Ebene“ (2010-2013)².

2 Das Projekt untersucht die Schulinspektionskonzepte in vier Bundesländern über ein multiperspektivisches und triangulatives Untersuchungsdesign. Neben einer quantitativen Schulleitungsuntersuchung umfasst das Verbundvorhaben eine Interviewstudie, bei der zentrale, regionale und schulische Akteure der Schulinspektion befragt werden. Das Interviewmaterial wird sowohl qualitativ-kodierend als auch rekonstruktiv mithilfe des Verfahrens der Objektiven Hermeneutik analysiert; der empirische Teil dieses Beitrags (Kapitel 3) stellt einen Auszug aus diesen objektiv-hermeneutischen Rekonstruktionen vor. Vgl. auch URL: <http://www.empirische-bildungsforschung-bmbf.de/zeigen.html?seite=8838>; Zugriffsdatum: 13.02.2012.

2. Evidenzen

Im Folgenden geht es darum, genauer zu bestimmen, was unter „Evidenz“ im Rahmen der Schulinspektionsprogrammatiken verstanden wird und welche Vorstellungen sich in Bezug auf die Generierung von Evidenzen im Schulinspektionsprozess daraus ergeben. Dazu erscheint es zunächst sinnvoll, sich den *idealtypischen Ablauf einer Schulinspektion* zu vergegenwärtigen (vgl. z.B. Döbert/Rürup/Dedering 2008): Kern jedes Schulinspektionsprozesses ist ein Schulbesuch durch ein externes Expertenteam, der sich in der Regel über mehrere Tage erstreckt. Während eines solchen Schulbesuchs sammeln die Inspektorinnen und Inspektoren eine Fülle von Daten und Informationen über die jeweilige Einzelschule. Dabei kommt eine Reihe von Instrumenten zum Einsatz; in erster Linie sind dies Interviews mit unterschiedlichen schulischen Akteuren sowie Unterrichtsbeobachtungen. Im Vorfeld werden häufig auch standardisierte Befragungen und Dokumentenanalysen vorgenommen. Die Informationssammlung, d.h. die Generierung von Evidenzen, richtet sich sowohl auf Prozesse der Schulorganisation und -verwaltung als auch auf die Unterrichtspraxis. Sie basiert auf einem schulübergreifenden Qualitätsmodell guter Schule, das in Kriterienkatalogen, sogenannten „Qualitätsrahmen“, kodifiziert ist. Ergebnis der Erhebung ist ein Inspektionsbericht als detaillierte Bewertung der Schule, welcher im Sinne der evidenzbasierten Steuerung die Grundlage für an die Inspektion anschließende Schulentwicklungsprozesse bildet und damit die steuernden Evidenzen liefern soll.

Innerhalb dieses idealtypischen Ablaufes lassen sich bundeslandspezifische Ausgestaltungen des jeweiligen Schulinspektionsverfahrens feststellen (vgl. Rürup 2008, S. 470-472). Dabei variiert auch das Evidenzkonzept, d.h. das, was unter Evidenzen verstanden wird. Evidenzen werden zwar häufig mit wissenschaftlich gewonnenen Erkenntnissen assoziiert; tatsächlich ist das Verhältnis von Schulinspektion bzw. – allgemeiner formuliert – von Evaluation und Wissenschaft jedoch ambivalent (vgl. z.B. Lüders 2006). Im Evaluationskontext scheint sich diesbezüglich die *Sprachregelung einer „systematischen“ Wissensgenerierung* durchzusetzen. So schreiben van Ackeren u.a. bezogen auf die Generierung von Evidenzen: „Letztendlich geht es um eine bestimmte (wissenschaftliche) Methode der Wissensgenerierung, d.h. um systematisches, empirisch gewonnenes und in seinen Voraussetzungen und seiner Gewinnung transparentes, objektiviertes Wissen“ (van Ackeren u.a. 2011, S. 172). Evaluationswissen soll demzufolge systematisch und empirisch – im Fall der Schulinspektion beispielsweise vor Ort – erhoben werden sowie transparent und objektivierbar sein. Diese Beschreibungen weisen zwar in die Richtung einer wissenschaftlichen Erzeugung von Evidenzen. Eine eindeutige Positionierung wird jedoch im Zitat vermieden: Der Begriff „wissenschaftlich“ wird lediglich eingeklammert verwendet und somit relativiert. Auch speziell im Schulinspektionskontext finden sich unterschiedliche konzeptionelle Bezugnahmen auf Wissenschaftlichkeit: In manchen

Bundesländern werden Schulinspektionen deutlich als „Verwissenschaftlichung“ der Schulsystemsteuerung aufgefasst (vgl. Rürup 2008, S. 470); andere Schulinspektionskonzepte grenzen sich z.T. dezidiert von Wissenschaftskonzepten ab. Rürup unterscheidet auf der Basis der Analyse der in einzelnen Bundesländern etablierten Schulinspektionen zwischen eher „wahrheits-“ und eher „angemessenheitsorientierten“ Inspektionsverfahren (ebd., S. 470ff.).

Im Fall der „wahrheitsorientierten“ *Schulinspektionsverfahren* wird die evaluative Urteilsbildung in erster Linie über die Standardisierung des Verfahrens legitimiert; die Bewertung soll hier auf Daten basieren, welche mit Hilfe einer Gleichförmigkeit und Geschlossenheit der methodisch kontrollierten Erhebungsverfahren erzeugt werden. Evidenz wird hier gleichgesetzt mit wissenschaftlichen Formen der Erkenntnisgewinnung; die Schulinspektion wird insbesondere an Verfahren der quantifizierenden empirischen Sozialforschung angelehnt. So heißt es beispielsweise in den offiziellen Dokumenten zur externen Schulevaluation in Mecklenburg-Vorpommern: „Evaluation setzt Fakten an die Stelle von Vermutungen und Mutmaßungen. Es entspricht den Prinzipien der Fairness und Transparenz, dass Evaluation sich in hohem Maße der in der empirischen und in der sozialwissenschaftlichen Forschung üblichen Verfahren bedient“ (LISA Mecklenburg-Vorpommern 2006, S. 6). Eine gute Schulinspektion misst Schulen in dieser Perspektive anhand eines externen kriterialen Bezugsrahmens und liefert so objektive und valide Daten über Schulen.

Im Fall der „angemessenheitsorientierten“ *Schulinspektionsverfahren* werden die evaluativen Schulbewertungen dagegen über die Professionalität der Schulinspektorinnen und Schulinspektoren legitimiert. Hier stehen nicht die Standardisierung des Verfahrens und die Quantifizierung schulischer Prozesse im Vordergrund; vielmehr soll die Schulinspektion eine zutreffende Deutung der Schule liefern und damit eher einer einzelfallbezogenen Bezugsnorm gerecht werden. Dieser Grundsatz wird am deutlichsten für das Bremer Schulinspektionsverfahren formuliert: „Die externe Evaluation soll ‚erzählen‘. Eine ‚gute Schule‘ ist mehr als das, was Vergleichsarbeiten, LAU-Daten oder quantitativ vorgehende Inspektionsverfahren erfassen, ‚zählen‘ können“ (Dr. Otto Seydel Institut für Schulentwicklung o.J., S. 8).

Die Unterscheidung zwischen „Wahrheits-“ und „Angemessenheitsorientierung“ ist jedoch missverständlich, wenn die Kategorien als sich widersprechend angelegt werden. *Letztlich verweisen die unterschiedlichen Konzeptionen auf differierende Relevanzvorstellungen, aus denen sich bezüglich der zu erzeugenden Evidenzen unterschiedliche Qualitätsvorstellungen ableiten:* Einmal geht es eher um eine schulübergreifende, durch Standardisierung erzeugte Vergleichbarkeit der Daten, einmal eher um eine Ausrichtung an den einzelschulischen Besonderheiten. In beiden Fällen basiert Schulinspektion jedoch auf der Idee, die generierten Evidenzen könnten im Sinne des zwanglosen Zwangs des besseren – empirischen – Arguments Einsicht und darüber

vermittelt Steuerungswirkung entfalten – anstelle einer hierarchischen Order, bei der die Einsicht in die Sinnhaftigkeit des Tuns zweitrangig ist.

Analog zu der unterschiedlichen Bezugnahme auf Evidenz im Rahmen der Programmatik verschiedener Schulinspektionskonzepte wird auch die Rolle der Schulinspektorinnen und Schulinspektoren unterschiedlich ausgedeutet. Im Fall der „*wahrheitsorientierten*“ Verfahren werden sie als distanzierte Beobachterinnen und Beobachter konzipiert. Sie sind dann, wie im Folgenden für die Niedersächsische Schulinspektion formuliert, „neutrale Fachleute [...], die in einem weitgehend standardisierten Verfahren und anhand transparenter Kriterien Analysen und Bewertungen vornehmen“ (Lohmann/Reißmann 2007, S. 20). Im Fall der „*angemessenheitsorientierten*“ Verfahren sollen sie eher als „teilnehmende Deuterinnen und Deuter“ der Situation der jeweiligen Einzelschule agieren. Daraus folgt auch, dass ihnen in der Regel größere Freiheiten in Bezug auf die Beurteilung schulischer Prozesse zuerkannt werden, d.h. die Erhebungsinstrumente, z.B. Unterrichtsbeobachtungsbögen, sind offener gestaltet als in „*wahrheitsorientierten*“ Verfahren. Dies soll eine schulangemessene Bewertung ermöglichen, wie im folgenden Zitat für die baden-württembergische Fremdevaluation zum Ausdruck kommt: „Abschließend sei darauf hingewiesen, dass dieses Bewertungssystem erlaubt, Schulen mit ganz unterschiedlichen Wegen und Maßnahmen eine ‚gute Praxis‘ je Merkmal zu bescheinigen, da die ‚Funktionalität‘ bzw. der ‚Nutzen‘ an der Schule im Mittelpunkt der Bewertung steht“ (LS Stuttgart 2009, S. 53).

Ungeachtet dieser differierenden Rollenverständnisse verweisen alle Programmatiken auf die Vorstellung einer „Schulentwicklung durch Einsicht“ und damit auf eine „*rationale Beziehungsstruktur*“. Eine solche rationale Beziehung wird zumindest programmatisch immer wieder durch die Verwendung der Formulierung „*auf Augenhöhe*“ beschworen. So formulierte der damalige niedersächsische Kultusminister Busemann anlässlich der Etablierung der dortigen Schulinspektion, diese solle „partnerschaftlich auf Augenhöhe erfolgen“ (Niedersächsisches Kultusministerium 2005). Baltruschat interpretiert den Ausdruck „auf Augenhöhe“ in Bezug auf die bayerische externe Schulevaluation als programmatisches Bekenntnis zu einem gleichberechtigten und freiwilligen Inspektionsprozess:

„Auf Augenhöhe“ bedeutet, dass sich zwei Partner gegenüber stehen, die sich gegenseitig als ebenbürtig betrachten, sowohl in Hinblick auf die Kompetenzen, die sie sich zuerkennen, als auch auf die symmetrisch verteilten Gestaltungsmöglichkeiten ihres Miteinanders. Eine derartige Begegnung impliziert darüber hinaus ganz selbstverständlich die selbstbestimmte, freiwillige Zustimmung der beiden einander anblickenden Parteien“ (Baltruschat 2010, S. 249).

Ein solches Bekenntnis zu einer Beziehung „auf Augenhöhe“ wird jedoch nur dann nötig, wenn es sich dabei eigentlich um ein hierarchisch strukturiertes Verhältnis handelt oder zumindest eine solche Deutung im Raum steht.

3. Menschen arbeiten mit Menschen

Bis zu dieser Stelle wurde herausgearbeitet, wie sich die mit der evidenzbasierten Steuerung einhergehende Gestaltungshoffnung programmatisch begründet und wie die Etablierung der Schulinspektionen legitimiert wird. Damit ist jedoch noch nichts über die Umsetzung dieser Programmatik gesagt. Offen ist, welches Verständnis von Evidenz in der Schulinspektionspraxis zum Tragen kommt und wie die herausgearbeitete Vorstellung einer „Steuerung durch Einsicht“ umgesetzt wird. Zur empirischen Klärung dieser Fragen soll die folgende objektiv-hermeneutische Rekonstruktion (vgl. Oevermann 2000; Wernet 2006) der Adaption der Programmatik durch einen Schulinspektor beitragen. Die Rekonstruktion richtet sich auf die Konzeption von Schulinspektion, welche aus den Praxiserfahrungen des Schulinspektors hervorgeht und in seiner Praxis handlungsleitend ist. Diese Konzeption, so die methodologische Annahme, drückt sich – teils unmittelbar offensichtlich, teils in unauffälligen Formulierungsnuancen – auch im Sprechen über die Schulinspektion aus.

Die Grundlage der Rekonstruktion bilden zwei Auszüge aus einem Interview mit einem Schulinspektor, in denen es um sein berufliches Selbstverständnis geht. Ausgangspunkt der ausgewählten Passage ist eine Interviewfrage danach, was für den Interviewten ein guter Schulinspektor ist. Auf diese Frage antwortet der Inspektor mit der Darstellung seines Idealbildes von einer Inspektion:

Ein guter Schulinspektor ist, dem es gelingt der Schule zu verdeutlichen, dass sie mit einem festen klar gegliederten Instrumentarium kommt, mit einer klaren Aufgabenstellung kommt und Zielsetzung kommt.

Für unseren Zusammenhang erscheinen an dieser Antwort verschiedene Aspekte bemerkenswert: Unmittelbar ins Auge springt der mehrfache Verweis auf die *Eindeutigkeit* der Interaktion, welche sich im Kontext der Schulinspektion entfaltet: Diese ist idealerweise durch *ein klar gegliedertes Instrumentarium* vorstrukturiert; es liegen eine wiederum *klare Aufgabenstellung* und *Zielsetzung* vor. Der Darstellung zufolge ist die im Rahmen des Inspektionsbesuchs stattfindende Interaktion also durch ein hohes Maß an Vorbestimmtheit und Transparenz geprägt. Entsprechend treten die Akteure allein in ihren Rollen in Erscheinung. In diese Richtung weist auch, dass der Sprecher nicht von den relevanten Akteuren, sondern von den Institutionen spricht: So kommt nicht der Schulinspektor in die Schule, sondern *sie*, also die Schulinspektion. Genau genommen beantwortet der Sprecher nicht die Frage nach einem idealen Inspektor, sondern die nach einer idealen Inspektion. Diese ver-

sucht, dem Wortlaut zufolge, das dargestellte Idealbild der Inspektion auch nicht den schulischen Akteuren zu verdeutlichen, also beispielsweise den Lehrerinnen und Lehrern bzw. der Schulleiterin oder dem Schulleiter, sondern der *Schule*. Wir müssen nicht wissen, in welchem Bundesland der Schulinspektor tätig ist, um von dieser Darstellung aus auf eine hohe Standardisiertheit des Inspektionsverfahrens zu schließen.

Dass sich diese Standardisiertheit tatsächlich ungebrochen in der Praxis realisiert, ist damit jedoch nicht gesagt. Unauffällig deutet sich in der Formulierung des Idealbildes jenseits der Betonung von Transparenz und Entschiedenheit bereits eine Gefährdung derselben an: So ist nicht die Rede davon, dass die Inspektion klar und transparent sein soll. Vielmehr gehe es darum, *dass es gelingen müsse, dieses zu vermitteln*. Darin deutet sich die Möglichkeit des Misslingens an, die sich in der Vermittlungsnotwendigkeit begründet. Demnach muss sich der Schulinspektor nicht nur klar und dem definierten Ziel entsprechend verhalten, er muss den adressierten schulischen Akteuren auch erfolgreich vermitteln, dass er dies tut. Damit verfügt er jedoch nicht mehr über die Voraussetzungen für den Erfolg seines Handelns. Seine Vermittlungsversuche können immer auch zurückgewiesen werden oder unverstanden bleiben. In der Konsequenz steht es außerhalb der Verfügungsmacht des Schulinspektors, ein guter Schulinspektor zu sein. Die damit verbundene Möglichkeit des Scheiterns scheint in der in den Blick genommenen Formulierung zwar auf, gleichzeitig drückt sich jedoch deren normative Unzulässigkeit aus: Dem Schulinspektor darf dieses Scheitern nicht passieren, will er ein guter Schulinspektor sein. Wie im Folgenden erkennbar wird, prägt dieser Widerspruch zwischen dem Risiko des Scheiterns und seiner normativen Unzulässigkeit die Konzeption des Steuerungsinstruments:

Und auf der Basis Menschen arbeiten mit Menschen an dieser Aufgabenstellung gemeinsam mit der Schule arbeiten möchte.

Zur vorausgehenden Betonung des institutionellen und rollenförmigen Charakters der Interaktion steht die hier formulierte Programmatik „Menschen arbeiten mit Menschen“ in einem krassen Gegensatz. Gehen wir von einer klaren Aufgabenstellung und von einer Strukturierung der Situation durch ein klar gegliedertes Instrumentarium aus, stellt sich die Frage, wo der Raum für die nun das Spielfeld betretenden Menschen sein soll. Während vorausgehend die Akteure komplett hinter den durch sie repräsentierten Institutionen zurücktraten, treten sie nun nicht als Rollenhandelnde in Erscheinung, sondern als ganze Menschen. Das Modell einer pointiert funktional strukturierten Interaktion wird damit durch das einer diffusen Sozialbeziehung ganzer Menschen abgelöst. Diese Verwerfung in der Darstellung ist nicht einfach ein zufälliger Fehler. Vielmehr kann sie als Ausdruck eines Bearbeitungsversuchs der vorausgehend herausgearbeiteten Problematik gedeutet werden: Es geht um die Bearbeitung des Widerspruchs, dass die Gelingensbedingungen

der Schulinspektion nicht in der Hand des Schulinspektors liegen, das Gelingen jedoch in seinem Verantwortungsbereich verortet wird. Darauf reagiert der Sprecher mit der hier erkennbaren Konzeption, indem er jene Akteure als Menschen adressiert, an deren Umgang mit der Inspektion sich sein Erfolg entscheidet. Diese Adressierung zielt darauf ab, die schulischen Akteure für die Schulinspektion zu gewinnen. Indem sich der Sprecher zudem selbst als *Mensch* einbringt, verschiebt sich die Akzentuierung der Interaktion weg von der vorausgehend betonten Transparenz in Richtung Authentizität. Als Menschen können sich der Schulinspektor und die schulischen Akteure vergemeinschaften, um *gemeinsam an der Aufgabenstellung zu arbeiten*. Diese Vergemeinschaftung dient als *Basis* der Sicherstellung des strukturell ungewissen Erfolgs der Schulinspektion.

Doch auch an dieser Stelle der Darstellung, an der der Schulinspektor sein Motto, *Menschen arbeiten mit Menschen*, propagiert, reproduziert sich das strukturelle Grundproblem der schulinspektorischen Praxis. Auch hier wird auf die Ungewissheit des Vermittlungsgeschehens verwiesen: Ein gemeinsames Arbeiten von Menschen mit Menschen wird als etwas, das der Schulinspektor *möchte*, also als Wunsch charakterisiert. Inwieweit dieser Wunsch erfüllt oder ausgeschlagen wird, bleibt offen. Nichtsdestotrotz richtet sich die Hoffnung auf den Modus der Vergemeinschaftung als Bearbeitungsweise der Strukturproblematik:

Dieses dieser Begriff gemeinsam. Es gibt ja so einen andren Begriff der spricht davon auf Augenhöhe, der begleitet mich auch seit fünf Jahren. Ich persönlich hinterfrage den sehr, weil das mag ja gerne unser Wunsch sein aber wir kömmt einer Schule nicht vorschreiben, dass sie uns auf Augenhöhe wahrnimmt. Der einen Schule wird das gelingen, aber der anderen Schule halt nicht. Für die andere Schule bleib ich immer der, der von der Fähre steigt. Und der kann ich zehn Mal sagen ‚ich komme zu Ihnen und wir reden auf Augenhöhe‘.

In dieser weiteren Erläuterung wird die Konzeption der Vergemeinschaftung der eingangs betonten rollenförmigen Bezugnahme gegenübergestellt: Statt auf *Augenhöhe* soll *gemeinsam* gehandelt werden. Diese explizite Abgrenzung gegen die administrativ bzw. politisch vorgegebene Terminologie verweist auf konkrete Erfahrungen des Scheiterns des Schulinspektors: Er kann die schulischen Akteure nicht zur intendierten Adaptionsweise des Steuerungsimpulses Schulinspektion zwingen. Er kann „zehn Mal sagen, ich komme zu Ihnen und wir reden auf Augenhöhe“, ohne dass dies eine erfolgreiche Vermittlung gewährleistet. Nicht zufällig erinnert diese Formulierung an entgleitende Szenen pädagogischer Interaktionen, etwa wenn ein Mathematiklehrer seinen Schülerinnen und Schülern vorwirft, er könne ihnen zehnmal die binomischen Formeln erklären, und sie würden es trotzdem nicht verstehen.

Bei genauerem Hinsehen findet die in einer derartigen Aussage aggressiv gewendete Ohnmacht in der sich protokollierenden Aussage unmittelbaren Ausdruck: Wie oben ausgeführt (Kapitel 2), wird von Augenhöhe nur gesprochen, wenn die Symmetrie der Kommunikation in Frage steht. Dabei wird nur der Interaktionspartner Augenhöhe einfordern, der sich von seinem Gegenüber von oben herab behandelt fühlt. Stellen wir uns beispielsweise vor, im Rahmen parteipolitischer Koalitionsverhandlungen wird auf Verhandlungen auf Augenhöhe insistiert, so würde dies in aller Regel seitens der Partei geschehen, welche den geringeren Stimmenanteil in die zu bildende Koalition einbringt.

