

Die Wirksamkeit der Neuen Steuerung – theoretische und methodische Probleme ihrer Evaluation

Helmut Fend

Zusammenfassung: Im bildungspolitischen Diskurs der deutschsprachigen Länder über die Konsequenzen nach PISA ist eine neue „evidenzbasierte Steuerung“ des Schulwesens vorgeschlagen, deren Kern in der Formulierung von Bildungsstandards und deren testbasierter Überprüfung sowie in einem neuen Wechselspiel von lokaler Autonomie und zentralen Kontrollformen besteht. Wer ein evidenzbasiertes Konzept verfolgt, der wird sich um eine empirische Untermauerung der Hypothese bemühen, dass mit der neuen Steuerung auch positive Wirkungen im Sinne verbesserter Lernergebnisse verbunden sind.

Im vorliegenden Beitrag wird deshalb in einem ersten Schritt versucht, diese Erwartungen, insbesondere jene an die Standards und ihre Messung als Instrumente der Steigerung von Lernergebnissen, in den Kontext der „school-effectiveness“-Forschung zu stellen. In einem zweiten Schritt wird geprüft, ob die neuen Konzepte der Steuerung, eingebettet in die politologischen Traditionen von „Governance“-Konzepten, neue Untersuchungschancen der Hypothese enthalten, dass durch Verfahren wie jene der Standardsicherung und Schulautonomie Verbesserungen im Bildungswesen zu erzielen sind. Vorausgeschickt wird eine kurze Klärung, was unter „neuer Steuerung“ heute verstanden wird.

Schlüsselwörter: Steuerung · Governance · Schulqualitätsforschung · Bildungsstandards · Forschungsmethodologie

The effectiveness of the new governance—theoretical and methodological problems of evaluating governance

Abstract: In the educational debate in German speaking countries after PISA a new “evidence-based governance” of schooling has been proposed which, in its core, consists of performance standards, standard-related testing and a new balance of local autonomy and central control. Whoever is advocating an evidence-based system, will also be interested in empirically substantiating the hypothesis that “new governance” is connected with positive effects, i. e. with improved results of learning.

Angenommen: 07.12.2010

© VS Verlag für Sozialwissenschaften 2011

Prof. Dr. H. Fend (✉)

Neuhauserstr. 6, 78484 Konstanz, Deutschland

E-Mail: Helmut.Fend@t-online.de

The paper at hand places these expectations—particularly, those related to performance standards and to their assessment as instruments of improving learning outcomes—into the context of school effectiveness research. In a second step the paper discusses whether or not new concepts of governance in education (which have been developed from roots in political sciences) offer new chances for testing the hypothesis that instruments, such as performance standards and school autonomy, will allow to achieve substantial improvements in education. As an introduction, the meaning of “new governance” in contemporary education is clarified.

Keywords: Governance · School effectiveness research · Performance standards · Research methodology

Im konsequenzenorientierten Diskurs nach PISA hat eine als Paradigmawechsel apostrophierte neue Bildungspolitik eine große Rolle gespielt. Sie sollte einen grundlegenden Wandel von einer Inputsteuerung zu einer Outputsteuerung vollziehen. Als Kern dieser „neuen Steuerung“ haben sich die Formulierung von Bildungsstandards und ihre testbasierte Überprüfung herauskristallisiert. Gleichzeitig wurden neue Formen der Regulierung des Schulwesens eingeführt. Im Wechselspiel von Autonomie und Kontrolle wurde vielerorts die klassische Form der Kontrolle durch Schulaufsicht ersetzt durch eine Neugestaltung der Entscheidungsräume auf Schulebene im Sinne einer größeren Autonomie und Entscheidungskompetenz der Schulleitungen bei einer gleichzeitig stärker schulextern organisierten Evaluation (Heinrich 2007).