Dieses Parallelbeispiel macht deutlich, dass sich in der Deutung des Schulinspektors geradezu eine Umkehrung der strukturellen Hierarchie vollzieht, welche in Bezug auf das Verhältnis zwischen ihm und den schulischen Akteuren erwartet werden kann: Der Prüfer bittet vergeblich darum, von den Geprüften auf Augenhöhe wahrgenommen zu werden. In der Konsequenz scheint dem Schulinspektor nur noch übrig zu bleiben, an die „Menschlichkeit“ der schulischen Akteure zu appellieren, spricht sich mit ihnen zu vergemeinschaften. Die Möglichkeit der Verweigerung der Schule, welche dem Schulinspektor in seiner fünfjährigen Praxis begegnet, dekonstruiert also die offizielle Konzeption der Schulinspektion als Instrument einer evidenzbasierten Steuerung, innerhalb derer letztlich hierarchiefrei allein der *zwanglose Zwang des besseren Arguments* steuere.

Wie herausgearbeitet, impliziert die Strategie der Adressierung der schulischen Akteure als Vergemeinschaftungsofferte jedoch auch die Möglichkeit, dass diese abgelehnt werden kann. Blicken wir also abschließend darauf, wie der Schulinspektor damit umzugehen versucht:

Also da ist ein ganz ganz wichtiger Punkt wo wir sagen müssen gut dass Schulinspektion sich immer noch viele viele Stunden auch in Schule aufhält immer wieder auch Zeit hat für Rückfragen, für ein kleines Zwischengespräch mit einem Schulleiter, sehr wohl auch Mal den kleinen Small Talkkontakt sucht zu Lehrkräften. Einfach um um aufzuzeigen dass man hohes Interesse hat dass man das genau herausfinden will, um um den Menschen mit denen man dort arbeitet zu zeigen wir geben uns hier größtmögliche Mühe euch ein wirklich sauberes Ergebnis abzuliefern, ein ehrliches aber ein sauberes Ergebnis und nicht ein Ergebnis wo ihr, über das ihr euch grämen müsst sondern wo wirklich Eckpunkte benannt werden die sehr sehr hilfreich sind für die für die weitere Arbeit in der Schule.

Zunächst führt der Sprecher verschiedene Formen der Kontaktpflege an – angefangen vom *Sich-Zeit-Nehmen* bis zum *Small Talk*, der explizit nicht dem Austausch von Inhalten, sondern der Beziehungspflege dient. Hier werden nicht nur die schulischen Akteure als *Menschen* adressiert, auch der Schulinspektor bringt sich als *Mensch* ein. Jedoch werden diese kommunikativen Signale nicht allein als atmosphärische

Begleitung der eigentlichen Inspektionspraxis konzipiert. Vielmehr strukturiert diese Gestaltung der Interaktion auch die systematisierte Erfassung und Bewertung der Schule, um die es im Kern geht: Das Ergebnis der Schulinspektion soll *ehrlich* sein. Dies verweist auf den Geltungsanspruch der *Wahrhaftigkeit*. Es geht also nicht darum, ob das Ergebnis zutreffend ist, sondern um die Authentizität der Rückmeldung: Das Gegenteil einer ehrlichen Rückmeldung ist nicht allein eine *falsche*, sondern eine *unehrliche*.

Gleich doppelt wird das Attribut „sauber“ verwendet. Kontextfrei könnte ein sauberes Ergebnis als eines verstanden werden, welches verfahrenstechnisch „sauber“ zu Stande kam. Tatsächlich verweist hier das Attribut jedoch auf etwas anderes: Ein *sauberes Ergebnis* ist hier eines, für das sich die Schule *nicht grämen* muss, welches sie also nicht beschmutzt. Nur in dieser Lesart kann der Sprecher das Kriterium der Ehrlichkeit dem der Sauberkeit sinnvoll gegenüberstellen: Das Ergebnis soll ehrlich *aber* sauber sein. Diese Gegenüberstellung wäre nicht sinnvoll erklärbar, wenn das Kriterium der Sauberkeit auf das Verfahren verwiese, denn ein sauberes Verfahren arbeitet einer Unehrlichkeit gerade entgegen. Die Gegenüberstellung verweist also auf den problematischen Fall, dass die Ehrlichkeit die Mitteilung eines Ergebnisses verlangt, welches die Schule beschämen könnte. Darin begründet sich die eigentümliche Verschiebung der Maßstäbe, an denen der Schulinspektor die Güte der von ihm zu generierenden Evidenzen bemisst: An Stelle *wahrer* und *vergleichbarer* bzw. *schulspezifisch angemessener* Daten versucht er, *ehrliche* und *saubere* Rückmeldungen zu geben, um die Steuerungswirkung der Evidenzen zu erzeugen, welche die Konzeption einer evidenzbasierten Steuerung voraussetzt.

4. Gute Evidenzen

Die rekonstruierte Verschiebung der auf die Inspektionsrückmeldung bezogenen Gütemaßstäbe könnte als ein kontingenter Einzelfall gedeutet werden, protokolliert sich in den rekonstruierten Interviewpassagen doch allein die fallspezifische Aneignung der Schulinspektionsprogrammatik durch den befragten Schulinspektor. Jedoch weist diese Adaption insofern über den Einzelfall hinaus, als dass der Schulinspektor mit ihr ein sich aus der Programmatik der Schulinspektion ergebendes Strukturproblem zu bearbeiten versucht: Nicht zufällig fokussiert er auf sein Verhältnis zu den adressierten schulischen Akteuren und damit auf jene bereits auf programmatischer Ebene im Topos der „Augenhöhe“ anklingende Voraussetzung einer „Schulentwicklung durch Einsicht“. Diese bedingt jedoch eine Gefährdung des Konzeptes einer evidenzbasierten Steuerung: Auf Augenhöhe können die schulischen Akteure die qua Schulinspektionsbericht zur Verfügung gestellten Daten ablehnen und diese für nicht *evident* erachten. Sie können sich – aus administrativer Perspektive betrachtet – uneinsichtig zeigen und damit den inten-

dierten Steuerungsmechanismus der Schulinspektion still stellen. Weil die evidenzbasierte Steuerung Einsicht voraussetzt, verlagert sich die Deutungsmacht darüber, welches Wissen als Evidenzen anerkannt wird, in Richtung der schulischen Akteure. Bereits das unterschiedliche Verständnis von Evidenz auf Ebene der Schulinspektionsprogrammatik verweist immanent darauf, dass es ganz unterschiedliche Quellen für Evidenz gibt. Für die schulischen Akteure eröffnet sich die Option, eigensinnig und vor dem Hintergrund ihrer schulischen Praxis über die Plausibilität und Relevanz der angebotenen Evaluationsergebnisse zu entscheiden: Sie können bewerten, inwiefern ihnen die Inspektionsergebnisse im alltagsweltlichen Sinne *evident*, also augenscheinlich plausibel („self-evidence“; vgl. Jornitz 2008, S. 206f.) erscheinen. Dies läuft nicht nur der empiristisch gewendeten aufklärerischen Idee entgegen, „Vermutungen und Mutmaßungen“ durch „Fakten“ und damit schlechte Evidenzen durch bessere Evidenzen zu ersetzen. Gleichzeitig eröffnet sich die Möglichkeit, sich zu der in die Qualitätsrahmen und in die Schulinspektionsergebnisse eingeschriebenen Normativität zu verhalten. An dieser Stelle setzen die Bemühungen des Schulinspektors an, durch seine Person und sein Auftreten für die Ergebnisse der Schulinspektion zu werben. Seine Authentizität und Parteinahme für die Schule sollen die generierten Daten aufwerten und sie mit Evidenz ausstatten. Diese Adaption kann also als ein latenter Versuch des Nachsteuerens gedeutet werden, der darauf zielt, an der mit der Schulinspektion verbundenen Steuerungshoffnung festhalten zu können.

Vor dem Hintergrund dieser Deutung erscheint die Diskussion um jenes den Schulinspektionsprogrammatiken zu Grunde liegende Evidenzverständnis in neuem Licht: Sowohl die Betonung der Wissenschaftlichkeit des Verfahrens und damit der Güte der Ergebnisse als auch die pointierte Abgrenzung von derartigen standardisierten Verfahren in Verbindung mit der Herausstellung einer Orientierung am Einzelfall können als Versuche gedeutet werden, auf unterschiedlichem Wege die Inspektionsergebnisse glaubhaft zu machen und so aufzuwerten, um sie zu steuerungswirksamen Evidenzen werden zu lassen. Erst wenn sie als solche anerkannt werden und wenn sie damit als Evidenzen nicht mehr hinterfragt werden, können sie im Sinne des *zwanglosen Zwangs der – vermeintlich oder tatsächlich – besseren Evidenz* Wirkung entfalten. Diese Überzeugungsbemühungen laufen jedoch Gefahr, ein Hinterfragen der Evidenzen zu verhindern. Damit widersprechen sie dem Programm einer rationalen „Steuerung durch Einsicht“, wie nicht zuletzt die in der Fallrekonstruktion erkennbaren Verwerfungen zeigen. Es gilt sich zu entscheiden: Will man Schulinspektion zu einem rationalen Instrument einer „Steuerung durch Einsicht“ machen, muss anerkannt werden, dass die schulischen Akteure im Sinne einer professionellen Autonomie eigensinnig mit den Evidenzen umgehen und diese im Zweifelsfall auch ablehnen. Andernfalls sind Evidenzen lediglich Vehikel einer latenten klassisch-hierarchischen Beeinflussung.

Literatur

- Baltruschat, A. (2010): Evaluation als zeremonielles Rollenspiel: Reflexion der externen Evaluation an bayerischen Schulen. In: Schönig, W./Baltruschat, A./Klenk, G. (Hrsg.): Dimensionen pädagogisch akzentuierter Schulevaluation. Baltmannsweiler: Schneider Verlag Hohengehren, S. 249-281.
- Bos, W./Voss, A. (2008): Empirische Schulentwicklung auf Grundlage von Lernstandserhebung. Ein Plädoyer für einen reflektierten Umgang mit Ergebnissen aus Leistungstests. In: Die Deutsche Schule 100, H. 4, S. 449-458.
- Böttcher, W./Keune, M. (2010): Funktionen und Effekte der Schulinspektion. Ausgewählte nationale und internationale Forschungsbefunde. In: Böttcher, W./Dicke, J.N./Hogrebe, N. (Hrsg.): Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit. Münster u.a.: Waxmann, S. 151-164.
- Böttger-Beer, M./Koch, E. (2008): Externe Schulevaluation in Sachsen – ein Dialog zwischen Wissenschaft und Praxis. In: Böttcher, W./Bos, W./Döbert, H./Holtappels, H.G. (Hrsg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Münster u.a.: Waxmann, S. 253-264.
- Dedering, K. (2009): Schulministerien und ihr Umgang mit Leistungsvergleichsstudien. In: Böttcher, W./Dicke, J.N./Ziegler, H. (Hrsg.): Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis. Münster u.a.: Waxmann, S. 201-210.
- Döbert, H./Rürup, M./Dedering, K. (2008): Externe Evaluation von Schulen in Deutschland. Die Konzepte der Bundesländer, ihre Gemeinsamkeiten und Unterschiede. In: Döbert, H./Dedering, K. (Hrsg.): Externe Evaluation von Schulen. Historische, rechtliche und vergleichende Aspekte. Münster u.a.: Waxmann, S. 63-151.
- Dr. Otto Seydel Institut für Schulentwicklung (o. J.): Externe Evaluation Bremer Schulen V: Information für die beteiligten Schulen. Dr. Otto Seydel Institut für Schulentwicklung, Überlingen: Otto Seydel Institut. URL: http://www.bildung.bremen.de/fastmedia/13/evaluation_info.pdf; Zugriffsdatum: 14.02.2012.
- Husfeldt, V. (2011): Wirkungen und Wirksamkeit der externen Schulevaluation. Überblick zum Stand der Forschung. In: Zeitschrift für Erziehungswissenschaft 14, H. 2, S. 259-282.
- Jornitz, S. (2008): Was bedeutet eigentlich „evidenzbasierte Bildungsforschung“?: Über den Versuch Wissenschaft für Praxis verfügbar zu machen am Beispiel der Review-Erstellung. In: Die Deutsche Schule 100, H. 2, S. 206-216.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Hrsg.) (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. München: LinkLuchterhand. URL: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_08_01-Gesamtstrategie-Bildungsmonitoring.pdf; Zugriffsdatum: 14.02.2012.
- Kussau, J./Brüsemeister, T. (2007): Educational Governance: Zur Analyse der Handlungskoordination im Mehrebenensystem Schule. In: Altrichter, H./Brüsemeister, H./Wissinger, J. (Hrsg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS, S. 15-54.
- Lambrecht, M./Rürup, M. (im Druck): Bildungsforschung im Rahmen einer evidence based policy. Das Beispiel „Schulinspektion“. In: Wacker, A./Mayer, U./Wissinger, J. (Hrsg.): Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen. Wiesbaden: VS.
- Lohmann, A./Reißmann, J. (2007): Von der Verstrickung der Rollen zum funktionalen Unterstützungsnetz – Die Schulinspektion im Gefüge von Schulleitung, Kollegium, Schulaufsicht und Unterstützungssystem. In: Journal für Schulentwicklung 11, H. 3, S. 15-24.

- LISA (Landesinstitut für Schule und Ausbildung) Mecklenburg-Vorpommern (Hrsg.) (2006): Gute Schule: Externe Evaluation von Schulen in Mecklenburg-Vorpommern. Schwerin: LISA. URL: http://www.bildung-mv.de/export/sites/lisa/de/schule/entwicklung/selbst_schule/dok_s_schule/GS_Ext_Evaluation.pdf; Zugriffsdatum: 14.02.2012.
- LS (Landesinstitut für Schulentwicklung) Stuttgart (Hrsg.) (2009): Fremdevaluation an allgemein bildenden Schulen in Baden-Württemberg: Informationen zur Regelphase (QE-11). Stuttgart: LS. URL: http://www.schule-bw.de/entwicklung/qualieval/qualiabst/fev/QE11_Fremdevaluation_2009-10.pdf; Zugriffsdatum: 14.02.2012.
- Lüders, C. (2006): Qualitative Evaluationsforschung: Was heißt hier Forschung? In: Flick, U. (Hrsg.): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek b. Hamburg: Rowohlt, S. 33-62.
- Maier, U. (2008): Was lernen Schulen aus zentralen Tests? Was sollten Bildungspolitiker lernen, wenn sie testen lassen? In: Die Deutsche Schule 100, H. 1, S. 66-72.
- Niedersächsisches Kultusministerium (2005): Schulinspektion Niedersachsen nimmt in Bad Iburg ihre Arbeit auf – Busemann: „Partnerschaftliche Unterstützung der Schulen auf Augenhöhe“ (Presseerklärung vom 03. Mai 2005). Hannover: Niedersächsisches Kultusministerium. URL: http://www.mk.niedersachsen.de/portal/live.php?navigation_id=1820&article_id=5903&psmand=; Zugriffsdatum: 14.02.2012.
- Oevermann, U. (2000): Die Methode der Fallrekonstruktion in der Grundlagenforschung sowie in der klinischen und pädagogischen Praxis. In: Kraimer, K. (Hrsg.): Die Fallrekonstruktion. Sinnverstehen in der sozialwissenschaftlichen Forschung. Frankfurt a.M.: Suhrkamp, S. 58-156.
- Posch, P. (2009): Zur schulpraktischen Nutzung von Daten. Konzepte, Strategien, Erfahrungen. In: Die Deutsche Schule 101, H. 2, S. 119-135.
- Rürup, M. (2008): Typen der Schulinspektion in den deutschen Bundesländern. In: Die Deutsche Schule 100, H. 4, S. 467-477.
- van Ackeren, I./Zlatkin-Troitschanskaia, O./Binnewies, C./Clausen, M./Dormann, C./Preisendörfer, P./Rosenbusch, C./Schmidt, U. (2011): Evidenzbasierte Schulentwicklung: Ein Forschungsüberblick aus interdisziplinärer Perspektive. In: Die Deutsche Schule 103, H. 2, S. 170-184.
- Wernet, A. (2006): Einführung in die Interpretationstechnik der Objektiven Hermeneutik. Wiesbaden: VS.

Dietrich, Fabian, Dr. des., geb. 1975, wissenschaftlicher Mitarbeiter im Arbeitsbereich Bildungsforschung am Institut für Erziehungswissenschaft der Leibniz Universität Hannover.

Anschrift: Schloßwender Str. 1, 30159 Hannover
E-Mail: fabian.dietrich@iew.phil.uni-hannover.de

Maike Lambrecht, geb. 1978, wissenschaftliche Mitarbeiterin im Arbeitsbereich Bildungsforschung am Institut für Erziehungswissenschaft der Leibniz Universität Hannover.

Anschrift: Schloßwender Str. 1, 30159 Hannover
E-Mail: maike.lambrecht@iew.phil.uni-hannover.de

Tobias Feldhoff/Luzia Durrer/Stephan G. Huber

Steuerung eines Schulsystems

Eine empirische Analyse, wie Akteure die Steuerungs- konfigurationen des Schulsystems wahrnehmen und sich deren zukünftige Gestaltung wünschen

Zusammenfassung

In diesem Beitrag wird exemplarisch anhand einer Studie eines Schweizer Kantons im Reformprozess aufgezeigt, wie neue Steuerungsmodelle in Schulsystemen anhand von Governance-Reglern analytisch und empirisch beschrieben werden können. Die Studie bedient sich hierfür unterschiedlicher methodischer Zugänge (Dokumentenanalysen von Schulgesetzen und Befragungen von schulischen Akteuren nach ihren Einschätzungen zu aktuellen und zukünftig gewünschten Steuerungskonfigurationen). Des Weiteren gibt der Beitrag über die Studie hinaus Impulse dafür, welche Potentiale eine Analyse von Governance-Reglern für die Bildungsforschung und Bildungsverwaltung im Allgemeinen haben kann.

Schlüsselwörter: Governance, Steuerung im Bildungswesen, Handlungskoordination

Governance of a School System

An Empirical Analysis of How Actors Perceive the Governance Configurations of the School System and How They Want Its Future Configuration to Be

Abstract

This article shows exemplarily with regard to a Swiss canton, which is in a reform process, how new governance models in the school system can be analytically and empirically described by means of governance rulers. Therefore, the study uses different methodological approaches (document analyses of school laws and interviews of school actors about their appreciation of current governance configurations and those desired in the future). Furthermore, the article gives impulses beyond the study with regard to the potentials, which an analysis of governance rulers may have for educational research and educational administration in general.

Keywords: governance, educational governance, action coordination

1. Einleitung – Anpassung und Umbau von Steuerungsstrukturen im Bildungswesen

Das Thema „Steuerung im Bildungsbereich“ hat in den letzten Jahren im deutschsprachigen Raum in der Politik, in der Wissenschaft und auch in der Öffentlichkeit zunehmend an Bedeutung gewonnen (vgl. von Kopp 2008). Als Reaktion auf das relativ schlechte Abschneiden dieser Länder bei internationalen Vergleichsuntersuchungen, wie der TIMS- und vor allem der PISA-Studie im Jahr 2000, wurden zahlreiche „Maßnahmen initiiert, um die Qualität des Bildungssystems und seiner Ergebnisse [...] weiterzuentwickeln und nachhaltig zu sichern“ (Altrichter/Brüsemeister/Wissinger 2007, S. 9). Abgesehen von diesem ‚schlechten‘ Abschneiden bei internationalen Vergleichsstudien gilt es ohnehin, Schul- und Bildungssysteme an laufende gesellschaftliche und kulturelle Veränderungen im Zuge der Globalisierung und des Übergangs in eine Wissensgesellschaft anzupassen; denn „[e]s ist [...] Aufgabe der Schule, gesellschaftliche Anforderungen und Entwicklungen zur Kenntnis zu nehmen und die Kinder und Jugendlichen auf die Welt von morgen vorzubereiten“ (Kussau 2005, S. 5). Mit dieser Anpassung ist der Umbau traditioneller Steuerungsstrukturen im Bildungs- und Schulwesen verbunden (vgl. Fend 2008; Kussau/Brüsemeister 2007; von Kopp 2008). Dabei befinden sich die jeweiligen Bundesländer und Kantone auf der Suche nach geeigneten Steuerungsmodellen, mit denen die Leistungen der Bildungssysteme zielgerichteter und effizienter erbracht werden können. Die bisherigen Bemühungen in diesem Kontext zielen tendenziell auf eine Umstrukturierung des klassischen Modells einer bürokratischen Inputsteuerung in Richtung einer evaluationsbasierten Outputsteuerung (vgl. Brüsemeister 2007).

In Bezug auf die initiierten Reformmaßnahmen und den damit verbundenen Umbau der Bildungssteuerung stellen sich aus der Sicht der Bildungsforschung folgende Fragen (vgl. Altrichter/Brüsemeister/Wissinger 2007):

- Wie sehen diese neuen Steuerungsmodelle aus bzw. wie lassen sie sich analytisch und empirisch beschreiben?
- Erfüllen diese neuen Modelle die Intentionen, die mit ihnen verbunden sind (v.a. in Bezug auf die Qualität des Bildungswesens)? Haben sie – erwünschte oder unerwünschte – Nebenwirkungen?

Der Beitrag versucht anhand einer Studie zur Steuerung und Handlungskoordination des Volksschulsystems im Kanton Uri (vgl. Feldhoff/Huber/Durrer 2009, 2011), exemplarisch Antworten auf die erste Frage – nach der analytischen und empirischen Beschreibung neuer Steuerungsmodelle – zu geben. Hierfür wird in Kapitel 2 zunächst die Governance-Perspektive anhand des Regler-Modells von Schimank (2007) zur empirischen Analyse von komplexen Steuerungs- und Regelungsstrukturen im Mehrebenensystem Schule vorgestellt. In Kapitel 3 folgt die Beschreibung der Studie und des Forschungsdesigns. In Kapitel 4 werden die Ergebnisse der empiri-

schen Analysen mittels des Regler-Modells vorgestellt. Neben einer Analyse der Schulgesetze (4.1) wurden die schulischen Akteure im Kanton nach ihren Einschätzungen der derzeitigen und der zukünftig gewünschten Steuerung (4.2) gefragt. Kapitel 5 setzt die Ergebnisse der einzelnen Analysen zueinander in Beziehung und interpretiert diese. Zum Schluss wird in Kapitel 6 von dem exemplarischen Beispiel abstrahiert und aufgezeigt, welche Potentiale eine Analyse von Governance-Reglern für die Bildungsforschung und Bildungsverwaltung im Allgemeinen haben kann.

2. Governance als Konzept zur empirischen Analyse von komplexen Regelungsstrukturen im Mehrebenensystem Schule

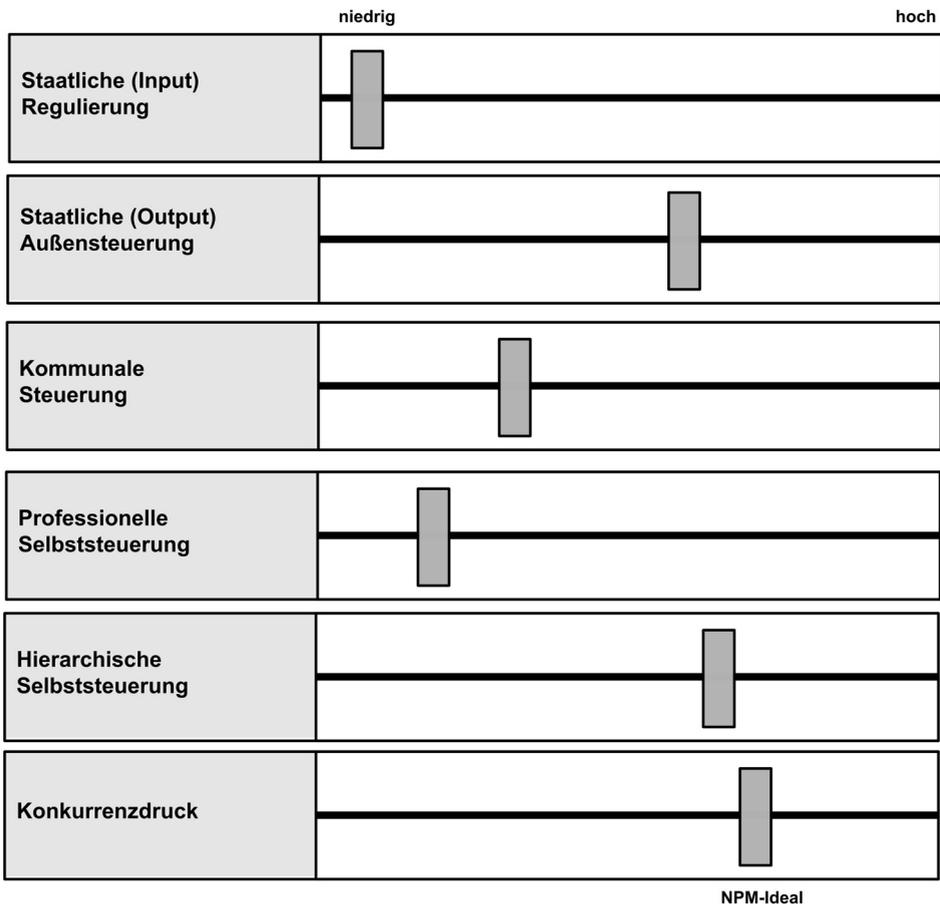
In der Politik- und Verwaltungswissenschaft lässt sich vermehrt feststellen, dass neben staatlichen auch andere gesellschaftliche Einrichtungen an der Auftragserfüllung im öffentlichen Sektor beteiligt sind (vgl. Benz 2004). Dies gilt in wachsendem Maße auch für den Bildungsbereich. Formen der Kooperation und Aushandlung gewinnen neben klassischen Formen der Steuerung durch Rechtsnormen oder Finanzmittel zunehmend an Bedeutung. In diesen neuen Regelungsstrukturen ist oftmals eine eindeutige Zuordnung von Steuerungsobjekt und -objekt, wie bei hierarchischen Steuerungsmodellen üblich, nicht mehr möglich bzw. führt zu einer verengten Sichtweise, die wesentliche Elemente der Steuerung ausblendet (vgl. Mayntz 1998).

Steuerung im Sinne von Governance ist als ein wechselseitiges Verhältnis zu verstehen, das von der Koproduktion von Leistungsbeiträgen verschiedener Akteure auf unterschiedlichen Ebenen des Schulsystems (Bildungsdirektion auf kantonaler Ebene, Schulträger auf Gemeindeebene, Schulhausleitung auf Schulebene, Lehrpersonen und Schülerinnen und Schüler auf Unterrichtsebene), die in vielfältigen, komplexen und wechselseitigen Beziehungen zueinander stehen, abhängig ist. Der Governance-Ansatz widmet sich explizit diesem „Mehrebenencharakter des Bildungssystems und [deren] Übergänge[n] zwischen den Ebenen“ (Altrichter/Brüsemeister/Wissing 2007, S. 11). Analytisch gesehen gilt es zu klären, welchen Beitrag die einzelnen Akteure und die Kombination dieser komplexen Regelungsstrukturen, basierend auf den unterschiedlichen Formen der Handlungskoordination, zur Qualität des Bildungswesens leisten.

Schimank hat, basierend auf Vorarbeiten von Clark (1998) und Braun (1999), ein „spezifisches Analyseinstrument von *mittlerem Abstraktionsgrad*“ (Schimank 2007, S. 231) entwickelt, mit dem staatliche oder staatsnahe Governance-Regime in ihrer zeitlichen Entwicklung oder im Vergleich zu anderen Governance-Regimen untersucht werden können. Als Analysegegenstand diente der Vergleich von Steuerungsmodellen von Hochschulen in verschiedenen Ländern. Das Modell beschreibt fünf grundlegende Governance-Dimensionen bzw. -Regler, mit Hilfe derer Hochschul-, aber auch Schulsysteme gesteuert werden können (vgl. Schimank 2007). Dieses

Modell lässt sich wie folgt auf den schulischen Kontext übertragen (vgl. Altrichter/Soukup-Altrichter 2008):

Abbildung 1: Governance-Dimensionen mit idealtypischer Reglerstellung nach dem New-Public-Management-Modell



Quelle: eigene Darstellung

Die staatliche Inputregulierung beschreibt die klassische Form der Steuerung mittels staatlicher Vorgaben in den Bereichen Finanzen, Personal, Organisation und Unterricht. Je stärker der Detailgrad der Regulierung ist, desto ausgeprägter ist dieser Steuerungsmechanismus.

Im Bereich der Finanzen sind zwei Gegenpole erkennbar: die Regelung durch ein Globalbudget, z.B. in Form einer Schülerpauschale, und die bisherige Praxis von an fixe Positionen gebundenen Finanzmitteln. Personal bezieht sich auf die staatliche Regulierung von Einstellung, Beschäftigungsdauer, Entlohnung, Umsetzungs-

und Kündigungsmöglichkeiten. Organisation umfasst die Regulierung von Aufbauorganisationen, Leitungs- und Beteiligungsstrukturen im schulischen Kontext sowie Instrumente der Schulentwicklung, wie Leitbilder, Schulprogramme etc., und Richtlinien, die die Größe der Organisation und ihrer Untereinheiten regeln, wie z.B. Klassengröße, Angaben zur Lehrer-Schüler-Relation etc. Unterricht beinhaltet die Regulierung durch Fächervorgaben, Studentafeln, Lehrpläne, Unterrichtsorganisation und -gestaltung, Versetzungsregeln, Disziplinarmaßnahmen und Lehrmittel.

Die staatliche Außensteuerung erfolgt im Gegensatz zur staatlichen Inputsteuerung nicht über Konditionalprogramme im Sinne von Wenn-Dann-Regelungen, sondern über Zweckprogramme in Form von Zielvorgaben, bei denen der Weg zur Zielerreichung den Schulen selbst überlassen wird (vgl. Luhmann 1964). Die staatliche Außensteuerung ist eine outputorientierte Form der Steuerung, die z.B. mit Hilfe von festgelegten Zielgrößen (externe Evaluation, Bildungsstandards usw.) erfolgt.

Schimank (2007) verwendete die fünf Governance-Dimensionen zur Beschreibung und zum Vergleich von Hochschulsystemen. Im Unterschied zu den Hochschulen liegt die Zuständigkeit im Schulbereich auf der kantonalen und zusätzlich auf der gemeindlichen Ebene beim Gemeinderat und Schulrat. Der Schulrat, in einigen Kantonen auch Schulpflege oder Schulkommission genannt, ist in der Schweiz ein von der Gemeinde gewähltes Laiengremium, das die Schule in Bezug auf die strategischen Ziele leitet. Um diese für den schulischen Kontext bedeutsame gemeindliche Ebene zu berücksichtigen, wurde das Modell um einen weiteren Regler erweitert, nämlich den der kommunalen Steuerung.

Die professionelle Selbststeuerung beschreibt die Einfluss- und Entscheidungsmöglichkeiten von schulischen Mitwirkungs- und Entscheidungsgremien bei schulrelevanten Fragen in den o.g. Bereichen (Kategorien analog zur staatlichen Regulierung). Dies betrifft auch die individuelle Autonomie der einzelnen Lehrpersonen.

Die hierarchische Selbststeuerung beschreibt hierzu analog den Einfluss und die Entscheidungsmöglichkeiten der schulischen Leitungspersonen in den o.g. Bereichen. Eine weitere Form der hierarchischen Selbststeuerung sind Instrumente wie Leitbild, Schulprogramm, Qualitätsbericht und Jahresbericht.

Der Konkurrenzdruck beschreibt, inwieweit es eine quasimarktförmige Konkurrenz, z.B. im Bereich von Schüleranmeldungen und Schulprofilen zwischen einzelnen Gemeinden, zwischen einzelnen Schulen oder auch innerhalb von Schulen, gibt. Bisherige Untersuchungen im deutschsprachigen Raum, wie z.B. von Büeler (2007) im Kanton Luzern, zeigen, dass marktförmige Mechanismen im schulischen Bereich ‚noch‘ kaum eine Rolle spielen. Dennoch könnte im Zuge einer freien Schulwahl und der Zusammenlegung von Schulen aufgrund der demografischen Entwicklung

sowie durch Privatschulen auch dieser Mechanismus an Bedeutung gewinnen (vgl. Altrichter/Soukup-Altrichter 2008).

In Anlehnung an Schimank (2007) gilt es in den Governance-Dimensionen bzw. -Reglern zwischen der formellen Regelung und dem kulturellen Aspekt, d.h. der konkreten Ausgestaltung der Formen der Handlungskoordination durch die beteiligten Akteure, zu unterscheiden.

Für die Analyse der Governance-Regler ist davon auszugehen, dass die Dimensionen aus theoretischer Sicht nicht unabhängig voneinander sind. Auch empirische Befunde legen dies nahe. Bis auf die Untersuchungen von Schimank (2007) und Altrichter (2010) gibt es jedoch bisher keine Studie, die den Zusammenhang der Regler untersucht. Schimank (2007) und auch Altrichter (2010) beschreiben gewisse Kombinationen der Governance-Regler als konfliktträchtig. Die auftretenden Spannungen dieser Kombinationen lassen zumindest ein dauerhaftes Bestehen einer solchen Steuerungskonfiguration als unwahrscheinlich erscheinen:

- starke staatliche (Input-)Regulierung und starker Konkurrenzdruck;
- starke staatliche (Input-)Regulierung und starke hierarchische Selbststeuerung der Schule;
- starke professionelle Selbstorganisation und starke Außensteuerung;
- starke professionelle Selbstorganisation und starke hierarchische Selbststeuerung der Schule;
- starke hierarchische Selbststeuerung der Schule und starke kommunale Steuerung.

Für eine Übergangszeit wäre ein paralleles ‚Hochfahren‘ von eigentlich gegensätzlichen Dimensionen durchaus denkbar.

3. Forschungsdesign

Zur Identifizierung von zentralen gesellschaftlichen und bildungspolitischen Herausforderungen an das Volksschulsystem und der gleichzeitigen Entwicklung von Strategien zum Umgang mit diesen hat der Kanton Uri das Projekt „Volksschule 2016“ ins Leben gerufen. Im Rahmen dieses Projekts wurde eine wissenschaftliche Expertise zur derzeitigen und zukünftigen Steuerung im Kanton erstellt (vgl. Feldhoff/Huber/Durrer 2009).

Basierend auf der oben skizzierten theoretischen Analyseperspektive wurden im Rahmen der Expertise drei zentrale Forschungsfragen formuliert:

1. Wie werden die derzeitigen und zukünftigen zentralen kantonalen, interkantonalen bzw. nationalen Anforderungen an das Volksschulwesen im Kanton Uri beschrieben und aus Sicht der beteiligten Akteure eingeschätzt?

2. Wie sehen die aktuellen Steuerungs- und Regelungsstrukturen der Volksschulen im Kanton Uri aus?
3. Wie sollten aus Sicht der beteiligten Akteure die Steuerungs- und Gestaltungsstrukturen der Volksschulen im Kanton Uri angesichts der momentan abzusehenden Entwicklungen sowie der Wahrnehmung und der Bewertung der derzeitigen Strukturen zukünftig gestaltet sein (bis zum Jahre 2016)?

Zur Beantwortung der Forschungsfragen wurde ein mehrstufiges Design gewählt. Dieses besteht aus einer Dokumentenanalyse zu den formalrechtlichen Regelungen der Steuerungs- und Regelungsstrukturen der Volksschule, einer ersten quantitativen Online-Befragung der schulischen Akteure¹ zur aktuellen Steuerung und Handlungskoordination sowie einer zweiten quantitativen Online-Befragung der schulischen Akteure zur derzeitigen und zukünftigen Steuerung und Gestaltung. Beide Erhebungen wurden als Vollerhebungen durchgeführt. Der Rücklauf über alle Akteursgruppen hinweg betrug in der ersten Erhebung 58% (N=393) und in der zweiten Erhebung 41% (N=277). Eine detaillierte Darstellung des Designs, methodischen Vorgehens und der Rückläufe findet sich bei Feldhoff, Huber und Durrer (2009).

4. Ergebnisse

Die nachfolgenden Ergebnisse beziehen sich auf Teilaspekte der zweiten Forschungsfrage zu formalrechtlichen Regelungen der Steuerungs- und Regelungsstrukturen mittels Dokumentenanalyse (4.1) und der dritten Forschungsfrage nach der derzeitigen und zukünftigen Steuerung anhand der zweiten Online-Befragung (4.2).

4.1 Steuerung im Kanton Uri auf Basis formaler Regelungen

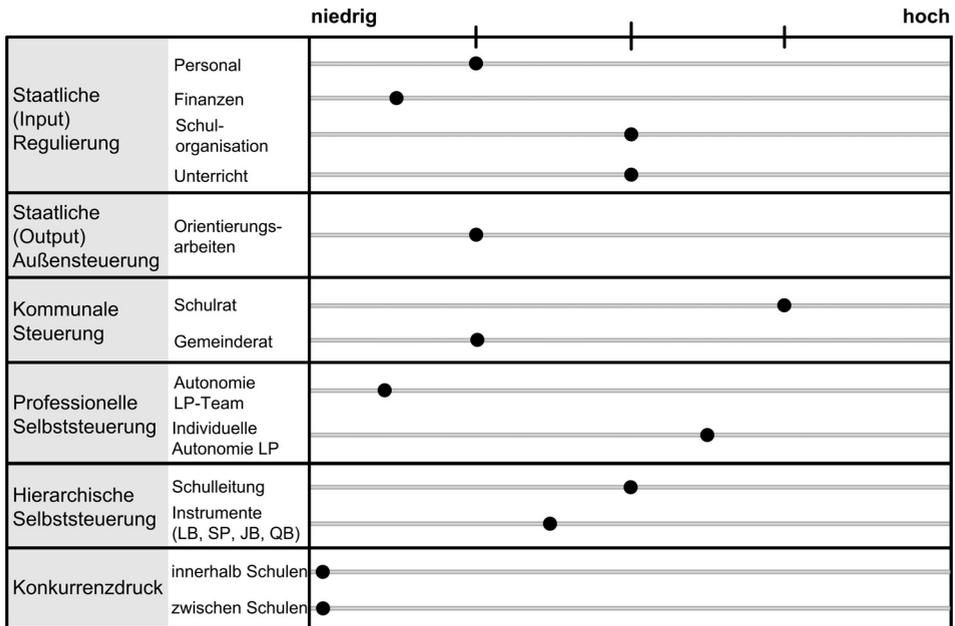
Zur Beantwortung der Frage nach den aktuellen Steuerungs- und Regelungsstrukturen der Volksschulen im Kanton Uri wurden die amtlichen Dokumente des Kantons mittels einer qualitativen Inhaltsanalyse analysiert (vgl. Mayring 2008). Zudem wurde in der ersten Erhebung die Wahrnehmung des Steuerungshandelns der unterschiedlichen Akteurskonstellationen aus Sicht der schulischen Akteure untersucht (vgl. Feldhoff/Huber/Durrer 2011). Die formalen Regelungen können in diesem Kontext als rechtlicher ‚Orientierungs‘-Rahmen verstanden werden, innerhalb dessen sich

¹ Die zu befragenden Expertinnen und Experten bestehen aus der Projektgruppe, den Schul(haus)leitungen, Lehrpersonen, Schulräten, Mitarbeiterinnen und Mitarbeitern des Amtes für Volksschulen, Gemeinderäten, Erziehungsräten sowie Vertreterinnen und Vertretern von „Schule & Elternhaus Uri“, der Elternorganisation des Kantons Uri.

das Steuerungshandeln der Akteure auf den unterschiedlichen Ebenen vollzieht bzw. durch den das Handeln der Akteure vorstrukturiert wird.

Auf Basis des theoretischen Rahmenmodells der Governance-Regler zur Erfassung der schulischen Steuerung wurden die schulrechtlichen Dokumente des Volksschulwesens im Kanton Uri in Zusammenarbeit mit dem Amt für Volksschulen analysiert. Kriterium für die Analyse war das Ausmaß an Steuerung (niedrig, mittel, hoch) durch den jeweiligen Regler (vgl. Abbildung 2).

Abbildung 2: Formale Regelung der Steuerungs- und Regelungsstrukturen des Volksschulsystems im Kanton Uri



LB = Leitbild, SP = Schulprogramm, JB = Jahresbericht, QB = Qualitätsbericht

Quelle: eigene Darstellung

Die staatliche Inputregulierung stellt sich in den vier Subkategorien Personal, Finanzen, Schulorganisation und Unterricht wie folgt dar: Die meisten Entscheidungen im Bereich Personal (Wahl, Einstellung, Besoldung und Beurteilung der Lehrpersonen sowie der Schulleitung) obliegen dem Schulrat. Die Vorgaben von Seiten des Kantons sind relativ gering.

Die Gemeinden erhalten zur Finanzierung der Schule vom Kanton eine Schülerpauschale. Damit verfügen sie über ein Globalbudget. Dies ist als sehr geringe Detailsteuerung einzustufen.

Vorgaben im Bereich Schulorganisation beinhalten die Einführung von Schulprogrammen, Leitbildern, Jahresberichten und Qualitätsbeauftragten. Dies ist als ein mittleres Ausmaß der Regelung zu interpretieren.

Im Bereich Unterricht erlässt der Erziehungsrat Lehrpläne, Stundentafeln, Versetzungsregeln und Disziplinarmaßnahmen sowie die Entscheidung über Lehrmittel in Kernfächern. Die Wahl der Lehrmethoden sowie die zeitliche Rhythmisierung liegen im Entscheidungsbereich der Lehrpersonen. Insgesamt ergibt sich ein mittleres Ausmaß an Detailregelungen. Über die verpflichtenden Orientierungsarbeiten hinaus gibt es im Kanton keine Maßnahmen der staatlichen Außensteuerung. Daher ist das Ausmaß als niedrig einzustufen.

Der Schulrat leitet das Schulwesen in der Gemeinde und ist für alle Entscheidungen des Schul- und Erziehungswesens auf Gemeindeebene zuständig, u.a. Wahl, Entlassung und Besoldung der Lehrpersonen, Überprüfung der Stundentafel und Schulpflicht sowie Verwaltung der Schulgebäude. Der Schulrat verfügt über ein hohes Ausmaß an Befugnissen.