Diese neue Konfiguration der Steuerung, die an der Oberfläche *Restrukturierungen der Kontrolle des Bildungswesens* enthält und weniger *ressourcenintensive Unterstützungsleistungen*, stand im Kontext einer bildungspolitischen Gesamtstrategie, die diese stärker als bislang „evidence-based“ machen wollte, also solche Maßnahmen als implementationswürdig ansieht, die nachweislich positive Wirkungen haben. Sie sollte aus dem häufig von Glaubensüberzeugungen geprägten Streit über die notwendigen Reformen im Bildungswesen herausführen.

Wer konsequent ein evidenzbasiertes Konzept verfolgt, der wird sich auf eine empirische Untermauerung der Hypothese verlassen wollen, dass mit der neuen Steuerung auch positive Wirkungen im Sinne verbesserter Lernergebnisse verbunden sein werden. Wenn man dies tut, bewegt man sich unwillkürlich in das große Feld der „school effectiveness“-Forschung, das sich seit vielen Jahrzehnten mit den wirksamen Faktoren für die Verbesserung von Schulleistungen beschäftigt (Creemers et al. 2000; Luyten 2003; Scheerens et al. 2006; Sørensen und Morgan 2000; Teddlie und Reynolds 2000; van der Wolf 1988).

Deshalb soll hier in einem ersten Schritt versucht werden, die obigen Erwartungen, insbesondere jene an die Standards und ihre Messung als Instrumente der Steigerung der Lernergebnisse, in den Kontext der „school-effectiveness“-Forschung zu stellen. In einem zweiten Schritt soll geprüft werden, ob die neuen Konzepte der Steuerung, eingebettet in politologische „Governance“-Traditionen (Benz 2007), neue Untersuchungschancen der Hypothese enthalten, dass durch Verfahren wie jene der Standardsicherung und Schulautonomie Verbesserungen im Bildungswesen zu erzielen sind. Vorausgeschickt sei eine kurze Klärung, was unter „neuer Steuerung“ heute verstanden wird (s. bs. Altrichter et al. 2007; Altrichter und Maag Merki 2010; Berkemeyer 2010).

1 „Neue Steuerung“ und die Überprüfung von „Systemsteuerungen“

Gibt es bereits eine „Neue Steuerung“ in der Praxis und welche sind die *Kernelemente der Reformpraxis im Sinne einer neuen Governance*? Es erscheint trivial, wenn festgestellt wird, dass die Bildungsforschung darauf angewiesen ist, dass es diese Interventionen *gibt*, dass Verfahren der Standardbeschreibung und der Standardmessung entwickelt und damit soziale und kulturelle Wirklichkeit werden. Sie zu schaffen erfordert aber viele Ressourcen und je nach Anspruch an Qualität einen langen Zeitraum.

In der Tat sind in allen deutschsprachigen Ländern große Anstrengungen unternommen worden bzw. sind immer noch unterwegs, Verfahren der Neuen Steuerung zu entwickeln und zu implementieren. In Österreich wird die Entwicklung von Standards und entsprechenden Testinstrumenten für die vierten und achten Klassen in den Fächern Deutsch, Mathematik und Englisch noch mehrere Jahre beanspruchen und einen zweistelligen Euromillionenbetrag benötigen. Ähnliches gilt für Instrumente zur objektiven Leistungsmessung auf Maturaebene, die testtheoretischen Kriterien der Item-response-Konzepte genügen sollen.

Deutschland und die Schweiz gehen hier etwas andere Wege. In Deutschland sind es sogenannte standardbasierte Vergleichsarbeiten, die zu verschiedenen Zeitpunkten der Schullaufbahn Auskunft geben sollen, was Schülerinnen und Schüler können. Im Hintergrund stehen Einigungen der Kultusministerkonferenz über sogenannte Standards, die erreicht werden sollen. Im Institut für Qualitätsentwicklung in Berlin werden die Testungen in verschiedenen Bundesländern gebündelt, formuliert und für Bundesländervergleiche aufbereitet. Auf Abiturebene sind mit der Ausnahme von Rheinland-Pfalz alle Bundesländer zum Zentralabitur übergegangen. Dabei werden Aufgaben von ausgewählten Expert/innen der Lehrer/innenschaft entwickelt, zentral zu gleichen Zeitpunkten gestellt und in einem ausgeklügelten Verfahren von Erstbegutachtung, Zweitbegutachtung durch Lehrpersonen aus anderen Schulen und bei abweichenden Voten durch Dritte bewertet.

In der Schweiz stehen Bemühungen im Vordergrund, die Standards in die Entwicklung neuer Lehrpläne zu integrieren, um Inhalte und Kompetenzen in ihren ergänzenden Funktionen sichtbar zu machen. Um Einblick in das mit anderen Schulklassen vergleichbare Niveau der Schulleistungen einer Schulklasse zu bekommen, können Lehrpersonen auf kostenpflichtige Tests, die über Internetplattformen (Cockpit, Stellwerk) erhältlich sind, zurückgreifen. Die Initiative dazu muss aber von den Lehrpersonen bzw. von der Schulpflege kommen, die schließlich die Kosten zu tragen hat.

An allen Orten werden große Investitionen in die neuen Verfahren getätigt. Dies macht den Wunsch verständlich, eine wissenschaftliche Erfahrungsbasis zur Wirksamkeit dieser Maßnahmen zu schaffen. Die Ausschreibung eines Forschungsschwerpunktes „Systemsteuerung“ durch das BMBF in Deutschland stellt eine erste Anstrengung in diese Richtung dar. Gleichzeitig wird dabei deutlich, was die *Hypothesen* dieser Forschung über Systemsteuerung sein können und was sich als Forschungsgegenstand im Umkreis von Governance herauskristallisiert hat. In der Ausschreibung heißt es:

„Gefördert werden theorie- und hypothesengeleitete empirische Forschungsvorhaben zu Fragen in folgenden Bereichen:

1. Umfassende Steuerungsinnovationen: Steuerungen (z. B. mittels Evaluation, Schulin-spektion, stärkerem Management, mehr Wettbewerb, Bildungsstandards und zentraler Abschlussprüfungen) erfordern in vielfacher Weise eine veränderte, oftmals erhöhte vertikale und horizontale Handlungskoordination zwischen Akteuren im Bildungswesen. Insbesondere auch im Zuge stärkerer Eigenverantwortung ist eine verstärkte Rückmeldung der Ergebnisse von spezifischen Steuerungsinstrumenten und -prozessen zwischen verschiedenen Ebenen notwendig... Untersucht werden soll die Frage, unter welchen Rahmenbedingungen welche Instrumente, die in den letzten Jahren im Bildungsbereich verwendet wurden, welche Effekte für den Erhalt und den Ausbau der Qualität von Bildungssystemen erzielen...
2. Umstrukturierung der Bildungsverwaltungen: Mit der Umstellung auf output-orientierte Steuerung verändern Bildungsverwaltungen ihre Ausrichtung, Tätigkeit und Funktion, um angesichts neuer Steuerungserfordernisse ihrer koordinierenden Funktion im Bildungssystem gerecht zu werden. In welcher Weise und mit welchem Erfolg werden diese Umstrukturierungsprozesse gestaltet?...
3. Steuerungswirkung im Hinblick auf Unterricht: Im Mittelpunkt steht die Frage, welche Wirkungen neue Formen der Steuerung im Hinblick auf innerschulische Koordination und Strukturbildung, auf Unterrichtsgestaltung sowie für Lernergebnisse der Schülerinnen und Schülern haben...“ (<http://www.bmbf.de/de/13417.php>).

In diesem Rahmenkonzept haben die Antragsteller/innen jeweils ihre Akzente gesetzt. Die zentralen Projekte, die gefördert wurden, sind dann im Umkreis der neuen Governance angesiedelt (<http://www.empirische-bildungsforschung-bmbf.de/zeigen.html?seite=8832>).