Der Gemeinderat hat mit Ausnahme der Genehmigung des Schulbudgets keine Befugnisse im Schulwesen; somit sind seine formalen Befugnisse als relativ niedrig einzuschätzen.

Das Lehrpersonenteam hat als schulisches Gremium keine formalen Befugnisse. Neben den bereits skizzierten Vorgaben im Unterricht haben die einzelnen Lehrpersonen Mitspracherecht bei der und eine Verpflichtung zur Gestaltung der Schule. Die Schulleitung kann die Autonomie der Lehrpersonen im Rahmen der Sicherstellung der schulischen Qualität, der individuellen Weiterbildung und der Lehrerbeurteilung einschränken. Die individuelle Autonomie der Lehrperson kann auf einem mittleren Niveau eingeschätzt werden.

Die Schulleitung hat eine Reihe von Kompetenzen und Autonomie, die durch den Schulrat eingeschränkt oder erweitert werden können. Insgesamt ergibt sich ein mittleres Ausmaß der hierarchischen Selbststeuerung durch die Schulleitung. Des Weiteren stehen der Schulleitung Instrumente (Leitbild, Jahresberichte, Schulprogramm, Qualitätsbeauftragte) zur Leitung der Schule zur Verfügung. Das Steuerungspotential der Instrumente liegt auf einem mittleren Niveau. Jedoch ist darauf hinzuweisen, dass die Schulleitung in den Volksschulen des Kantons Uri noch ein junger Akteur ist, der erst 2003 eingeführt wurde. Vorher lag die alleinige Leitung des Schulwesens beim Schulrat.

Im Kanton Uri existieren keine freie Schulwahl oder andere marktförmige Mechanismen wie Schulprofile. Folglich gibt es auf formaler Basis keine Steuerung über quasimarktförmige Konkurrenzmechanismen.

4.2 Derzeitige und zukünftige Steuerungs- und Gestaltungsstrukturen im Volksschulsystem

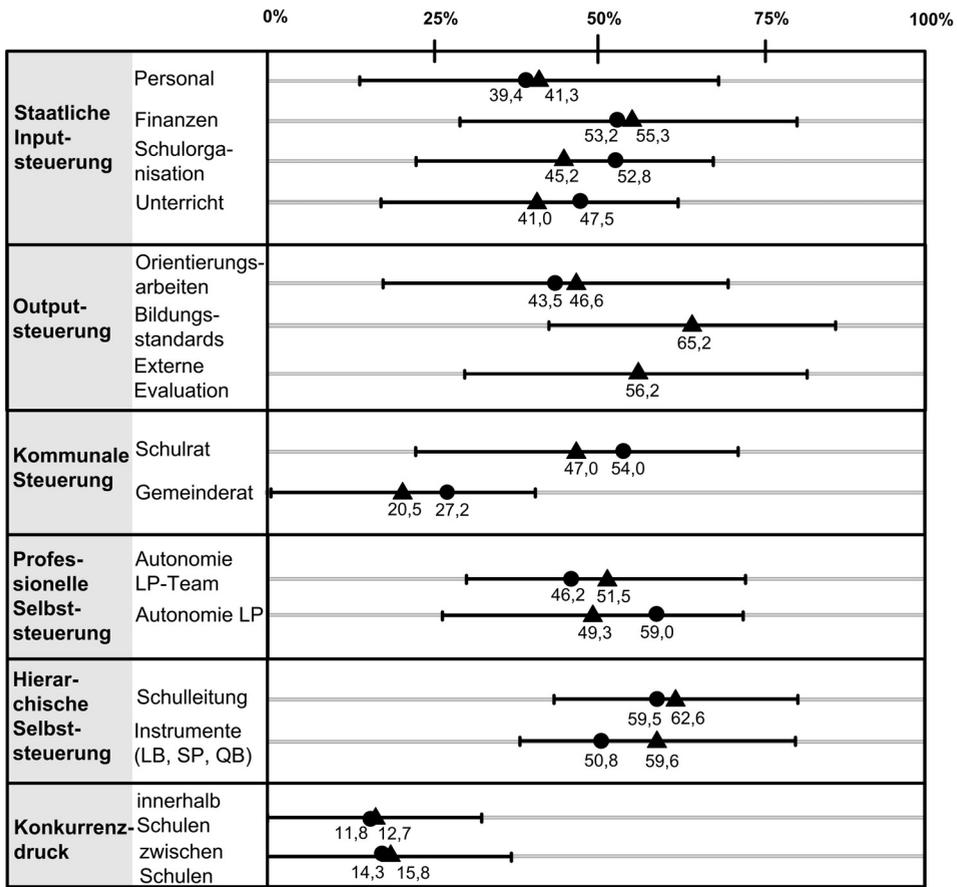
Im Rahmen der dritten Forschungsfrage wurde untersucht, wie aus Sicht der beteiligten Akteure die Steuerungs- und Gestaltungsstrukturen der Volksschulen im Kanton Uri angesichts der momentan abzusehenden Entwicklungen sowie der Wahrnehmung und der Bewertung der derzeitigen Strukturen zukünftig (bis zum Jahre 2016) gestaltet sein sollten. Zur Beantwortung dieser Forschungsfrage wurden aus den Ergebnissen der ersten Erhebung Themenbereiche für die zweite Erhebung ausgewählt, bei denen die zukünftige Gestaltung der Volksschule im Mittelpunkt steht. Den Akteuren wurden u.a. die Ergebnisse der Dokumentenanalyse zur formalen Regelung präsentiert. Sie sollten mittels eines Schiebereglers einschätzen, in welchem Ausmaß (0% = niedrig, 100% = sehr hoch) sie die Steuerung im jeweiligen Bereich wahrnehmen und wie sie sich diese zukünftig wünschen (vgl. Abbildung 3).

Um den Rahmen dieses Beitrags nicht zu sprengen, werden die Ergebnisse über alle Akteursgruppen hinweg präsentiert. Auch wenn es bei einzelnen Reglern statistisch bedeutsame Unterschiede zwischen Akteursgruppen gibt, so sind die Ergebnisse der Gesamtanalyse von der Tendenz her mit den detaillierten Analysen vergleichbar (vgl. Feldhoff/Huber/Durrer 2009). Da in den Analysen keine Gewichtungen nach Akteursgruppen vorgenommen wurden, haben die Einschätzungen der Lehrpersonen als größte Personengruppe in der Untersuchung einen größeren Einfluss auf den Gesamtmittelwert und die Streuung als die Einschätzung von Personen anderer Akteursgruppen.

Generell gilt, dass die Varianz bei der Einschätzung fast aller Governance-Dimensionen – ausgenommen Konkurrenz – groß ist. Die Ergebnisse zeigen, dass dies sowohl auf unterschiedliche Einschätzungen innerhalb der jeweiligen Akteursgruppen als auch zwischen ihnen zurückzuführen ist.

Die staatliche Inputregulierung im Bereich Personal wird von den Akteuren höher eingeschätzt als die formale Regelung auf Basis der Dokumentenanalyse. Der Unterschied zwischen derzeitiger und zukünftig gewünschter Regulierung ist statistisch nicht bedeutsam. Die Streuung nimmt im Vergleich mit den Einschätzungen der derzeitigen Regulierung noch zu. Die Akteure schätzen das Ausmaß der Regulierung durch die Finanzen deutlich höher ein, als sich dies auf formaler Ebene zeigt. Auch hier ist der Unterschied zwischen derzeitiger und zukünftig gewünschter Regulierung nicht signifikant, mit zunehmenden Streuungen. Die Schulorganisation wird von den Akteuren in einem ähnlichen Ausmaß wahrgenommen, wie sie sich durch die formale Regelung abbildet. Die Streuung ist deutlich geringer als bei der Einschätzung der Indikatoren Personal und Finanzen. Zukünftig wünschen sich die Akteure eine geringere Steuerung durch die Schulorganisation. Diese Einschätzungen unterscheiden sich signifikant mit einer geringen Effektstärke ($d=0.36$). Das Ausmaß der staatlichen

Abbildung 3: Einschätzung der aktuellen und zukünftigen Steuerung der Governance-Regler



● Wahrnehmung derzeit ▲ Einschätzung zukünftig — Standardabweichung zukünftig

Quelle: eigene Darstellung

Inputregulierung im Unterricht wird von den Akteuren sehr ähnlich wahrgenommen wie die formale Einschätzung. Zukünftig wünschen sie sich ein signifikant geringeres Ausmaß an Regulierung im Unterricht mit einer geringen Effektstärke ($d=0.29$). Die Streuung ist wiederum relativ groß.

Das Ausmaß der staatlichen Outputregulierung durch Orientierungsarbeiten wird im Vergleich zur formalen Regelung von den Akteuren deutlich höher wahrgenommen und zukünftig noch höher gewünscht. Die zukünftige Regulierung durch Bildungsstandards sowie durch die externe Evaluation wird von den Akteuren niedriger eingeschätzt, als diese von der Begleitforschung prognostiziert wird.

Die Akteure nehmen das derzeitige Ausmaß der Befugnisse des Schulrats weniger stark wahr, als dies die formale Regelung zeigt. Die derzeitigen Befugnisse des Gemeinderats werden in einem ähnlichen niedrigen Ausmaß wie die formale Regelung wahrgenommen. Zukünftig wünschen sich die Akteure signifikant weniger Befugnisse für die Schul- und Gemeinderäte im Bereich der kommunalen Steuerung mit einer geringen Effektstärke ($d=0.30/d=0.33$).

Im Bereich der professionellen Selbststeuerung wird das Ausmaß der derzeitigen Autonomie des Lehrpersonenteams von den Akteuren im Vergleich zur formalen Regelung deutlich höher eingeschätzt. Zudem wird das Ausmaß für die zukünftige Autonomie des Lehrpersonenteams in einem signifikant höheren Ausmaß gewünscht, als es derzeit ist ($d=0.25$). Die individuelle Autonomie der Lehrpersonen wird über alle Akteure hinweg weitgehend der formalen Regelung entsprechend eingeschätzt. Zukünftig wird diese signifikant niedriger gewünscht, mit einer mittleren Effektstärke ($d=0.45$).

Das derzeitige Ausmaß der Steuerung durch die Schulleitung im Bereich der hierarchischen Selbststeuerung wird im Vergleich zur formalen Regelung nur etwas höher wahrgenommen. Die Wahrnehmungen und Einschätzungen der derzeitigen und zukünftigen Befugnisse der Schulleitung unterscheiden sich signifikant mit einer mittleren Effektstärke ($d=0.56$). Zukünftig sollen die Schulleitungen mehr Befugnisse erhalten.

Das Ausmaß der Steuerung durch Organisations- und Steuerungsinstrumente auf der Schulebene wird von den Akteuren stärker wahrgenommen, als dies die formale Regelung zeigt. Die Streuung ist relativ groß. Die derzeitige und zukünftige Einschätzung unterscheiden sich signifikant mit einer geringen Effektstärke ($d=0.38$).

Das Ausmaß der Konkurrenz innerhalb von Schulen sowie zwischen ihnen wird im Vergleich zur formalen Regelung höher wahrgenommen – im Gesamtbild jedoch immer noch relativ niedrig. Über alle Akteure hinweg wird das zukünftige Ausmaß der Konkurrenz innerhalb von Schulen sowie zwischen Schulen nur gering etwas höher gewünscht als derzeit.

5. Zusammenfassung und Interpretation der Ergebnisse

Die Ergebnisse zeigen zunächst einmal, dass die Akteure im Kanton Uri eine sehr unterschiedliche Einschätzung haben, über welche Regler mit welcher Intensität das Schulsystem gesteuert wird und zukünftig gesteuert werden sollte. Für den anstehenden Reformprozess im Zuge des Projekts „Volksschule 2016“ wäre es gewinnbringend,

in dem geplanten Vernehmlassungsprozess² diese unterschiedlichen Einschätzungen (auch zur formalen Regelung) zu thematisieren und zu ergründen (zur hierfür ziel-führenden Fragestellung siehe 6.).

Lässt man die Streuung aus dem Blick und interpretiert nur die Mittelwerte, so wird der Kanton Uri derzeitig vorrangig mithilfe einer Mischung aus kantona-ler Inputsteuerung, kommunaler sowie professioneller und hierarchischer Selbst-steuerung gesteuert. Die Ausprägung der Steuerung aller Regler befindet sich auf mittlerem Niveau.

Die zukünftige Entwicklung könnte nach Einschätzung der Akteure anders aussehen: Im Kontext der staatlichen Inputregulierung lässt sich eine Tendenz erkennen. Die befragten Akteure wünschen sich eine geringere kantonale Input-Steuerung in den Bereichen Schulorganisation und Unterricht. Dies kann als Forderung nach einer stärkeren Autonomie der Einzelschule interpretiert werden. Die Befugnisse auf der kommunalen Ebene (Gemeinde- und Schulräte) sollen ebenfalls etwas redu-ziert werden. Auf der Ebene der Einzelschule wünschen sich die Akteure eine weitere Stärkung. Die professionelle Selbststeuerung soll bezogen auf die Autonomie und die Partizipationsmöglichkeiten des Lehrpersonenteams gestärkt werden. Dagegen soll die Autonomie der einzelnen Lehrpersonen verringert werden. Dies kann im Sinne von Rolff (1995) als Gestaltungsautonomie – eine zentrale Gelingensbedingung von Schulentwicklung – verstanden werden. Dagegen sollen die hierarchischen Befugnisse der Schulleitungen nicht vergrößert werden. Eine höhere Steuerungswirkung der Instrumente der Schulentwicklung wird jedoch gewünscht. Die Außensteuerung über Bildungsstandards und externe Evaluation wird auf einem mittleren Niveau gesehen. Konkurrenz wird als Steuerungsinstrument dagegen nicht gewünscht.

Analysiert man die Steuerungskonfiguration – bestehend aus der Kombination und Stellung der sechs Regler – in Bezug auf mögliche Spannung (vgl. 2), so deut-et sich ein potentieller Konflikt bei der derzeitigen Reglerstellung der kommuna-len Steuerung (Schulrat) und der hierarchischen Selbststeuerung an. Beide sind re-lativ stark ausgeprägt, und die hierarchische Selbststeuerung befindet sich noch im Aufbau, wodurch Spannungen zusätzlich verstärkt werden könnten. Ein solches Konfliktpotential bestätigen auch die Befunde der detaillierten Analyse der derzeiti-gen Formen der Handlungskoordination (vgl. Feldhoff/Huber/Durrer 2011). Von da-her kann der Wunsch nach Reduzierung der Befugnisse des Schulrates evtl. als ein

2 „Als Vernehmlassungsverfahren wird in der Schweiz diejenige Phase innerhalb des Vorver-fahrens der Gesetzgebung bezeichnet, in der Vorhaben des Bundes von erheblicher politi-scher, finanzieller, wirtschaftlicher, ökologischer, sozialer oder kultureller Tragweite auf ihre sachliche Richtigkeit, Vollzugstauglichkeit und Akzeptanz hin geprüft werden. Die Vorlage wird zu diesem Zweck den Kantonen, den in der Bundesversammlung vertretenen Parteien, den Dachverbänden der Gemeinden, Städte und der Berggebiete, den Dachverbänden der Wirtschaft sowie weiteren, im Einzelfall interessierten Kreisen unterbreitet“ (URL: <http://www.admin.ch/dokumentation/gesetz/pc/index.html?lang=de>; Zugriffsdatum: 01.02.2012).

Indiz für eine sensible Wahrnehmung sich andeutender Konflikte interpretiert werden.

Wenn im Kanton die Steuerungskonfiguration des Schulsystems gemäß der skizzierten Wünsche geändert würde, könnten sich aus folgenden Reglerkombinationen Spannungen ergeben: einer relativ starken professionellen Selbstorganisation bei gleichzeitig relativ starker hierarchischer Selbststeuerung sowie einer relativ starken professionellen Selbstorganisation bei gleichzeitig relativ starker Außensteuerung.

6. Perspektiven für Verwaltung und Forschung im deutschsprachigen Kontext

Auf den ersten Blick wirkt eine Übertragung einer Studie in einem kleinen Schweizer Bergkanton auf die Entwicklung und Forschung im Kontext der Systemsteuerung in Deutschland oder Österreich befremdlich.

Richtet man jedoch den Blick auf die Analyse der Steuerungskonfiguration mit Hilfe der Governance-Regler, so ergeben sich für die Bildungsforschung und -verwaltung im Kontext des Umbaus der Steuerungssysteme neue und spannende Perspektiven. Das erweiterte Set an Governance-Reglern eignet sich für die Analyse vieler Schulsysteme (vgl. Altrichter 2010 für Österreich), wenn auch Modifikationen auf der Ebene der einzelnen Indikatoren hierfür notwendig sind. Aus den Befunden lassen sich folgende Perspektiven und weitergehenden Fragen für Forschung und Verwaltung ableiten:

1. Eine Analyse der Schulgesetze mit Hilfe der Governance-Regler ermöglicht zunächst einmal, auf deskriptiver Ebene einen systematischen Überblick über die Gesamtheit der Maßnahmen zu gewinnen, in welcher Form, auf welcher Ebene (Regler), mit welcher Ausprägung ein Schulsystem (formal) gesteuert wird. Solche Analysen sind bisher in Forschung und Verwaltung selten.
2. Ein anschließender Abgleich der formalen Regelung mit Einschätzungen der beteiligten Akteure (quantitativ und/oder qualitativ) bietet die Möglichkeit, tiefere Erkenntnisse über die Steuerung und deren Wirkung zu gewinnen: 1. In Bezug auf die Diskrepanz zwischen formaler und wahrgenommener Regelung: Wie lassen sich starke Abweichungen zwischen formaler und wahrgenommener Steuerung mit Hilfe einzelner Regler erklären? Sind hierfür lokale Anpassungsleistungen auf den einzelnen Ebenen verantwortlich? Verstärken Anpassungsleistungen die Wirkungen oder stehen sie diesen entgegen? 2. In Bezug auf die Diskrepanz zwischen den Einschätzungen der Akteure oder Akteursgruppen: Sind Unterschiede „nur“ auf die verschiedenen Blickwinkel zurückzuführen oder wirken Maßnahmen anders, als intendiert bzw. postuliert wird? Lassen sich regions- oder schulformspezifische Unterschiede in der Wahrnehmung fest-

stellen, die evtl. darauf hindeuten, dass die Maßnahmen je nach lokalem Kontext andere Wirkungen erzielen?

3. Ein besonders großes Potential der Governance-Regler liegt in der Analyse der Steuerungskonfigurationen in ihrer Gesamtheit und des Verhältnisses der einzelnen Regler zueinander: Welche Zusammenhänge sehen Verwaltungsexperten zwischen einzelnen Steuerungsmaßnahmen und Reglern? Welche Maßnahmen und Regler ergänzen und komplementieren sich? Welche Maßnahmen rufen Widersprüche bzw. Spannungen hervor? Welche Auswirkungen können diese haben? Sind sie temporärer Natur im Zuge des Umbaus oder dauerhaft? Wie gehen Akteure mit möglichen Spannungen um? Werden auf der Basis von Spannungen Modifikationen vorgenommen oder Konsequenzen gezogen?
4. Neben der Analyse einzelner Schulsysteme wäre gerade für die Forschung ein Vergleich verschiedener Schulsysteme anhand der Governance-Regler von Bedeutung. Neben der Identifizierung von Gemeinsamkeiten und Unterschieden wäre vor allem die Frage gewinnbringend, inwieweit gleiche Maßnahmen in unterschiedlichen Systemen zu unterschiedlichen Wirkungen führen bzw. unterschiedliche Maßnahmen zu gleichen Wirkungen und worauf sich diese zurückführen bzw. wodurch sie sich erklären lassen.

Literatur

- Altrichter, H. (2010): Lehrerfortbildung im Kontext von Veränderungen im Schulwesen. In: Müller, F.H./Eichenberger, A./Lüders, M./Mayr, J. (Hrsg.): Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung. Münster u.a.: Waxmann, S. 17-34.
- Altrichter, H./Brüsemeister, T./Wissinger, J. (Hrsg.) (2007): Educational Governance: Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS.
- Altrichter, H./Soukup-Altrichter, K. (2008): Changes in Educational Governance Through More Autonomous Decision-Making and In-school Curricula? In: International Journal of Contemporary Sociology 45, H. 2, S. 33-48.
- Benz, A. (Hrsg.) (2004): Governance – Regieren in komplexen Regelsystemen. Wiesbaden: VS.
- Braun, D. (1999): New Managerialism and the Governance of Universities in a Comparative Perspective. In: Braun, D./Merrien, F.-X. (Hrsg.): Towards a New Model of Governance for Universities? A Comparative View. London: Jessica Kingsley, S. 239-261.
- Brüsemeister, T. (2007): Steuerungsakteure und ihre Handlungslogiken im Mehrebenensystem der Schule. In: Kussau, J./Brüsemeister, T. (Hrsg.): Governance, Schule und Politik. Zwischen Antagonismus und Kooperation. Wiesbaden: VS, S. 62-96.
- Büeler, X. (2007): School Governance – Die Fallstudie Luzern. In: Altrichter, H./Brüsemeister, T./Wissinger, J. (Hrsg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS, S. 131-155.
- Clark, B.R. (1998): Creating Entrepreneurial Universities. Organizational Pathway of Transformation. Oxford: Pergamon Press.
- Fend, H. (2008): Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität. Wiesbaden: VS.