So formuliert eines der geförderten Projekte:

Zentrale standardbasierte Tests von Schülerleistungen bilden das Kernstück einer outputorientierten Steuerung von Bildungsqualität. Sie dienen der empirischen Überprüfung von Bildungsstandards und sollen – vermittelt über die Rückmeldung von Testergebnissen – die Leistungsfähigkeit des Bildungssystems sichern bzw. steigern.

Hier wird eine klare Wirkungsthese formuliert, die mit der *Intervention*, mit dem „*Treatment*“, verbunden ist.

In Bezug auf diese großen Interventionen der Neuen Steuerung lässt sich in der Tat, z. B. in der Schweiz, in Deutschland und in Österreich, auf Governance-Ebene ein *faktischer Konsens* in der derzeitigen Qualitätssicherung im Bildungswesen beobachten. *Zentrale Prüfungen, überschulische Vergleichsarbeiten*, auf vereinbarte *Standards* und *Kompetenzen* ausgerichtete Bildungspläne werden in allen Ländern akzeptiert. Begleitet sind sie von *Qualitäts-Referenzrahmen*, die in eindrucksvoller Entfaltung *Qualitätsstandards für einzelne Schulen* präzisieren, und von weitgehenden Verfahren der *Autonomie* und *Führungsfähigkeit* von Schulen vor Ort, die nicht selten in einem produktiven Qualitätswettbewerb stehen. Standardsicherungsmodelle und Autonomiemodelle ergänzen sich dabei, um die *Qualitätsentwicklung* und die *Qualitätssicherung* funktionsfähig zu machen.

So scheinen wir auf dem Wege zu einer neuen *Grammatik der Qualitätssicherung*: Auf Systemebene sichert ein *kontinuierliches Monitoring* (Arbeitsgruppe Bildungsbe-

richterstattung 2008; Böttcher et al. 2008) den Wissensstand über Stärken und Schwachstellen im Bildungswesen. Standards und Kompetenzerhebungen lassen dies bis auf die Schule- und Klassenebene ausweiten. Da sich die geschlossenen Systeme und die *Rückmeldekultur* als defizitär erwiesen hatten, kann das Qualitätsbewusstsein in Bezug auf die Leistungen durch *überschulische Vergleichs- und Diagnosearbeiten* gestärkt werden. Damit hätten auch die Schulleitungen die Chance, Einblicke in die Qualität der Lehre, aber auch Einblicke in die Lernvoraussetzungen der Schüler/innenschaft zu gewinnen. *Qualitätsarbeit* würde erstmals sichtbar und verhandelbar. Damit hätten sich der verobjektivierbare Stand des Wissens über das Bildungswesen und vor allem auch das *Qualitätsbewusstsein* wesentlich verbessert.

In der Summe sind also zwei Kernelemente der Interventionen sichtbar: solche der Standardsicherung und solche der Qualitätssicherung auf schulischer Ebene über Autonomiegewährung und Evaluation, die wiederum im Kontext einer rechtlichen Veränderung von Verantwortlichkeiten steht, die verwaltungstechnisch im Rahmen des New Public Management platziert ist (s. dazu Dubs 1996; Schedler 1995). Hier geht es auch um die Veränderung der Schulaufsicht, um neue Führungsstrukturen, um die Entwicklung eines Qualitätsrahmens für die Schulen und um Verfahren der internen und externen Evaluation. Sie konnten auf die schon länger laufenden Initiativen und Entwicklungen von „selbständigen Schulen“ aufbauen (Berkemeyer 2010).¹