- Feldhoff, T./Huber, S.G./Durrer, L. (2009): Expertise zur Steuerung und Handlungs-
koordination im Volksschulsystem des Kantons Uri. URL: [http://www.bildungsma-
nagement.net/pdf/Poster_als_pdf-Download/forschungsprojekte/FeldhoffHuberDurrer-
Uri2016-SteuerungundHandlungskoordinationimVolksschulsystem.pdf](http://www.bildungsma-
nagement.net/pdf/Poster_als_pdf-Download/forschungsprojekte/FeldhoffHuberDurrer-
Uri2016-SteuerungundHandlungskoordinationimVolksschulsystem.pdf); Zugriffsdatum:
16.01.2012.
- Feldhoff, T./Huber, S.G./Durrer, L. (2011): Steuerung durch Handlungskoordination im
Schulwesen – eine empirische Analyse. In: Aufenanger, S./Hamburger, F./Ludwig,
L./Tippelt, R. (Hrsg.): Bildung in der Demokratie. Beiträge zum 22. Kongress der
Deutschen Gesellschaft für Erziehungswissenschaft, Band 2. Opladen: Barbara Budrich,
S. 69-81.
- Kussau, J. (2005): Schulpolitik, schulische Governance und Schule. Programmatische
Notizen. Bochum/Konstanz: Progos – Projektgruppe Governance in Schulen. URL:
<http://www.progos-institute.com/Kussau%20Schulpolitik%20und%20Governance.pdf>;
Zugriffsdatum: 01.02.2012.
- Kussau, J./Brüsemeister, T. (2007): Educational Governance: Zur Analyse der Hand-
lungskoordination im Mehrebenensystem der Schule. In Altrichter, H./Brüsemeister,
T./Wissinger, J. (Hrsg.): Educational Governance. Handlungskoordination und
Steuerung im Bildungssystem. Wiesbaden: VS, S. 15-54.
- Luhmann, N. (1964): Lob der Routine. In: Luhmann, N. (Hrsg.): Politische Planung.
Opladen: Westdeutscher Verlag, S. 113-142.
- Mayntz, R. (1998): New Challenges to Governance Theory. Jean Monnet Chair Papers 50.
In: Bang, H. (Hrsg.): Governance, Governmentality and Democracy. Robert Schuman
Center at the European University Institute, Badia Fiesolana. Manchester: University
Press.
- Mayring, P. (2008): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim/
Basel: Beltz.
- Rolff, H.-G. (1995): Autonomie als Gestaltungsaufgabe. Organisationspädagogische Perspek-
tiven. In: Daschner, P./Rolff, H.-G./Stryck, T. (Hrsg.): Schulautonomie – Chancen und
Grenzen. Weinheim u.a.: Juventa, S. 31-54.
- Schimank, U. (2007): Die Governance-Perspektive: Analytisches Potenzial und anstehen-
de konzeptionelle Fragen. In: Altrichter, H./Brüsemeister, T./Wissinger, J. (Hrsg.):
Educational Governance. Handlungskoordination und Steuerung im Bildungssystem.
Wiesbaden: VS, S. 231-260.
- von Kopp, B. (2008): Bildungssteuerung: Vom Drehen an der Stellschraube zu Governance.
(Trends in Bildung international – TiBi 19.) Frankfurt a.M.: DIPF. URL: [http://www1.
dipf.de/publikationen/tibi/tibi19_kopp.pdf](http://www1.
dipf.de/publikationen/tibi/tibi19_kopp.pdf); Zugriffsdatum: 01.02.2012.

Tobias Feldhoff, Prof. Dr., geb. 1976, Juniorprofessor für Erziehungswissenschaft
mit dem Schwerpunkt empirische Bildungsforschung und Schulentwicklung am
Deutschen Institut für Internationale Pädagogische Forschung (DIPF) und an der
Goethe Universität Frankfurt am Main.

Anschrift: Schloßstraße 29, 60486 Frankfurt a.M.
E-Mail: feldhoff@dipf.de

Luzia Durrer, lic. phil., geb. 1982, Beauftragte für Bildung, Kinder und Jugend im Amt für Bildung, Kultur und Sport der Stadt Langenthal (Schweiz), bis 12/2011 wissenschaftliche Assistentin am Institut für Bildungsmanagement und Bildungsökonomie der Pädagogischen Hochschule Zentralschweiz Zug.

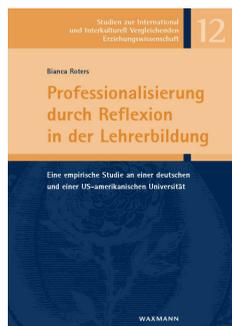
Anschrift: Jurastrasse 22, 4901 Langenthal, Schweiz
E-Mail: luzia.durrer@langenthal.ch

Stephan Gerhard Huber, Prof. Dr., geb. 1971, Leiter des Instituts für Bildungsmanagement und Bildungsökonomie (IBB) der Pädagogischen Hochschule Zentralschweiz (PHZ) Zug; Mitglied der Erfurt School of Education (ESE).

Anschrift: Institut für Bildungsmanagement und Bildungsökonomie (IBB), Pädagogische Hochschule Zentralschweiz (PHZ) Zug, Zugerbergstrasse 3, 6300 Zug, Schweiz
E-Mail: Stephan.Huber@phz.ch

WAXMANN

In einer vergleichend angelegten Studie werden die Professionalisierungsdiskurse in Deutschland und den USA anhand von Dokumentenanalysen, Interviews mit Lehrerausbildern und dokumentierten studentischen Reflexionen untersucht. Zentrale Ergebnisse der Studie sind sechs empirisch nachweisbare Reflexionsniveaus, durch die in Form einer Typologie eine Charakterisierung studentischer Reflexionsfähigkeit vom deskriptiv-pauschalisierenden bis hin zum transformativ-reflexiven Novizen möglich ist. Anhand der empirischen Befunde werden hochschuldidaktische Empfehlungen für ein reflexives Professionalisierungskonzept in der Lehrerbildung ausgesprochen.



Bianca Roters
**Professionalisierung
durch Reflexion
in der Lehrerbildung**
Eine empirische Studie an
einer deutschen und einer
US-amerikanischen Universität
2012, Studien zur International
und Interkulturell Vergleichenden
Erziehungswissenschaft, Bd. 12,
328 S., br. 34,90 €,
ISBN 978-3-8309-2662-7

Waxmann
order@waxmann.com
www.waxmann.com

Uwe Maier/Harm Kuper

Vergleichsarbeiten als Instrumente der Qualitätsentwicklung an Schulen Überblick zum Forschungsstand

Zusammenfassung

Vergleichsarbeiten wurden vor ca. zehn Jahren in Deutschland als Qualitätsentwicklungsmaßnahme eingeführt und sollen vor allem auf Schul-, Fachkonferenz- und Lehrerebene zur Reflexion über und Verbesserung von Unterricht genutzt werden. Qualitätsentwicklung und -sicherung durch zentrales Testen ist jedoch keine neue Idee. In diesem Überblicksartikel werden Ergebnisse der internationalen und deutschsprachigen Literatur zu Effekten zentraler Tests zusammengefasst und Implikationen für die Weiterführung der deutschsprachigen Forschung in diesem Bereich diskutiert.

Schlüsselwörter: Vergleichsarbeiten, Leistungsmessung, Qualitätsentwicklung, Schulentwicklung, Unterrichtsentwicklung, Evaluation

Mandatory Testing as an Instrument for School Quality Development

A Research Survey from an International Perspective

Abstract

Ten years ago, German states implemented mandatory testing policies which aim at instructional improvement on school, department and classroom level. However, the concept of test-based quality assurance and quality development is not a new idea. This research survey summarizes main empirical findings on how mandatory testing affected teaching and student learning in different educational contexts. The paper eventually discusses implications for further research on mandatory testing effects in Germany.

Keywords: mandatory testing, student assessment, quality development, school improvement, school evaluation

1. Vergleichsarbeiten als Schulreforminstrument

Vergleichsarbeiten sind Teil einer Gesamtstrategie der Kultusministerkonferenz für das Bildungsmonitoring. Ihnen liegt die Idee zugrunde, nicht nur auf internationaler bzw. bildungspolitischer Ebene Leistungsdaten zur Qualitätssicherung und Qualitätsentwicklung zu nutzen; vielmehr möchte man mit Instrumenten wie den Vergleichsarbeiten die von vielen Bildungspolitikern und -politikerinnen als erfolgreich wahrgenommene Logik einer datengestützten Qualitätsentwicklung durch Leistungsmessung flächendeckend auf alle Schulen eines Bundeslandes übertragen.

Die Idee, mit zentralen Leistungstests Schule oder Unterricht zu verbessern oder gar die Schüler und Schülerinnen zu vertieftem und nachhaltigem Lernen zu motivieren, ist nicht neu. Zunächst einmal muss man die historische Tatsache zur Kenntnis nehmen, dass die zentrale Messung und Zertifizierung von Schülerleistungen bereits immer ein besonderes Qualitätsmerkmal eines Schulsystems war. Die Attraktivität des Gymnasiums für das aufstrebende Bürgertum im 19. Jahrhundert war wesentlich an die Einführung des Abiturs als Zulassungsvoraussetzung für ein Studium gekoppelt. Asiatische Länder haben eine Jahrtausende alte Tradition des zentralen Prüfens und Selektierens, um die besten Köpfe für den Staatsdienst zu gewinnen. In den USA spielen zentrale Abschluss- oder Eingangstests seit über 100 Jahren eine Rolle für den akademischen Karriereweg.

International vergleichende Studien zeigen immer wieder, dass diese Form von testbasierter *student accountability* (Schüler und Schülerinnen werden getestet und tragen die Konsequenzen des Testergebnisses) in der Regel zu mehr Anstrengung (bei Schülern und Schülerinnen, Eltern und Lehrkräften) und besseren Leistungsergebnissen (wiederum gemessen mit zentralen Vergleichstests) führt (vgl. z.B. Bishop 1995; Wößmann 2007). Vergleichsarbeiten folgen jedoch nicht der Logik von zentralen Abschlussprüfungen. Adressaten der Testrückmeldungen sind vielmehr Schulen, Fachkonferenzen und einzelne Lehrkräfte (*school accountability*). Diese sollen die Ergebnisse analysieren und daraus Konsequenzen für die Optimierung des Unterrichts ziehen. Diese Funktion zentraler Tests findet man erstmals in den USA und in England in den 1980er-Jahren. In dieser Blütezeit des Neoliberalismus hielten verstärkt Ideen der leistungsorientierten Gestaltung öffentlicher Dienste Einzug. Auch Schulen sollten sich an harten Leistungsdaten orientieren und diese als Ausgangspunkt für Veränderungen heranziehen. Diese Denkweise gipfelte in der *No-Child-Left-Behind*-Gesetzgebung (NCLB) (2001) aus der Regierungszeit von G.W. Bush. Alle US-Bundesstaaten wurden zur Einführung von testbasierter Rechenschaftslegung für Schulen verpflichtet, und mit ausgefeilten Sanktionssystemen wurde versucht, Lehrkräfte zur Arbeit zu motivieren.

Dieser Hintergrund ist wichtig, um verschiedene Forschungsbereiche zur Thematik verstehen und ihre Relevanz für die Bewertung von Vergleichsarbeiten in Deutschland (z.B. VERA) einschätzen zu können. In Deutschland ist man zwar weit entfernt von harten Sanktionen für Schulen, die unterdurchschnittliche VERA-Testergebnisse aufweisen. Dennoch sollte man zur Kenntnis nehmen, dass zumindest der Schulverwaltung VERA-Testergebnisse auf Schul- und auch Lehrerebene bekannt sein dürften. Ebenso gehört zur VERA-Logik, dass Schulleitungen oder Fachkonferenzleiter und -leiterinnen die Testergebnisse kennen und damit Unterrichtsentwicklung anstoßen sollen. Damit richten sich an deutsche Vergleichsarbeiten und „high-stakes tests“ in den USA oder England prinzipiell die gleichen Funktionserwartungen: Sie sollen dazu dienen Qualität zu sichern und gleichzeitig Lehrkräfte zur Verbesserung des Unterrichts stimulieren. Verschieden sind jedoch die angestrebten Steuerungsmechanismen, die im Falle der „high-stakes tests“ auf administrativ induzierte Leistungskonkurrenz und im Falle der Vergleichsarbeiten auf die informationelle Unterstützung professioneller Schulentwicklung setzen.

2. Forschungsstand

Vor dem Hintergrund der Eingangüberlegungen kann man sagen, dass über die Effekte zentraler Tests in einem „high-stakes“ Kontext bereits sehr viel geforscht wurde. Auch wenn diese bildungspolitischen Regelungsstrukturen nicht direkt auf deutsche Verhältnisse übertragbar sind, sollte man diese Befunde für eine Einschätzung von standardbasierten Vergleichsarbeiten im „low-stakes“ Kontext Deutschland dennoch zur Kenntnis nehmen (2.1). Anschließend werden Studien skizziert, die sich mit der Frage beschäftigen, ob und wie über zentrale Leistungsmessungen didaktische Innovationen transportiert werden können (2.2). In einem dritten Schritt geht es um Forschungsarbeiten, die Effekten von Testrückmeldungen auf Schulebene nachgehen (2.3). Abschließend werden zentrale Ergebnisse der deutschen Forschung zu Vergleichsarbeiten skizziert.

2.1 Forschung zu zentralen Tests im „High-Stakes“-Kontext

In einer Vielzahl empirischer Untersuchungen wurden negative und nicht erwünschte Effekte von „high-stakes testing“ untersucht (vgl. Stecher 2002; Cheng/Curtis 2004; Herman 2004). Die große Mehrheit der Studien entstand im Zusammenhang mit der ersten Welle staatlicher Tests, die als Folge allgemeiner Unzufriedenheit mit dem US-Schulsystem ab den 1970er-Jahren sukzessive eingeführt wurden. Weitere Studien beschreiben die nicht intendierten, negativen Konsequenzen der Testsysteme von US-Bundesstaaten im Rahmen der aktuellen NCLB-Gesetzgebung.

Ein empirisch sehr gut abgesichertes Ergebnis ist die Reduktion der verfügbaren Unterrichtszeit durch Testvorbereitungsaktivitäten. Lehrkräfte setzen veröffentlichte Aufgaben früherer Tests oder kommerzielle Materialien ein, um Schüler und Schülerinnen auf die Prüfungen vorzubereiten. Nicht getestete Fächer und Inhalte verlieren an Bedeutung bzw. werden gestrichen. Analog dazu führen Tests mit weitreichenden Sanktionen für Lehrkräfte zu einer verstärkten Fokussierung auf test-spezifische Inhaltsgebiete eines Faches. Für die Testvorbereitung werden vor allem simple, lehrergelenkte Unterrichtsmethoden und testähnliche Aufgabenformate eingesetzt (vgl. Ketter/Pool 2001; Vogler 2005; Assaf 2006; Grant 2007; Olson 2007; Valli/Chambliss 2007; Watanabe 2007).

Trotz des Anspruchs, mit testbasierter Rechenschaftslegung die soziale Selektivität zu reduzieren, gibt es in den USA einen konstanten „test score gap“: Schüler und Schülerinnen ethnischer Minderheiten und sozioökonomisch benachteiligter Gruppen erreichen im Schnitt wesentlich niedrigere Werte bei zentralen Leistungstests als ihre weißen Mitschüler und Mitschülerinnen aus Mittelschichtfamilien. Zahlreiche Analysen zeigen, dass „high-stakes tests“ dabei eher zu einer Verstärkung sozialer Disparitäten beitragen (vgl. z.B. Sloan 2007; Diamond 2007). Wie in einem Teufelskreis sind diese Schulen dann immer weniger attraktiv für Eltern und qualifizierte Lehrkräfte. Studien wiesen auch die Korrumpierung der externen Leistungsindikatoren nach, beispielsweise durch eine clevere Kategorisierung von Schülern und Schülerinnen. Um Testwertsteigerungen zu erhalten, wurden Angehörige von Minderheitengruppen in sog. „disability programs“ überführt.

Eine Serie von Forschungsarbeiten beschäftigte sich in den USA mit den Effekten zentraler Leistungstests auf Schülerleistungen (deutschsprachige Zusammenfassung: Maier 2010). Ausgangspunkt dieser Studien war die Arbeit von Amrein und Berliner (2003). Sie teilten die US-Bundesstaaten in „high-stakes testing“ (HST)-Staaten und „low-stakes testing“ (LST)-Staaten und verglichen die Schülerleistungszuwächse in diesen beiden Gruppen auf Basis nationaler Vergleichsdaten. Das Fazit dieser Analyse war ernüchternd: Die Werte der HST-Staaten waren in nationalen Längsschnittstudien (z.B. *National Assessment of Educational Progress*: NAEP) insgesamt rückläufig. Rosenshine (2003) reagierte mit einer kritischen Reanalyse dieser Studie durch Hinzunahme einer Kontrollgruppe von 15 LST-Staaten. Rosenshine fand einen moderaten Leistungszuwachs bei den NAEP 4-Mathematiktests und einen substanziellen Effekt bei den NAEP 8-Mathematik- und NAEP 4-Lesetests jeweils zugunsten der HST-Staaten. Weitere Forschergruppen folgten diesem Design und legten jeweils auf Basis unterschiedlicher Leistungsvergleichsdaten und mit unterschiedlicher forschungsmethodischer Präzision Forschungsergebnisse vor (vgl. Carnoy 2005; Lee/Wong 2004). Diese Studien weisen auf differentielle Effekte hin. Eine Kombination aus hohem Rechenschaftsdruck auf Schulebene und zentralen Leistungsmessungen in Abhängigkeit von Randbedingungen in einzelnen Teilbereichen kann zu Schülerleis-

tungssteigerungen führen. Letztlich wird damit ein empirischer Beleg für die Kontingenz der Wirkungen testbasierter Schulreform gegeben.

2.2 Unterrichtsreformen über zentrale Tests

Die zum Teil massive Kritik an negativen Konsequenzen von „high-stakes testing“ auf der Basis eines zu eng gefassten Leistungsbegriffs führte in zahlreichen US-Bundesstaaten zu Testreformen und einer neuen Generation von Testaufgaben. Dabei wurden auch aus administrativer Sicht die Tests immer mehr als Hebel für die positive Veränderung der Unterrichtspraxis betrachtet (vgl. Popham 1987; Cheng 1999; Stecher 2002). Diese Form der staatlichen Standards und Tests unterscheidet sich von Vorgängerversionen in der Regel durch folgende Merkmale: (1) Die Tests sind mit Bildungsstandards verknüpft, die neben Grundfertigkeiten auch höherwertige Denkprozesse und Informationsverarbeitungsstrategien betonen. (2) Zunehmend finden alternative Aufgabenformate Einzug in die standardisierten Tests („performance based assessment“): offene Fragen, Textproduktion, Experimente oder Portfolios. (3) Die Tests sollen den für das Lehren Verantwortlichen eine Rückmeldung geben und damit den Unterricht positiv beeinflussen.

Die intendierten Effekte dieser neuen Generation staatlich verordneter Leistungsmessungen (z.B. in Kentucky oder Maryland) wurden in einer Reihe von Studien aus den Bereichen „general education“ und angewandter Linguistik („washback“-Studien) untersucht (vgl. McDonnell/Choisser 1997; Stecher/Barron 2001; Faulkner/Cook 2006; Parke/Lane/Stone 2006; Firestone/Winter/Fitz 2000). Diese Studien lassen sich folgendermaßen zusammenfassen: Es gibt nachweisbare, intendierte Effekte staatlicher Testreformen auf Unterrichtsprozesse. Lehrkräfte nutzen innovative Schreib- oder Problemlöseaufgaben auch in ihrem regulären Unterricht. Allerdings zeigten die Studien ebenso, dass dieser Reformimpuls deutlich begrenzt ist. Es gab keine empirischen Belege dafür, dass ein nennenswerter Anteil der Lehrerschaft aufgrund fachdidaktisch innovativer Testsysteme ihre didaktischen Ansätze grundsätzlich überdenkt. Auch die „washback“-Forschung zu Testreformen in China und Hong Kong kommt zu ähnlichen Schlüssen (vgl. Cheng 1999, 2003; Luxia 2007).

2.3 Studien zum schulinternen Umgang mit Testrückmeldungen

In Studien zu Effekten zentraler Tests auf institutioneller Ebene werden Kommunikationsstrukturen und die Nutzung der Rückmeldungen zur Ableitung von Verbesserungsmaßnahmen innerhalb von Einzelschulen analysiert (vgl. Yang u.a. 1999; Hayes/Rutt 1999; Wikeley/Stoll/Lodge 2002; Demie 2003). Ein wichtiges Ergebnis dieser Literatur ist, dass die schulinterne Nutzung externer Leistungsdaten von außen gezielt unterstützt werden muss. Wikeley, Stoll und Lodge (2002) berichten beispiels-

weise über Lehrkräfte, die aufgrund der Teilnahme an spezifischen Fortbildungen zunehmend in die Lage versetzt wurden, externe Leistungsdaten schüler- und schülergruppenspezifisch zu interpretieren. Daraufhin konnten Programme für spezifische Schülergruppen entwickelt und durchgeführt werden. Wenn allerdings aufgrund eines schmalen Projektbudgets keine Fortbildungen und Besprechungen stattfanden, konnten Schulen mit ohnehin geringen Veränderungskapazitäten die externen Leistungsinformationen nicht nutzen.