2 Neue Steuerung im Rahmen des empirisch-analytischen Forschungsparadigmas der school effectiveness-Forschung

Die Neue Steuerung steht im Kontext einer Auffassung, nach der Veränderungen und Entwicklungen im Bildungswesen auf der Grundlage ihrer nachgewiesenen Wirksamkeit, also „evidence-based“ erfolgen sollen. Diesem Anspruch muss sie sich konsequenterweise in gleicher Weise aussetzen. Das Erfolgskriterium ist dabei die Steigerung der Lernergebnisse. Sie ordnet sich damit in den Kontext der school effectiveness-Forschung ein, die sich mit der Prädiktion von Schulleistungen beschäftigt. Im Kontext eines umfassenden Wirkungsmodells müsste dann lokalisiert werden, welche „Wirkkraft“ Verfahren der Standarderreichung über flächendeckende Testverfahren zukommt, um Lernergebnisse positiv zu beeinflussen.

Welche Modelle und Forschungsergebnisse stehen dabei zur Verfügung? Als allgemeines Wirkmodell hat sich das Input-Prozess-Output und Kontext-Modell der OECD als hilfreich für die Ordnung der zentralen Faktorengruppen erwiesen. Bestmöglicher Input, bestmögliche Prozesse und günstige Kontexte fördern die Lernergebnisse, den „Output“.

Welche Faktoren sind dabei entscheidend? Die Grundlinien moderner Antworten, wie sie z. B. in der holländischen Schulforschung formuliert werden (Creemers et al. 2000; Creemers und Kyriakides 2008; Scheerens und Bosker 1997) sind erstaunlich einfach und einhellig. Auf Seiten der Institution sind vor allem *Zeitinvestitionen* und *inhaltliche Lerngelegenheiten* ausschlaggebend, die auf Makro-, Meso- und Mikroebene angesiedelt sein können. Dazu kommen *Qualitätsindikatoren* des Angebotes, die auf Unterrichts-

ebene (Helmke 2009) über verschiedene Studien, z. B. über Videostudien, immer wieder erarbeitet wurden (Klieme et al. 2006, 2009): Es sind dies

- ein effizientes und funktionales Klassenmanagement mit möglichst wenig Zeitverlust,
- eine gute Strukturierung des Unterrichtes,
- ein kognitiv aktivierender Unterricht,
- eine gute Abstimmung des Unterrichtes auf die motivationalen und emotionalen Bedürfnisse der Schüler/innenschaft,
- ein Unterricht, der differenzierend und die Schüler/innen beteiligend auf die Heterogenität der Lernstände eingeht.

Quantität und Qualität des Angebotes erweisen sich als die zentralen Größen, die die Lernergebnisse bestimmen. Eine zweite große Gruppe von Faktoren, die Lernergebnisse beeinflussen, ist dann auf Schüler/innenseite zu suchen. Kognitive Lernvoraussetzungen und motivationale Konstellationen stehen dabei im Mittelpunkt. Sie beeinflussen die Qualität der Nutzung von Lernangeboten. An der Schnittstelle von Angebot und Nutzung zeigen sich zusätzliche Qualitätsmerkmale. Sie bestehen in der *Adaption* von gewährter mit benötigter Lernzeit (Carroll 1963). Dieses Prozessmerkmal guten Unterrichts realisiert sich in aufeinander abgestimmten ko-konstruktiven Prozessen zwischen Lehrpersonen und Schüler/innen (Reusser 1998).

Dieses Modell der Leistungswirksamkeit ließe sich formalisieren (Fend 2008, S. 24 ff.):

- (1) $LE = f(A, N, A * N)$
LE=Lernergebnisse
A=Angebot
N=Nutzung
- (2) $A = f(\text{QuantA}, \text{QualA})$
QuantA=Allokation von Zeit und Lerngelegenheiten
QualA=Qualitätsmerkmale von Unterricht
- (3) $N = f(\text{QuantN}, \text{QualN})$
QuantN=Zeitinvestitionen in Lernen
QualN=kognitive Lernvoraussetzung, Motivation
- (4) $A * N = \text{Interaktion von Angebot und Nutzung (Adaptivität)}$

Wenn also mehr und qualitativ bessere Angebote gemacht werden und diese mit mehr Zeitinvestitionen und Motivation auf Schüler/innenseite genutzt werden, dann resultieren daraus auch Steigerungen der Lernergebnisse.

Die school effectiveness-Forschung, die hier angesprochen ist, konzentriert sich schwerpunktmäßig auf den Kern des schulischen Geschehens, auf Lehren und Lernen. Die Qualität der Mikrostruktur schulischer Lehrprozesse steht damit im Vordergrund (s. dazu Creemers und Kyriakides 2008). Die Neue Steuerung scheint von dieser Ebene des Handelns weit entfernt. Welchen Stellenwert können im Rahmen eines mikroorientierten, quantitativ arbeitenden Modelldenkens aus der Makrosteuerung des Bildungswesens kommende Standards und deren Testung für die Steigerung der Lernergebnisse haben? Nach dem obigen Modell wäre dies nur dann der Fall, wenn die Makrosteuerung Kernbedingungen für die Steigerung von Lernergebnissen beeinflussen würde.

Die empirische Forschungslage dazu hat Maag Merki (2010) im Handbuch zur Neuen Steuerung referiert. Sie ist nicht so eindeutig, wie man es wünschen würde. Internationale Vergleiche sind eine zentrale Quelle, da in ihnen Variationen in Makrofaktoren des Bildungswesens gefunden werden können. Wößmann (2002) hat die Datenbasis der TIMSS-Studien genutzt, um zu überprüfen, welche Makrofaktoren einen Einfluss auf die Leistungsergebnisse haben. Er glaubte nachweisen zu können, dass jene Länder anderen überlegen sind, die standardisierte Messungen auf Abiturniveau eingerichtet haben. Bei den Maßnahmen zur Stärkung der Schulautonomie war dies nicht der Fall. Doch auch hier kovarierte das Instrument der zentralen Prüfungen als Instrument auf der Makroebene mit anderen Faktoren, so dass eine exakte kausale Zuordnung schwierig war. Ähnliches gilt für die Übertragbarkeit von amerikanischen Studien zum „high stakes testing“, wonach getestete Schulleistungen gravierende Konsequenzen für Lehrpersonen und auch Schüler/innen haben. Auf deutsche und österreichische Bildungssysteme sind die Ergebnisse schon deshalb nicht übertragbar, weil ein solches Testen bislang nicht umfassend implementiert ist. Eine Ausnahme bilden jedoch die Abiturprüfungen in den meisten Bundesländern in Deutschland. Da sie noch nicht lange eingeführt wurden, eröffnet sich die Chance, ihre Wirksamkeit zu überprüfen (Maag Merki 2010).

Bei einer *kausalen* Zuordnung müssten sich Faktoren auf der Makroebene auf Quantität oder Qualität des Lernangebotes und die bestmögliche Zeitnutzung auswirken. Sie müssten das Qualitätsbewusstsein in Bezug auf zu erreichende Lernziele stärken und so zu einer Fokussierung der Lehranstrengungen und zu einer Stärkung der Lehrer/innenmotivation führen. Auf Schüler/innenseite müsste nachgewiesen werden, dass mehr Motivation und Zeit in die Nutzung der besseren Lernangebote investiert werden.

Kompetenzzuwächse müssten nach den Forschungen zu „school effectiveness“ über die Veränderung jener proximalen Faktoren vermittelt werden, die mit Kompetenzsteigerungen zusammenhängen. Auf diesem Hintergrund hat es die empirisch analytische Forschung schwer, nachzuweisen, dass mit der neuen Standard-Governance auf *Makroebene* Steigerungen in den Lernergebnissen auf der *Mikroebene* erzielt werden können. Sie wird auf dem Hintergrund ihrer Forschungsansätze keine starken *direkten* Wirkungen der neuen Steuerung auf Leistungsergebnisse vermuten und Hypothesen eher für *indirekte* Wirkungen, vermittelt über die proximalen Faktoren, formulieren (Scheerens et al. 2006).