Die vertrauensvolle Zusammenarbeit mit den lokalen Schulbehörden ist eine ebenso wichtige Voraussetzung für eine sinnvolle Dateninterpretation auf Schulebene (vgl. Rudd/Davies 2002). Als besonders hilfreich stellte sich heraus, wenn lokale Schulbehörden sinnvolle Leistungsindikatoren auswählen, mit weiteren organisationalen Variablen verknüpfen und in lesbarer Form den Schulen zur Verfügung stellen (vgl. Demie 2003; Louis/Febey/Schroeder 2005). Die Schnittstelle dieser Zusammenarbeit zwischen Schulverwaltung und Einzelschule sind Experten und Expertinnen, die sich auf die Dateninterpretation und -nutzung spezialisiert haben (vgl. Hayes/Rutt 1999).

Auch innerhalb der Einzelschule kommt es darauf an, dass Lehrkräfte mit einer gewissen statistischen Expertise bei der Übersetzung und Interpretation von Daten helfen können (vgl. Yang u.a. 1999; Wikeley/Stoll/Lodge 2002). Fachabteilungsleitern und -leiterinnen muss es gelingen, fachdidaktisch orientierte Visionen im Kollegium zu bündeln und diese mit den Implikationen externer Leistungsindikatoren zu verknüpfen. Zentrale Testergebnisse werden auf Abteilungsebene vor allem dann systematisch analysiert, wenn sich das Lehrerkollegium bereits über die Notwendigkeit von Reformmaßnahmen verständigt hat. Das generelle Innovationsklima an Schulen ist eine weitere, wichtige Kontextvariable. In Schulen mit geringen Veränderungskapazitäten beispielsweise standen die entwickelten Testrückmeldesysteme recht unverbunden anderen Initiativen der Qualitätsentwicklung gegenüber.

Es gibt allerdings auch empirische Studien, die grundsätzlich in Frage stellen, dass Rückmeldungen aus „high-stakes tests“ valide Indikatoren für Schulqualität darstellen und von Lehrkräften zur Weiterentwicklung von Unterricht genutzt werden können. Mintrop und Trujillo (2007) kommen in einer mehrperspektivischen Studie zu dem Ergebnis, dass die absoluten Testwerte des kalifornischen Leistungstests in keiner systematischen Verbindung mit weiteren Kriterien zur Schul- und Organisationsqualität stehen. Ähnlich kritisch argumentieren Ingram, Louis und Schroeder (2004). Die Autoren untersuchten ebenfalls in einer mehrperspektivischen Studie datenbasierte Entscheidungsprozesse in US-amerikanischen *high schools*, die Teilnehmer eines staatlichen Qualitätsmanagements sind und als „best practice“-Schulen ausgewiesen wurden. Selbst in diesen Schulen sind Diskrepanzen zwischen professionellem Wissen und Evaluationswissen deutlich sichtbar. Lehrer und Lehrerinnen haben in der Regel einen eigenen Qualitätsmaßstab für ihren Unterricht, der sich deutlich von

externen Indikatoren unterscheidet und auf anekdotischer Evidenz und informellem Wissen basiert.

2.4 Deutschsprachige Forschung zur Rezeption und Nutzung von Vergleichsarbeiten

Mit der Einführung von Vergleichsarbeitsprojekten auf Länderebene entstand eine eigenständige Forschungslinie, die sich mit der Akzeptanz, Rezeption und Nutzung von Rückmeldedaten durch Lehrkräfte und Schulen beschäftigt.

Im Rahmen der Thüringer Kompetenztests wurden beispielsweise regelmäßige Fragebogenstudien zur Evaluation der Nutzung von Rückmeldedaten durchgeführt (vgl. Nachtigall/Jantowski 2007). Es zeigte sich, dass die organisatorisch-technische Realisierung der Dateneingabe und -auswertung für Lehrkräfte in der Regel kein größeres Problem darstellt. Durch die Einführung von Sofortrückmeldungen konnte die Nutzung der Leistungsdaten erhöht werden. Lehrkräfte berichteten ebenfalls über eine Verlagerung von unterrichtlichen Inhalten, eine Anpassung von Methoden und eine Intensivierung der Zusammenarbeit im Kollegium als Folge der Kompetenztests. Allerdings sehen Lehrer und Lehrerinnen den Nutzen der Kompetenztests vorrangig im Bereich der Leistungsdiagnostik und des Leistungsvergleichs. Als Evaluationsinstrument für den Unterricht werden sie nur teilweise wahrgenommen und genutzt.

Auch in Online-Befragungen zur Nutzung und Rezeption der nordrhein-westfälischen Lernstandserhebungen in der Sekundarstufe I äußern die Lehrkräfte insgesamt eine hohe Bereitschaft, sich mit den Daten auseinanderzusetzen (vgl. z.B. Kühle/Peek 2007; Kühle 2010). Die Bereitschaft zur nutzungsorientierten Rezeption der Ergebnismeldungen hängt allerdings von der eingeschätzten Nützlichkeit und von der Akzeptanz ab und nicht von der Verständlichkeit der Rückmeldungen bzw. der Intensität der Auseinandersetzung. Auch im Rahmen des Verbundprojektes VERA 3/4 wurden regelmäßige Online-Lehrerbefragungen zur Rezeption und Nutzung der Testergebnisse durchgeführt (vgl. z.B. Groß Ophoff/Hosenfeld/Koch 2007). Die Vergleichsarbeiten wurden von den Grundschullehrkräften als informativ wahrgenommen, und die Durchführung sowie Auswertung bereiteten kaum Probleme. Im Vordergrund stand aus Lehrersicht die Ableitung von geeigneten Fördermaßnahmen für einzelne Schüler und Schülerinnen. Nur in einem geringeren Maße wurde die Bedeutung der Tests für die Reflexion des eigenen Unterrichts und eine ergebnisorientierte Unterrichtsentwicklung gesehen.

Die bis hierher referierten Befunde entstanden im Umkreis der Vergleichsarbeitsprojekte und können aus diesem Grund als „usability“-Studien bzw. Evaluationsvorhaben bezeichnet werden. Hierzu gehört beispielsweise auch die von Koch (2011) entwickelte und evaluierte Fortbildungsveranstaltung zur Verbesserung der

Datenkompetenz von Grundschullehrkräften. Darüber hinaus gibt es weitere empirische Arbeiten, die Rezeption und Nutzung von Leistungsrückmeldungen aus einer externen Perspektive analysieren.

Maier (2007, 2008a) führte eine qualitative Studie zur Nutzung von Rückmeldedaten in Baden-Württemberg durch. Lehrkräfte an Sekundarschulen in Baden-Württemberg nutzten die Vergleichsarbeitsrückmeldung vor allem als weitere Klassenarbeitsnote. Eine formativ-diagnostische Nutzung wurde aufgrund der eingeschränkten curricularen Validität des Tests und des ungünstigen Testzeitpunkts abgelehnt. Die von Lehrkräften in den Interviews berichteten Unterrichtsentwicklungsmaßnahmen waren eher rudimentär und zufällig. In einer weiteren Studie fand Maier (2008b) deutliche Unterschiede bei der Rezeption und Nutzung von Leistungsrückmeldungen in Abhängigkeit verschiedener Vergleichsarbeitsysteme. Bei den Kompetenztests in Thüringen (Klasse 6 und 8) erhielten die Lehrkräfte im Vergleich zu den baden-württembergischen Vergleichsarbeiten (Klasse 6) kriteriale und faire Vergleiche als Rückmeldung. Es zeigte sich, dass Thüringer Lehrkräfte sowohl die Akzeptanz, die curriculare Validität als auch die Nutzung der Leistungsdaten für Lerndiagnosen und die zukünftige Unterrichtsplanung höher einschätzten. In Thüringen waren die Leistungsrückmeldungen auch wesentlich häufiger Gegenstand systematischer Diskussionen in Fach- und Gesamtlehrerkonferenzen.

Kuper und Hartung (2007) fragten vor einem professions- und organisationstheoretischen Hintergrund, welche Rezeptions- und Deutungsmuster bei Lehrkräften im Umgang mit Vergleichsarbeitsrückmeldungen zu beobachten sind. Hierzu wurden „Orientierung an technologischen Gründen (empirisches Wissen)“ und „Orientierung an normativen Gründen“ theoretisch unterschieden. Mit qualitativen Daten (Fallanalysen in NRW zu Lernstand 9) zeigten Kuper und Hartung, dass beide Überzeugungen in Bezug auf Lernstandserhebungen vorkommen. Es kommt sowohl zur retrospektiven Ursachenzuschreibung durch Leistungsrückmeldungen (technologisches Handeln) als auch zur Betonung der Bedingungen professionellen Handelns (Ablehnung des Schlusses von Leistungsergebnissen auf Unterrichtsqualität). Kuper und Hartung gehen in ihrer Studie allerdings nicht von einer direkten Beeinflussung des pädagogischen Interaktionsverhältnisses aus. Vielmehr wäre denkbar, dass Lernstandserhebungen die Reflexion professioneller pädagogischer Arbeit organisatorisch unterstützen.

3. Fazit und Forschungsperspektiven

Der Überblick zum Forschungsstand zeigt eine Vielzahl von Ergebnissen auf, die institutionelle und organisatorische Kontingenzen der Rezeption und Verwendung von Ergebnissen aus zentralen Leistungsüberprüfungen in den Schulen belegen. Eine professions-, organisations- und/oder institutionentheoretische Rahmung empirischer Projekte, die sich mit den Folgen testbasierter Schulreform befassen, erscheint daher geboten und aussichtsreich. Eine Fokussierung auf die Auswirkungen schulübergreifender institutioneller Regelstrukturen ist insbesondere in der US-amerikanischen Forschung erfolgt. Diese Forschung ist deutlich mit den Folgen von „high-stakes testing“ befasst; die Ergebnisse sind daher auf die Konzepte testbasierter Schulreform in den deutschen Ländern nicht unmittelbar übertragbar. Wohl aber sind ihnen theoretische und methodische Anregungen zu entnehmen. Darüber hinaus regt der internationale Vergleich der Befunde institutionentheoretische Interpretationen der Auswirkungen regulativer Rahmenbedingungen auf die Verwendung der Ergebnisse in den Schulen an. Der deutsche Forschungsstand nimmt diese Anregungen zum Teil auf, enthält aber allenfalls ansatzweise und kaum systematische Ergebnisse, die über die Betrachtung der Schulebene hinausgehen. Forschung zu den Folgen testbasierter Schulreform im Kontext der institutionellen Strukturen eines Mehrebenensystems ist ein Desiderat.

Das von den beiden Autoren geleitete Forschungsprojekt „Testbasierte Schulreformen im Mehrebenensystem des Schulsystems“ reagiert auf dieses Forschungsdefizit. Das Projekt wird vom BMBF im Rahmen der Ausschreibung „Steuerung im Bildungssystem“ gefördert. Ein Ziel ist die Untersuchung des Zusammenspiels verschiedener Akteure und Akteursgruppen bei der Rezeption und Nutzung von Vergleichsarbeitsrückmeldungen in der Organisation Schule. Die Daten werden in zusammenhängenden Organisations- und Verwaltungseinheiten erhoben (Schulen in einem Schulverwaltungsbezirk; Interviews mit Lehrkräften, Fachgruppenleitern und -leiterinnen und Schulleitungen innerhalb einer Schule); die Leitfadenterviews arbeiten Relationen zwischen den Rezeptions- und Nutzungsformen auf unterschiedlichen Ebenen heraus. Ein zweites Ziel ist die Klärung der Frage, inwiefern sich institutionelle Regelungskontexte auf die Rezeption und Nutzung von Vergleichsarbeitsdaten auswirken. Um dieser Fragestellung nachgehen zu können, führen wir die qualitativen Befragungen an Gymnasien in vier Bundesländern (Berlin, Brandenburg, Baden-Württemberg und Thüringen) durch. Wir gehen davon aus, dass die institutionelle Umwelt einer Einzelschule sowohl durch die bundeslandspezifischen Regelungen zu Vergleichsarbeiten als auch durch eher implizite Vorgaben und Botschaften der lokalen Schulverwaltung geprägt wird. In vergleichenden Fallstudien soll herausgearbeitet werden, inwiefern diese Regelungskontexte die Rezeption und Nutzung von Rückmeldedaten an einzelnen Schulen bedingen.

Literatur

- Amrein, A.L./Berliner, D.C. (2003): The Effects of High Stakes Testing on Student Motivation and Learning. In: *Educational Leadership* 60, H. 5, S. 32-38.
- Assaf, L. (2006): One Reading Specialist's Response to High-Stakes Testing Pressures. In: *Reading Teacher* 60, H. 2, S. 158-167.
- Bishop, J.H. (1995): The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning. In: *International Journal of Educational Research* 23, H. 8, S. 653-752.
- Carnoy, M. (2005): Have State Accountability and High-Stakes Tests Influenced Student Progression Rates in High School? In: *Educational Measurement: Issues and Practice* 24, H. 4, S. 19-31.
- Cheng, L. (1999): Changing Assessment: Washback on Teacher Perceptions and Actions. In: *Teaching and Teacher Education* 15, H. 3, S. 253-271.
- Cheng, L. (2003): Looking at the Impact of a Public Examination Change on Secondary Classroom Teaching: A Hong Kong Case Study. In: *Journal of Classroom Interaction* 38, H. 1, S. 1-10.
- Cheng, L./Curtis, A. (2004): Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In: Cheng, L./Watanabe, Y./Curtis, A. (Hrsg.): *Washback in Language Testing. Research Contexts and Methods*. Mahwah/London: Lawrence Erlbaum, S. 3-17.
- Demie, F. (2003): Using Value-added Data for School Self-Evaluation: A Case Study of Practice in Inner-City Schools. In: *School Leadership & Management* 23, H. 4, S. 445-467.
- Diamond, J.B. (2007): Where the Rubber Meets the Road: Rethinking the Connection Between High-Stakes Testing Policy and Classroom Instruction. In: *Sociology of Education* 80, H. 4, S. 285-313.
- Faulkner, S.A./Cook, C.M. (2006): Testing vs. Teaching: The Perceived Impact of Assessment Demands on Middle Grades Instructional Practices. In: *Research in Middle Level Education* 29, H. 7, S. 1-13.
- Firestone, W.A./Winter, J./Fitz, J. (2000): Different Assessments, Common Practice? Mathematics Testing and Teaching in the USA and England and Wales. In: *Assessment in Education* 7, H. 1, S. 13-37.
- Grant, S.G. (2007): High-Stakes Testing: How Are Social Studies Teachers Responding? In: *Social Education* 71, H. 5, S. 250-254.
- Groß Ophoff, J./Hosenfeld, I./Koch, U. (2007): Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. In: *Empirische Pädagogik* 21, H. 4, S. 411-427.
- Hayes, S.G./Rutt, S. (1999): Primary Analysis for Secondary Schools: An LEA Research Officer's Perspective on Helping Secondary Schools Interpret Assessment Data for School Improvement Purposes. In: *Improving Schools* 2, H. 2, S. 44-52.
- Herman, J.L. (2004): The Effects of Testing on Instruction. In: Fuhrman, S.H./Elmore, R.F. (Hrsg.): *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press, S. 141-166.
- Ingram, D./Louis, K.S./Schroeder, R.G. (2004): Accountability Policies and Teacher Decision Making: Barriers to the Use of Data to Improve Practice. In: *Teachers College Record* 106, H. 6, S. 1258-1287.
- Ketter, J./Pool, J. (2001): Exploring the Impact of a High-Stakes Direct Writing Assessment in Two High School Classrooms. In: *Research in the Teaching of English* 35, H. 3, S. 344-393.

- Koch, U. (2011): Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnisrückmeldungen aus Vergleichsarbeiten. Münster u.a.: Waxmann.
- Kühle, B. (2010): Zentrale Lernstandserhebungen – Ergebnisorientierte Unterrichtsentwicklung? Schulische Strategien beim Umgang mit Ergebnissen aus den Schulrückmeldungen im Kontext der ersten Lernstandserhebungen 2004/2005 in Nordrhein-Westfalen. Berlin: Köster.
- Kühle, B./Peek, R. (2007): Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. In: Empirische Pädagogik 21, H. 4, S. 428-447.
- Kuper, H./Hartung, V. (2007): Überzeugungen zur Verwendung des Wissens aus Lernstandserhebungen. In: Zeitschrift für Erziehungswissenschaft 10, H. 2, S. 214-229.
- Lee, J./Wong, K.K. (2004): The Impact of Accountability on Racial and Socioeconomic Equity: Considering Both School Resources and Achievement Outcomes. In: American Educational Research Journal 41, H. 4, S. 797-832.
- Louis, K.S./Febey, K./Schroeder, R. (2005): State-Mandated Accountability in High Schools: Teachers' Interpretations of a New Era. In: Educational Evaluation and Policy Analysis 27, H. 2, S. 177-204.
- Luxia, Q. (2007): Is Testing an Efficient Agent for Pedagogical Change? Examining the Intended Washback of the Writing Task in a High-Stakes Writing Test in China. In: Assessment in Education 14, H. 1, S. 51-74.
- Maier, U. (2007): Welche Konsequenzen ziehen Mathematiklehrkräfte aus verpflichtenden Diagnose- und Vergleichsarbeiten? In: *mathematica didactica* 30, H. 2, S. 5-31.
- Maier, U. (2008a): Vergleichsarbeiten im Vergleich – Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. In: Zeitschrift für Erziehungswissenschaft 11, H. 3, S. 453-474.
- Maier, U. (2008b): Rezeption und Nutzung von Vergleichsarbeiten aus der Perspektive von Lehrkräften. In: Zeitschrift für Pädagogik 54, H. 1, S. 95-117.
- Maier, U. (2010): Effekte testbasierter Rechenschaftslegung auf Schule und Unterricht: Ist die internationale Befundlage auf Vergleichsarbeiten im deutschsprachigen Raum übertragbar? In: Zeitschrift für Pädagogik 56, H. 1, S. 112-128.
- McDonnell, L.M./Choisser, C. (1997): Testing and Teaching: Local Implementation of New State Assessments (CSE Tech. Rep. No. 442). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Mintrop, H./Trujillo, T. (2007): The Practical Relevance of Accountability Systems for School Improvement: A Descriptive Analysis of California Schools. In: Educational Evaluation and Policy Analysis 29, H. 4, S. 319-352.
- Nachtigall, C./Jantowski, A. (2007): Die Thüringer Kompetenztests unter besonderer Berücksichtigung der Evaluationsergebnisse zum Rezeptionsverhalten. In: Empirische Pädagogik 21, H. 4, S. 401-410.
- Olson, K. (2007): Lost Opportunities to Learn: The Effects of Education Policy on Primary Language Instruction for English Learners. In: Linguistics and Education 18, H. 2, S. 121-141.
- Parke, C.S./Lane, S./Stone, C.A. (2006): Impact of a State Performance Assessment Program in Reading and Writing. In: Educational Research and Evaluation 12, H. 3, S. 239-269.
- Popham, W.J. (1987): The Merits of Measurement-Driven Instruction. In: Phi Delta Kappan 68, H. 9, S. 679-682.
- Rosenshine, B. (2003): High-Stakes Testing: Another Analysis. In: Education Policy Analysis Archives 11, H. 24. URL: epaa.asu.edu/epaa/v11n24/; Zugriffsdatum: 01.02.2012.
- Rudd, P./Davies, D. (2002): A Revolution in the Use of Data? The LEA Role in Data Collection, Analysis and Use and Its Impact on Pupil Performance. Slough: NFER.

- Sloan, K. (2007): High-Stakes Accountability, Minority Youth, and Ethnography: Assessing the Multiple Effects. In: *Anthropology & Education Quarterly* 38, H. 1, S. 24-41.
- Stecher, B.M. (2002): Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practice. In: Hamilton, L.S./Brian, M./Stecher, S./Klein, P. (Hrsg.): *Making Sense of Test-Based Accountability in Education*. Santa Monica, CA: RAND Education, S. 79-100.
- Stecher, B.M./Barron, S. (2001): Unintended Consequences of Test-Based Accountability when Testing in "Milepost" Grades. In: *Educational Assessment* 7, H. 4, S. 259-281.
- Valli, L./Chambliss, M. (2007): Creating Classroom Cultures: One Teacher, Two Lessons, and a High-Stakes Test. In: *Anthropology & Education Quarterly* 38, H. 1, S. 57-75.
- Vogler, K.E. (2005): Impact of an Accountability Examination on Tennessee Social Studies Teachers' Instructional Practices. In: *Research in the Schools* 12, H. 2, S. 41-55.
- Watanabe, M. (2007): Displaced Teacher and State Priorities in a High-Stakes Accountability Context. In: *Educational Policy* 21, H. 2, S. 311-368.
- Wikeley, F./Stoll, L./Lodge, C. (2002): Effective School Improvement: English Case Studies. *Educational Research and Evaluation*. In: *School Effectiveness and School Improvement in a European Context* 8, H. 4, S. 363-385.
- Wößmann, L. (2007): International Evidence on School Competition, Autonomy and Accountability: A Review. In: *Peabody Journal of Education* 82, H. 2-3, S. 473-497.
- Yang, M./Goldstein, H./Rath, T./Hill, N. (1999): The Use of Assessment Data for School Improvement Purposes. In: *Oxford Review of Education* 25, H. 4, S. 469-483.

Uwe Maier, Prof. Dr., geb. 1971, Lehrstuhl für Schulpädagogik an der Universität Erlangen-Nürnberg.

Anschrift: Regensburgerstraße 160, 90478 Nürnberg
E-Mail: uwe.maier@ewf.uni-erlangen.de

Harm Kuper, Prof. Dr., geb. 1966, Arbeitsbereich Weiterbildung und Bildungsmanagement an der Freien Universität Berlin.

Anschrift: Arnimallee 12, 14195 Berlin
E-Mail: harm.kuper@fu-berlin.de

Knut Schwippert

Empirische Bildungsforschung: Perspektiven der Erziehungswissenschaft

Zusammenfassung:

Die erziehungswissenschaftliche Bildungsforschung hat nicht nur betreffend ihrer inhaltlichen, sondern auch betreffend ihrer methodologischen Ausrichtung eine bewegte Geschichte. In diesem Beitrag werden neben einem kursorischen Rückblick auf diese historische Entwicklung auch Überlegungen zu aktuellen Entwicklungen und zukünftigen Herausforderungen vorgestellt. Es wird aufgezeigt, dass die erziehungswissenschaftliche Bildungsforschung aufgrund ihrer Besinnung auf ihre Wurzeln und aufgrund ihrer Offenheit für interdisziplinäre Zusammenarbeit für diese Anforderungen gut gerüstet ist. Schlüsselwörter: Mixed-Methods, erziehungswissenschaftliche Bildungsforschung, theoretisch begründete Bildungsforschung

Empirical Educational Research: Perspectives of Educational Sciences

Abstract:

The empirical educational research has an eventful history, not only because of its content but also because of its methodological orientation. The article offers in addition to a brief historical review also some thoughts about the actual situation of empirical educational research. Reflections on future developments of and challenges for empirical educational research as an actual field of research are presented. The author shows how empirical educational research, due to reflections on its historical roots and its openness for interdisciplinary cooperation, is well prepared for future challenges.

Key words: mixed methods, empirical educational research, theoretically founded educational research

In diesem Beitrag wird es nicht um die (erneute) detaillierte und vollständige historische Aufarbeitung der Gründung, Entwicklung und Ausdifferenzierung der erziehungswissenschaftlichen Bildungsforschung gehen. Es wird auch nicht um eine übergreifende Betrachtung „der“ Erziehungswissenschaft gehen, sondern vielmehr um die empirische Bildungsforschung innerhalb der Erziehungswissenschaft und der an die Erziehungswissenschaft angrenzenden Disziplinen. Die Historie wird nur insoweit vorgestellt, als es für das Verständnis aktueller Bewegungen und Probleme sowie zukünftiger Herausforderungen notwendig erscheint. Der Fokus wird damit auf aktuellen Möglichkeiten und Grenzen und auch auf notwendigen Veränderungen in der Erziehungswissenschaft liegen.¹

1. Zwischen Historie und Moderne: Wo findet sich die Erziehungswissenschaft in der empirischen Bildungsforschung wieder?

Zunächst sei der Blick zurückgewendet, und es sei der Frage nachgegangen, ob die Erziehungswissenschaft eine Disziplin ist, die jetzt erst den Wert (quantitativ) empirischer Bildungsforschung für sich erkannt hat und diese erst zögerlich als Ergänzung der geisteswissenschaftlichen Ausrichtung versteht. Nach König und Zedler (1998) gab es „erste Versuche, Pädagogik nach dem Vorbild der Naturwissenschaften zu entwerfen, [...] [bereits] in der zweiten Hälfte des 18. Jahrhunderts“ (S. 36). Nach ihren Recherchen setzt sich bereits Trapp (1745-1818) mit der Anlehnung an die Naturwissenschaften auseinander, die von Beobachtungen und Erfahrungen ausgehen. Damit sind frühe Entwicklungen empirischer Sozial- und Humanwissenschaften markiert, die sich schließlich seit dem 19. Jahrhundert von ihrer anfänglichen Anlehnung an naturwissenschaftliche Konzepte der Philosophie zunehmend emanzipiert haben. Als weiteres Beispiel sei auf den rund 100 Jahre später (1912 bereits in der zweiten Auflage) erschienenen Band „Experimentelle Pädagogik“ von Lay (1912) verwiesen, in dem er verschiedene empirische Methoden beschrieben hat. Lay spricht sich hier für ein breites empirisches Repertoire aus: die Beobachtung, die Statistik und das Experiment. Sie waren für ihn Mittel zum Zweck, um Forschungsfragen unterschiedlicher Reichweiten und Paradigmen nachgehen zu können. Auch Fend (2009) stellt in seiner historischen Rückschau Pädagogen vor, die sich im letzten Jahrhundert immer wieder mit empirischen Fragestellungen auseinandergesetzt haben (hier seien exemplarisch Flitner, Nohl, Litt und Spranger, die z.T. auch interdisziplinär verankert waren, genannt), und zeigt damit, dass sich nicht erst mit Heinrich Roths „realistischer Wende“ in den 1960er-Jahren empirische Forschung in der Erziehungswissenschaft etablierte, auch wenn der Umbruch gerade in dieser Zeit für das Verständnis der aktuellen Entwicklung eine wichtige Rolle spielt.

1 Ich bedanke mich herzlich bei Eckhard Klieme (DIPF, Frankfurt a.M.), der mir in unseren Diskussionen zahlreiche wichtige und vor allem konstruktive Hinweise zu dem Beitrag gegeben hat.

Im Verlauf des letzten Jahrhunderts traten unterschiedliche Forschungsparadigmen jeweils phasenweise in den Vordergrund, während andere, nicht „favorisierte“ Paradigmen in den Hintergrund traten – was jedoch nie verhindert hat, dass auch weniger „populäre“ Paradigmen weiter angewandt und weiterentwickelt wurden. Vor der aktuellen Phase, in der eine quantitativ-empirische Schwerpunktsetzung bei der erziehungswissenschaftlichen Bildungsforschung zu verzeichnen ist, gab es eine Phase, in der – wenn empirisch geforscht wurde – öfter auf einen qualitativen empirischen Zugang zurückgegriffen wurde. Da in der Erziehungswissenschaft jeweils beide Ansätze kontinuierlich – wenn auch in unterschiedlichem Umfang – genutzt und weiterentwickelt wurden, hat die Erziehungswissenschaft für die Methodenintegration qualitativer und quantitativer empirischer Zugänge einige Möglichkeiten zu bieten. Nach Fend (2009) erfordert auch und gerade eine verstehensorientierte Konzeption des Bildungswesens eine multimodale Erschließung – dies gilt damit auch für die Identifizierung, Erfassung und Interpretation von Merkmalen des Bildungswesens. „So gilt das pädagogische Feld als hoch komplex, so dass beispielsweise nicht alle intervenierenden Variablen vorhersagbar sind“ (Tippelt 2009); gerade deshalb sind auch explorative empirische Methoden gefordert. Angeregt durch die geisteswissenschaftlich orientierte Erziehungswissenschaft wurden insbesondere in der Nachkriegszeit hermeneutisch verstehende Ansätze verfolgt. Unter knapper Erläuterung des Zeitgeistes der späten 1960er-Jahre verweist Fend (2009) pointiert auf die scharfe Kritik an den quantitativen empirischen Methoden, die jedoch nicht methodologisch, sondern eher politisch motiviert war, und verweist zudem auch auf den bis heute im Gedächtnis gebliebenen „oppositionellen Impetus“ gegen quantitative Methoden. Aber auch gegenüber qualitativen empirischen Ansätzen wurden und werden bis heute aus quantitativ orientierter Perspektive kontinuierlich Vorbehalte formuliert. Zentral ist hierbei insbesondere die Kritik an der Generalisierung von Befunden und an der bis heute strittig und kontrovers geführten Diskussion um eine generelle Akzeptanz, Bestimmung und Nutzung von Gütekriterien. Bei einer Rekonstruktion dieser Verhältnisse lässt sich aus heutiger Sicht feststellen, dass Kritik, insbesondere aus Unwissenheit über die jeweiligen Stärken der „anderen Ansätze“, häufig auch despektierlich vorgebracht wurde. Hierbei sind Begriffe wie „Impressionisten“ und „Fliegenbeinzähler“ noch harmloser Natur. Gerade vor dem Hintergrund der wechselseitig vorgebrachten Kritik ist dem Verlauf der Diskurse zu entnehmen, dass in der erziehungswissenschaftlichen Bildungsforschung sowohl quantitative empirisch-erklärende Ansätze als auch hermeneutisch verstehende und die sich daraus entwickelnden qualitativ empirischen Ansätze über das gesamte letzte Jahrhundert hinweg durchgängig betrieben und weiterentwickelt wurden – die empirische Pädagogik hat in dieser Zeit wichtige methodologische Erweiterungen sowohl in den qualitativen als auch den quantitativen Ansätzen erfahren.

Aktuell tritt in der erziehungswissenschaftlichen Bildungsforschung der Ansatz quantitativer empirischer Methoden in den Vordergrund. Grund hierfür ist, dass die Bildungsforschung sich als zunehmend interdisziplinäres Forschungsfeld entwickelt

und sich durch den methodologischen Austausch der Erziehungswissenschaft mit Nachbardisziplinen wie der Psychologie, der Soziologie und auch der Ökonomie, bei denen diese Verfahren verbreitet und etabliert sind, gemeinsame methodische Standards herausbilden. Wie zuvor beschrieben, ist die Nutzung quantitativer empirischer Methoden auch der forschenden Erziehungswissenschaft nicht neu. Einem Teil der erziehungswissenschaftlichen Bildungsforschung war dieser Zugang stets immanent (vgl. Tippelt/Schmidt 2010).

Mit dem Blick auf die beiden letzten Dekaden schließlich reflektieren Zedler und Döbert (2010) die theoretische und methodologische Reichweite der unter den Begriffen Bildungsforschung und erziehungswissenschaftliche Forschung subsumierten Verständnisse. Je nach Auslegung der beiden Begriffe rücken sie sowohl inhaltlich als auch methodisch einmal näher zusammen und einmal weiter auseinander. Dies zeigen die beiden Autoren durch einen Abriss unterschiedlicher Forschungsschwerpunkte, die mit unterschiedlichen Wertgewichtungen aufgrund jeweils aktueller gesellschaftlicher, politischer und auch wissenschaftlicher Konjunkturen alternierten. Für die aktuelle Bildungsforschung arbeiten sie drei Schwerpunkte heraus, bei denen das Bemühen, empirisch fundierte Steuerungsentscheidungen im Bildungssystem zu treffen, aktuell – gerade auch durch Bildungspolitik und Bildungsadministration – in den Vordergrund gerückt wird; diese prägen auch die derzeitige empirische Forschungslandschaft: (1) Lehr-Lern-Forschung, (2) Large Scale Assessments und (3) Bildungsmonitoring in Deutschland. Aus dieser Perspektive sehen die Autoren zukünftig insbesondere im Bereich der Steuerung von Bildungssystemen einen Bedarf an systematischen Daten und Indikatoren. Das Indikatorentableau wird im interdisziplinären Feld der empirischen Bildungsforschung weiter auszubauen sein und somit auch Gegenstand erziehungswissenschaftlicher Bildungsforschung bleiben. Auch wenn dieser Nutzungsgedanke mit seinem Schwerpunkt auf Systemmerkmalen aus politischer und administrativer Perspektive nachvollziehbar erscheint, so stehen im Forschungsfeld der erziehungswissenschaftlichen Bildungsforschung doch die Bildungs- und Erziehungsprozesse – und damit eine empirisch begründete, auf professionelle Praxis bezogene Forschung – im Vordergrund. Daher wären außer den drei genannten zentralen Forschungsbereichen unter anderem auch Forschung über Bildungsverläufe und professionelles Handeln in Institutionen (z.B. Schulqualität und Schulentwicklung) zu nennen. Hier geht es häufig um Grundlagenforschung über Zusammenhänge, zunehmend auch über Wirkungen im Bildungssystem bzw. im Bildungsverlauf. Damit entzieht sich erziehungswissenschaftliche Bildungsforschung ggfs. den Ansprüchen der „Steuerleute“ des Bildungssystems. Es spricht nichts gegen die Nutzung der Forschungsbefunde durch Bildungspolitik und Bildungsadministration – empirische Bildungsforschung darf sich jedoch nicht darauf einengen (lassen).

Betrachtet man den bis hier skizzierten zeitlichen Wandel der Konjunkturen empirischer Bildungsforschung, kann man mit Gräsel (2011, S. 23) feststellen, dass

es sich hierbei trotz dieses Wandels „nicht um eine Modeerscheinung handelt: International hat sich eine ‚evidence-based education‘ durchgesetzt, also eine Gestaltung von Bildungssystemen und -institutionen, die sich an Ergebnissen empirischer Forschung orientiert“. Gräsel sieht für die empirische Bildungsforschung zusammenfassend positive Entwicklungsmöglichkeiten, wenn sich die Akteure der empirischen Bildungsforschung auch der eigenen Forschungs- und Handlungslogiken und -grenzen bewusst sind (vgl. hierzu auch Tillmann u.a. 2008). Tippelt und Schmidt (2010, S. 12) stellen den Stand der empirischen Bildungsforschung im Vorwort des Handbuchs für Bildungsforschung folgendermaßen dar: „Man kann sagen, dass empirische Bildungsforschung als Orientierungsforschung immer dann Konjunktur hat, wenn Symptome sozialer Probleme gesellschaftlich wahrgenommen werden [...]. Eine entsprechende Bildungsforschung entsteht also aus einem besonderen Informationsbedürfnis einer interessierten Öffentlichkeit. Bildungsforschung kann in diesem Sinne auch Aufklärung leisten, weil stereotypen Urteilen und Vorurteilen sachliche Information entgegengesetzt werden kann.“ Und da mit einem steten Wandel der Gesellschaft mit den damit einhergehenden Problemen und Herausforderungen zu rechnen ist, wird die empirische Bildungsforschung sicherlich auch weiterhin Bestand haben und sich – wie in der Vergangenheit – auch zukünftig kontinuierlich weiterentwickeln. Aus diesem Selbstverständnis heraus kann eine empirische Bildungsforschung nur interdisziplinär gelingen.

2. Bildungsforschung als interdisziplinäres Forschungsfeld

Aus dem kurzen historischen Abriss wird deutlich, dass es für die empirische Bildungsforschung konstitutiv ist, dass sie sich sowohl quantitativer als auch qualitativer empirischer Methoden im Rahmen vereinbarter Qualitätsstandards bedient. Darüber hinaus wird die Bildungsforschung als interdisziplinäres Forschungsfeld verstanden, welches sich schon in der langjährigen Zusammenarbeit und aktuell – also 2012 – in der geplanten Gründung einer gemeinsamen Fachgesellschaft ausdrückt, in der insbesondere Vertreter und Vertreterinnen der Erziehungswissenschaft, der Fachdidaktiken, der Ökonomie, Psychologie und Soziologie (in alphabetischer Reihenfolge) ein gemeinsames Forum für empirische Bildungsforschung sehen. Jede der involvierten Disziplinen wird mit den jeweiligen Stärken ihren Teil dazu beitragen, Theorien und Methoden im Forschungsfeld weiter zu entwickeln, wobei der interdisziplinäre Austausch die Möglichkeit bietet, disziplinspezifische Paradigmen zu hinterfragen und produktiv weiterzuentwickeln, aber dadurch sicherlich auch die „Gefahr“ birgt, dass einzelne Paradigmen dieser Überprüfung nicht standhalten.

Die Perspektive auf eine disziplinübergreifende Bildungsforschung in einer sich wandelnden Gesellschaft aufgreifend, sei hier auf die Überlegungen von Terhart (2009, S. 36) verwiesen, der der Frage nachgeht, „was überhaupt Fortschritt in der Wissenschaft ist bzw. sein kann“ und welche konstituierenden bzw. immanenten

Merkmale die Gesellschaft bzw. deren Wandel beschreiben. Diese Merkmale können einerseits definitorisch festgeschrieben, andererseits aber auch im kontinuierlichen Wandel der Wertesysteme fortlaufend hinterfragt werden. Hierbei scheint die durchaus kontroverse Diskussion um den „Kompetenzbegriff“ in der Erziehungswissenschaft ein gutes Beispiel zu sein, dessen Relation einerseits zum Bildungsbegriff, andererseits zur psychometrischen Erfassung von Kompetenz im Rahmen von Kompetenztests (mit den entsprechenden Implikationen für das Individuum, den Unterricht und das System) und dem angelsächsischen *Literacy*-Ansatz diskutiert wird (vgl. etwa Klieme/Hartig 2007). Hier scheint weder mit Blick auf Methoden noch auf Fachdisziplinen eine normativ fundamentale Diskussion zielführend zu sein – vielmehr sollte das Konstrukt jeweils kontextabhängig und damit in Bezug auf die zu beantwortenden Forschungsfragen im Lichte unterschiedlicher Disziplinen, Forschungsparadigmen und Methoden betrachtet werden. Dies erscheint als ein forschungspragmatischer Weg, der die jeweilige Präzision in der Argumentation weder in den Traditionen spezifischer Fächer noch in der geisteswissenschaftlich bzw. empirisch orientierten Bildungsforschung beschneidet – und sich somit letztlich gegen eine Interpretationshoheit wendet.

Eine Balance zwischen Modernisierung und Bewahrung fachspezifischer Forschungsparadigmen unter dem Dach der Bildungsforschung zu finden, erscheint als zentrale Herausforderung der empirischen Bildungsforschung für die nächsten Jahre. Die Grenzen der Modernisierung wären in einer Konvergenz aller Fächer erreicht, bei der dann Bezüge zu älteren Forschungen und alternativen bzw. komplementären Paradigmen der jeweiligen Disziplinen verloren gehen würden. Die Grenzen der Bewahrung fänden sich in fachspezifischen dogmatischen Haltungen, die sich zum Beispiel im unbedingten Festhalten an vermeintlich „heimischen“ fachspezifischen Paradigmen ausdrücken könnten, die dann jedoch ggfs. auch zur Isolation der jeweiligen Fächer führen könnten. Damit die Zusammenarbeit von Forschern und Forscherinnen aus unterschiedlichen Forschungsdisziplinen, d.h. auch von Forschenden, die unterschiedlichen Methoden und Theorien nahestehen, funktionieren kann, muss ein disziplinäres Selbstbewusstsein mit Toleranz und Offenheit gepaart sein. Sollte sich hingegen eine der Disziplinen dominant positionieren (wollen), werden erwünschte und gewollte Synergien im Rahmen der Zusammenarbeit nicht nur minimiert – vielmehr würde es unter diesen Umständen auch zu einer Einengung und damit Verarmung möglicher Forschungsperspektiven und -methoden in der empirischen Bildungsforschung kommen. Gräsel (2011) verweist in ihrem Beitrag „Was ist empirische Bildungsforschung?“ explizit darauf, dass neben dem problemorientierten und interdisziplinären Anspruch der Bildungsforschung auch eine Offenheit sowohl für qualitative als auch für quantitative Methoden bestehen muss. Die Entwicklung von (insbesondere methodischen) Standards stellt eine Gemeinschaftsaufgabe dar, die nicht an Einzelne delegiert werden kann. Wobei – um es deutlich zu sagen – die Setzung von Standards hier nicht im Sinne des Anpassens unterschiedlicher Ansprüche bzw. Niveaus verstanden werden darf (hier sind immer

die höchsten Standards – auch guter wissenschaftlicher Arbeit – anzustreben), sondern es muss vielmehr eine Verständigung über die Konstitution der Standards geben.

Ernsthafte Probleme für das Aushandeln gemeinsamer Standards stellen Unkenntnis und Ignoranz dar. Probleme können sich somit einerseits aus Unkenntnis über die jeweils anderen methodischen Forschungsansätze heraus ergeben, aber andererseits auch – und dies gilt ausdrücklich sowohl für quantitative als auch für hermeneutisch-qualitative Methoden – durch Ignorierung angebrachter Methodenkritik, die sich dann in unzureichend durchgeführter Forschung ausdrückt.

Für die Etablierung einer interdisziplinären Bildungsforschung wäre es somit fatal, wenn aus dem Blickfeld der eigenen tradierten Forschungsdisziplin jeweils Forschungen der anderen Disziplinen mit dem Ziel herausgegriffen würden, die hierin gewählten methodischen und methodologischen Ansätze pauschal zu diskreditieren. Eine solche Haltung, wie sie leider in verschiedenen „Lagern“ immer wieder zu beobachten ist, muss selbstverständlich als unangemessen gelten, geht es doch vielmehr darum, die jeweiligen Stärken der unterschiedlichen Ansätze wertzuschätzen und zu verbinden und nicht – aus einem eingeschränkten Blickfeld heraus – nur die jeweils eigenen Ansätze als angemessen zu postulieren und isoliert die Schwächen der anderen herauszuarbeiten. Schwächen sollten sicherlich benannt, dann aber in einem konstruktiven Prozess der Verbesserung aufgehoben werden. Gerade in einer Phase der Unsicherheit und in Auseinandersetzung mit „schlecht“ durchgeführten Untersuchungen hat sich die Interaktion und damit auch disziplinübergreifende Kooperation bewährt, da sich hier bereits theoretische und auch methodische Qualitätsstandards etablieren konnten, die zurückgespiegelt in allen beteiligten Disziplinen zur Weiterentwicklung angewandter Methoden geführt haben. Ein Beispiel hierfür ist die Kritik neurowissenschaftlicher Ansätze aus erziehungs- und sozialwissenschaftlicher Perspektive (vgl. Stern/Grabner/Schumacher 2007). Diese setzt Offenheit und Selbstkritikfähigkeit nicht nur seitens der Individuen, sondern ebenso seitens der Fachgesellschaften als Ganzer voraus. Von der erziehungswissenschaftlichen empirischen Bildungsforschung ist daher zu erwarten, dass methodenkritische Betrachtungen sowohl für quantitative als auch für hermeneutisch-qualitative empirische Ansätze anerkannt werden. Dieser Prozess der wechselseitigen methodischen und meta-theoretischen Kritik verlangt nicht zuletzt eine nachvollziehbare und vollständige Dokumentation des Forschungsprozesses, der angewandten Methoden und der Forschungsbefunde, denn nur so lässt sich ein Verständnis dafür entwickeln, was als nachvollziehbare Interpretation und akzeptable Forschung gelten kann.

Welchen Platz hat die erziehungswissenschaftliche Bildungsforschung in den Annalen der Bildungsforschung? Auch wenn – wie gezeigt werden konnte – ein Teil der Erziehungswissenschaft sich trotz der langdauernden Dominanz geisteswissenschaftlicher Forschungstraditionen seit jeher auch dem quantitativen Forschungspara-

digma verpflichtet gefühlt hat (vgl. Tippelt/Schmidt 2010), so ist doch über einige Zeit die Aktivität der Kolleginnen und Kollegen in diesem Feld unter die Wahrnehmungsschwelle gerutscht. Hierfür mit ein Grund kann auch sein – und dies zeigt sich bis heute in den Publikationsstrategien von Journalen und Sammelbänden –, dass es einen missverstandenen „*state of the art*“ gibt, der zu einseitig fokussierten Forschungsarbeiten, Methoden oder Befunden führen kann. Als Beispiel sei hier das aus Forschungen zu Meta-Analysen bekannte *file-drawer*-Problem (vgl. Drinkmann 1990) genannt. Gleiches gilt für den schon „gebetsmühlenartig“ vorgebrachten Anspruch, bestimmte Analysen seien NUR mit dem einen oder dem anderen Ansatz durchführbar. Hier wird – insbesondere fachspezifisch – pauschalisierend eine *conditio sine qua non* postuliert, die das mögliche Ausloten neuer oder seltener angewandter Methoden unnützerweise erschwert. Statt die (Nicht-)Anwendung spezifischer Methoden als „Totschlagargument“ zu nutzen, wäre hier vielmehr eine konstruktive Einschätzung der tatsächlichen Reichweite von Befunden angebracht. Zur Vergegenwärtigung: Standards allgemein und damit auch methodische Standards verändern sich über die Zeit und werden sich – so das wissenschaftstheoretische Ideal von Versuch und Irrtum anerkannt wird – kontinuierlich verbessern. Diese sich verändernden Standards lassen aber ältere oder unter anderen Umständen realisierte Forschungen (insbesondere auch anderer Fachdisziplinen) nicht als unnützlich, falsch oder hinfällig erscheinen, wenn sie, zu ihrer Zeit bzw. in ihrem Rahmen *state of the art*, den Forschungsstand mit befruchtet und vielleicht ja erst die Eröffnung neuerer und besserer Standards ermöglicht haben bzw. mit dem Blick auf entsprechende zukünftige Forschungen diese ermöglichen werden.

Zu fordern sind also eine Akzeptanz auch der nicht im *mainstream* liegenden Forschung und die Bereitschaft, die hierbei gewonnenen Befunde und Erkenntnisse in die empirische Bildungsforschung einfließen zu lassen. Terhart (2009, S. 38) macht dies an dem Beispiel aktueller Bildungsforschung deutlich: „Sehr stark ist derzeit die in der empirischen Bildungs-/Schulforschung dominierende produktionsistische Denkweise, die der Frage nach dem Verhältnis von Aufwand und Ertrag folgt und alles in Modelle bringt, mit deren Hilfe Bedingungen, Prozesse und Resultate unterschieden und analysiert werden, und deren Verhältnis zueinander optimiert werden soll“. Weiter heißt es bei Terhart: „Im Bildungskontext stellt das Modell die Basis für das Bemühen um evidenzbasierte Bildungspolitik dar [...]. Dieses produktionsistische Variablenmodell liefert sicherlich eine gute Grundlage für empirische Bildungsforschung, bleibt aber ein Modell – und ein Modell ist noch keine Theorie“. Terhart (2011, S. 248) bemerkt mit Blick auf bildungstheoretische, schulpädagogische und schulorganisatorische Forschung an anderer Stelle: „Sie deshalb [weil sie sich nicht psychometrischer Forschungsmethodiken bedient (Anm. KS)] als weniger wichtig oder gar inexistent zu betrachten hieße, durch methodische Imperative die Blickperspektive ohne Not und fahrlässig allzu eng zu führen“. Dieser Kritik mangelnder Theoriebildung kann mit dem Blick auf das alte Produkt-Prozess-Paradigma weitgehend zugestimmt werden. Jedoch greift sie zu kurz, wenn sie sich auf empi-

rische Bildungsforschung im weiteren Sinne bezieht. Gerade die aktuelle Schul- und Unterrichtsforschung ist sehr wohl theoretisch fundiert. Klieme und Rakoczy (2008) legen beispielsweise Wert auf den Unterschied zwischen einer bloßen empirischen Beschreibung und Klassifizierung einerseits und einem empirisch fundierten Verstehen von Zusammenhängen und Wirkungen andererseits. Als Prototyp des beschriebenen Ansatzes verweisen sie auf das Rahmenmodell, wie es zum Beispiel Scheerens und Bosker (1997) beschrieben haben, mit seiner Systematisierung von relevanten Merkmalen mit unterschiedlichem Status (Kontext/Input-, Prozess- bzw. Ergebnis/Output-Variablen) auf unterschiedlichen Ebenen (Schule, Klasse und Lehrperson, Schüler/Schülerin und Herkunftsfamilie). Für die Schul- und Unterrichtsforschung sei es aber von zentraler Bedeutung zu verstehen, in welchem Zusammenhang Prozessqualitäten der übergeordneten Ebenen (insbesondere Schul- und Unterrichtsqualität) mit den Lehr-Lern-Prozessen der Schülerinnen und Schüler stehen. Klieme und Rakoczy (2008) weisen darauf hin, dass hierzu ein Rückgriff auf spezifische, etwa schulpädagogisch, psychologisch und nicht zuletzt fachdidaktisch verankerte Theorien notwendig ist, und verweisen damit deutlich auf den Anspruch einer theoretisch fundierten empirischen Forschung, die selber auch zur Weiterentwicklung von Theorien beiträgt.

3. Innovationen und Perspektiven der erziehungswissenschaftlichen Bildungsforschung am Beispiel der Methodendiskussion

Einen neuen und wichtigen Impuls für die erziehungswissenschaftliche Bildungsforschung, wie sie sich aktuell im Bereich von Leistungsvergleichsuntersuchungen konstituiert, hat es ab Anfang der 1990er-Jahre durch Rainer Lehmann gegeben. Ihm ist es zu verdanken, dass Deutschland die gut zwanzigjährige Abstinenz von internationalen Schulvergleichsuntersuchungen durch die Teilnahme an der Internationalen *Reading Literacy* Studie (IRL) überwunden hat (vgl. Schwippert/Goy 2008). Aber erst die nachfolgend durch Rainer Lehmann und Jürgen Baumert initiierten, durch Bund und Länder unterstützten Untersuchungen im Rahmen der *Third International Mathematics and Science Study* (TIMSS/II und TIMSS/III) wurden von einer breiteren Öffentlichkeit und einer durch die Ergebnisse dieser Studien aufgeschreckten Bildungspolitik wahrgenommen. Die Befunde der TIMS-Untersuchung stellten den Ausgangspunkt für die Beteiligung Deutschlands an PISA (*Programme for International Student Assessment*) dar. Durch PISA schließlich hat die empirische Bildungsforschung nunmehr ein sehr hohes Maß an Aufmerksamkeit erreicht, so dass sich die quantitative empirische Bildungsforschung aktuell als Protagonist in der erziehungswissenschaftlichen empirischen Forschung darstellt. Mit dem Mut, auch die traditionellen Grenzen zu überschreiten, haben es Rainer Lehmann und Jürgen Baumert geschafft, nicht nur IRL, TIMSS und PISA – also Deutschlands Teilnahme an internationalen Schulvergleichsstudien – zu initiieren, sondern auch ein Interesse

an quantitativ empirisch fundierten Befunden zu wecken, welches die Forschung und als Folge auch die Theorieentwicklung in der Schul- und Unterrichtsforschung unter anderem in den Bereichen der Schulpädagogik, Fachdidaktik und Psychologie angeht und befruchtet hat. Mit diesen Studien wurde die in der Erziehungswissenschaft immer noch vorherrschende geisteswissenschaftliche Forschung um quantitative empirische Perspektiven bereichert. Was die sich seitdem zunehmend etablierende (erziehungswissenschaftliche) Bildungsforschung erfolgreich gemacht hat, ist sicherlich auch die Tatsache, dass sie den Mut aufbrachte, nicht mehr nur erziehungswissenschaftliche Theorien und Methoden aufzugreifen, sondern – sich am eigentlichen Forschungsinteresse ausrichtend – sich auch disziplinübergreifend zu orientieren. Somit wurden in der sich mit der Zeit zunehmend etablierenden erziehungswissenschaftlichen Bildungsforschung auch Arbeiten der Soziologie (sozialer Hintergrund; Bourdieu), Psychologie (Interessen und Motivation; Schiefele/Krapp) und Ökonomie (Produktionsmodell; Walberg) berücksichtigt und damit auch ein wichtiger Beitrag zur Etablierung einer interdisziplinären Bildungsforschung geleistet (vgl. auch Deutscher Bildungsrat 1974 – nach Gräsel 2011). Fend (2009, S. 25) bringt es so auf den Punkt: „Es ist unübersehbar, dass in den letzten fünfzig Jahren Bildungsforschung noch nie so einflussreich war und die politische Agenda prägen konnte wie im Rahmen der PISA-Studien. Sie repräsentiert eine ungewöhnliche Renaissance der Erziehungswissenschaft empirischer Provenienz, die nicht zuletzt von einer hohen Professionalität in der Forschung und ihrer Präsentation geprägt war.“

Dieser Mut, auch über eigene Forschungstraditionen und Paradigmen hinweg Neues in die eigene Arbeit zu integrieren, scheint auch die Stärke der Erziehungswissenschaft zu sein. Neben der zuvor erwähnten Integration von Theorien z.B. der Soziologie, Psychologie oder der Ökonomie gilt dies auch für die Integration und Anwendung verschiedener empirischer Methoden. Aktuelle Integrationsbemühungen drücken sich hierbei in der Ausbildung aufwachsender methodischer und methodologischer Qualität nicht nur in quantitativen, sondern auch in den qualitativen Methoden aus.²

2 Die Bemühung, empirische Bildungsforschung mit ihrer Methodik und auch ihren Befunden in der Erziehungswissenschaft zu verankern und auch zu verstetigen, wird in dem 2008 von der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE) vorgestellten „Kerncurriculum Erziehungswissenschaft“ deutlich (vgl. Deutsche Gesellschaft für Erziehungswissenschaft [DGfE] 2008). Hier werden sowohl für grundständige erziehungswissenschaftliche Studiengänge als auch für die Lehramtsausbildung systematisch empirische Methoden eingeführt, damit sich die Studierenden im weiteren Studienverlauf mit empirischen Studien sowohl methodologisch als auch inhaltlich kritisch auseinandersetzen können und in Abschlussarbeiten auch die Möglichkeit bekommen, die Anwendung eines Forschungsverfahrens zu vertiefen.

Dieses Studienmodell ist in Hamburg für die 2007 eingeführten BA-/MA-Studiengänge zumindest in Teilen aufgegriffen worden. In der grundständigen erziehungswissenschaftlichen Ausbildung zeigt sich dies an dem Angebot, den Studierenden Grundbegriffe, Theorien und Methoden insbesondere auch der empirischen Bildungsforschung in entsprechenden Seminaren bzw. Modulen nahezubringen. Ein besonderes Augenmerk wurde hierbei auf die Vermittlung sowohl qualitativer als auch quantitativer Methoden und auf die in der Synergie beider Ansätze liegenden Möglichkeiten von Mixed-Methods gelegt. Dieses Angebot zieht

Diese Offenheit gegenüber unterschiedlichen Methoden verschiedener Paradigmen wird heute in der Erziehungswissenschaft im Rahmen von Diskursen zu Mixed-Methods deutlich, bei denen die jeweiligen Stärken der unterschiedlichen Ansätze integrativ herausgearbeitet werden. So zeichnet sich schon heute eine zunehmende Akzeptanz – insbesondere in der erziehungswissenschaftlichen Bildungsforschung – des Verständnisses ab, wonach nicht Methoden die Forschungsfragen determinieren sollten, sondern umgekehrt je nach Forschungsfrage angemessene Methoden genutzt werden sollten. Der Diskussion um Mixed-Methods in der erziehungswissenschaftlichen Bildungsforschung ist – dies sei hier nochmals betont – der unbedingte Anspruch immanent, für die jeweils eingesetzten (empirischen) Methoden die jeweils höchsten Qualitätsstandards zu erfüllen – seien sie nun qualitativ und/oder quantitativ. „Der Kern ihrer Bemühungen liegt in der datengestützten Fehlerreduktion von Aussagen. Das ist in konkreter Form mit ‚Wahrheitsfindung‘ gemeint. Dazu entwickelt sie methodische Verfahren, seien es Experimente, gezielte Beobachtungen, Textanalysen oder große Surveys, und setzt sich so den Daten als ‚Gegner‘ der eigenen Annahmen aus“ (Fend 2009, S. 27). Hierbei erscheint ein empirisch begründetes theoretisches Verstehen und Erklären als fundamentales Ziel von Wissenschaft. Aufgrund der langjährigen Erfahrungen im Überschneidungsbereich explorativer und überprüfender empirischer Forschung liegt die Stärke erziehungswissenschaftlicher Bildungsforschung in der Anwendung von Mixed-Methods als *state of the art*.

Die oben beschriebene interdisziplinär orientierte und methodologische Offenheit der erziehungswissenschaftlichen Bildungsforschung – die sich möglicherweise aus einer streitbaren Diskussionskultur innerhalb der Erziehungswissenschaft ergeben haben könnte – hat das Fach weiterentwickelt, anschlussfähig gemacht und es auch für die Zukunft gerüstet. Was wichtig erscheint, ist, dass erziehungswissenschaftliche Forschungstraditionen bewahrt werden, jedoch nicht ohne diese unter den modernen und interdisziplinär akzeptierten methodischen Qualitätsanforderungen zu prüfen

sich für die Studierenden, die in Hamburg ihr erziehungswissenschaftliches Masterstudium absolvieren, wie ein roter Faden durch die Module. Einen besonderen Schwerpunkt setzt die Universität Hamburg mit dem Master-Profilbereich „Bildungstheorie und Empirische Bildungsforschung“. In diesem Profilbereich wird ein besonderer Fokus auf die Verschränkung beider sich über Generationen hinweg eher unversöhnlich gegenüberstehenden Forschungstraditionen gelegt. Leider ist es in den lehramtsbezogenen Studiengängen bisher nicht gelungen, das von der DGfE empfohlene Kerncurriculum in dieser Form umzusetzen – in Hamburg ist die Notwendigkeit einer gut fundierten methodischen Ausbildung mit den sich daraus ergebenden Möglichkeiten für die Lehrerinnenausbildung (noch) nicht aufgegriffen worden. Das Bewusstsein, dass eine fundierte fachliche Auseinandersetzung mit der für das Fach relevanten Literatur nur unter vorangehender Vermittlung methodischen und methodologischen Basiswissens gelingen kann, scheint aber – und hierbei spielen die empirisch forschenden Fachdidaktiken in Hamburg eine wichtige Rolle – im Rahmen der Lehrerinnenausbildung zunehmend als notwendig erkannt zu werden. Der hierfür erforderliche Austausch zwischen den Erziehungswissenschaften und den Fachdidaktiken ist in Hamburg durch das sogenannte „Hamburger Modell“ erleichtert, welches sich an der Zuordnung der Fachdidaktiken zur Fakultät der Erziehungswissenschaft und nicht zu den Fakultäten der Fächer ablesen lässt.

und weiterzuentwickeln, um so zunehmend methodisch unzureichend abgesicherte, lückenhaft dokumentierte und überinterpretierte subjektive Erfahrung durch theoretisch und methodisch angemessen fundierte Forschungsbefunde zu ersetzen. An angemessenen wissenschaftlichen Standards, die auch disziplinübergreifend anzuerkennen sind, führt kein Weg vorbei.

Für die Erziehungswissenschaft im Allgemeinen wäre wünschenswert, dass insbesondere die leider immer noch geführten ideologischen Glaubenskriege zwischen Empirikern und Hermeneutikern, den „Fliegenbeinzählern“ und den „Impressionisten“, beendet werden. Ein Blick auf die sich derzeit etablierende und ausbildende Generation von erziehungswissenschaftlichen Bildungsforschern und -forscherinnen stimmt aber optimistisch, dass bald die gewählten Forschungstheorien und Forschungsmethoden nur durch ein handlungsleitendes Forschungsinteresse impliziert sind.

Literatur

- Deutsche Gesellschaft für Erziehungswissenschaft (DGfE) (2008): Kerncurriculum Erziehungswissenschaft. Opladen: Barbara Budrich.
- Drinkmann, A. (1990): Methodenkritische Untersuchungen zur Metaanalyse. Weinheim: Deutscher Studien Verlag.
- Fend, H. (2009): Bildungsforschung von 1965 bis 2008. In: Wischer, B./Tillmann, K.J. (Hrsg.): Erziehungswissenschaft auf dem Prüfstand. Weinheim u.a.: Juventa, S. 15-33.
- Gräsel, C. (2011): Was ist Empirische Bildungsforschung? In: Reinders, H./Ditton, H./Gräsel, C./Gniewosz, B. (Hrsg.): Empirische Bildungsforschung. Strukturen und Methoden. Wiesbaden: VS, S. 13-27.
- Klieme, E./Hartig, J. (2007): Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In: Zeitschrift für Erziehungswissenschaft, Sonderheft 8, S. 11-29.
- Klieme, E./Rakoczy, K. (2008): Empirische Unterrichtsforschung und Fachdidaktik. In: Zeitschrift für Pädagogik 54, H. 2, S. 222-237.
- König, E./Zedler, P. (1998): Theorien der Erziehungswissenschaft. Weinheim/Basel: Beltz.
- Lay, W.A. (1912): Experimentelle Pädagogik. Leipzig: Teubner.
- Scheerens, J./Bosker, R.J. (1997): The Foundations of Educational Effectiveness. Oxford: Pergamon.
- Schwippert, K./Goy, M. (2008): Leistungsvergleichs- und Schulqualitätsforschung. In: Helsper, W./Böhme, J. (Hrsg.): Handbuch der Schulforschung. 2., durchges. und erw. Auflage. Wiesbaden: VS, S. 387-421.
- Stern, E./Grabner, R./Schumacher, R. (2007): Lehr-Lern-Forschung und Neurowissenschaften: Erwartungen, Befunde und Forschungsperspektiven. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Terhart, E. (2009): Theorie der Schule. In: Wischer, B./Tillmann, K.J. (Hrsg.): Erziehungswissenschaft auf dem Prüfstand. Weinheim u.a.: Juventa, S. 35-50.
- Terhart, E. (2011): Zur Situation der Fachdidaktiken aus Sicht der Erziehungswissenschaft: konzeptionelle Probleme, institutionelle Bedingungen, notwendige Perspektiven. In: Bayrhuber, H./Harms, U./Muszynski, B./Ralle, B./Rothgangel, M./Schön, L.-H./

- Vollmer, H.J./Weigand, H.-G. (Hrsg.): Empirische Fundierung in den Fachdidaktiken. Münster u.a.: Waxmann, S. 241-256.
- Tillmann, K.J./Dedering, K./Kneuper, D./Kuhlmann, C./Nessel, I. (2008): PISA als bildungspolitisches Ereignis. Wiesbaden: VS.
- Tippelt, R. (2009): Steuerung durch Indikatoren!? In: Tippelt, R. (Hrsg.): Steuerung durch Indikatoren. Opladen: Barbara Budrich, S. 7-15.
- Tippelt, R./Schmidt, B. (2010): Einleitung der Herausgeber. In: Tippelt, R./Schmidt, B. (Hrsg.): Handbuch Bildungsforschung. 3. Aufl. Wiesbaden: VS, S. 9-19.
- Zedler, P./Döbert, H. (2010): Erziehungswissenschaftliche Bildungsforschung. In: Tippelt, R./Schmidt, B. (Hrsg.): Handbuch Bildungsforschung. 3. Aufl. Wiesbaden: VS, S. 23-42.

Knut Schwippert, Prof. Dr., geb. 1965, Universitätsprofessor für Erziehungswissenschaft mit dem Schwerpunkt Internationales Bildungsmonitoring und Bildungsberichterstattung an der Universität Hamburg.

Anschrift: Universität Hamburg, Fakultät für Erziehungswissenschaft, Psychologie und Bewegungswissenschaft, Binderstr. 34, 20146 Hamburg
E-Mail: knut.schwippert@uni-hamburg.de