Der Nachweis der Wirksamkeit der Neuen Steuerung ist jedoch ein anspruchsvolles Vorhaben und zwar nicht zuletzt deshalb, weil die wissenschaftlichen Kriterien, Wirkungen nachzuweisen, sehr rigoros sind:

1. Ein empirischer Nachweis wäre darauf angewiesen, ein *quasi-experimentelles Design* zu realisieren, das Kontrollgruppen oder zumindest Vergleichsgruppen als Bezugsgrößen erfordern würde und in einem Pre-Post-Testdesign positive Veränderungen auf der Mikroebene des unterrichtlichen Angebotes durch Makrofaktoren nachzuweisen hätte. Wenn ein ganzes Bildungswesen von einer Neuen Steuerung betroffen ist, dann fehlen diese Möglichkeiten, es sei denn, die zeitliche Umstellung in verschiedenen Sektoren des Bildungswesens bzw. in verschiedenen Ländern kann dazu klug genutzt werden. Ansonsten setzt dieser Ansatz den Vergleich mehrerer Bildungssysteme mit unterschiedlichen Steuerungsmodalitäten voraus. Durch die internationalen Schul-

Leistungsvergleiche ist allerdings die Datenlage so gut wie noch nie in der Geschichte der Bildungsforschung.

2. Erfahrungsgemäß lässt sich eine Intervention wie jene der „Standardmessung“ nur schwer als „reine“ Variable isolieren. In der Regel sind solche Interventionen *konfundiert* mit anderen Faktoren (curriculare Veränderungen, neue Lehrer/innenbildung, finanzielle Ressourcen, rechtliche Kompetenzverteilungen). Besonders akut ist dieses Problem bei der Versuchung, Ergebnisse aus anderen Ländern auf das jeweilige nationale Bildungssystem zu übertragen. Einzelne Maßnahmen stehen in der Regel in einem systemischen Kontext, im Kontext einer „stimmigen“ *Grammatik* eines Bildungswesens, der es erschwert, sie als Einzelvariablen zu isolieren und in ihren Wirkungen zu bewerten.
3. Entsprechende Implementationen auf Makroebene – wie z. B. jene der Standardmessung – sind *langfristige* Unternehmungen; dies nicht nur was die Entwicklung, sondern auch was die Zeiträume für Effekte angeht. Fortschritte werden jedoch häufig schnell erwartet, und Reformen können nicht so lange hinausgeschoben werden, bis entsprechende Forschungsergebnisse vorliegen. Damit besteht für ein Forschungsdesign die Gefahr, dass sich das „Treatment“, also die Intervention in der Gestalt der Steuerung des Bildungswesens, fortlaufend ändert.

Diese Schwierigkeiten der praktischen Realisierung von empirischen quantitativen Studien zur Wirksamkeit unterschiedlicher Makrofaktoren sprechen nicht grundsätzlich gegen den empirisch-analytischen Ansatz. Sie dämpfen lediglich die Erwartung, klare kausale Zuordnungen von Makrostrukturen zu Effekten auf der Lernebene finden zu können und dies in einem Zeitrahmen, der mit den politischen Entscheidungen zu Veränderungen im Bildungswesen kompatibel ist. Es dürfte sich aber lohnen, hier alle Möglichkeiten auszuschöpfen.²

Schwerer als die Schwierigkeiten der praktischen Realisierung klarer Wirksamkeitsdesigns wiegt die *theoretische* Frage, wie die *Mechanismen*, die die erwünschten Steigerungen in den Lernergebnissen bewirken könnten, gefunden und belegt werden können. Die lange *Wirkungskette* von Standardformulierungen und daran gebundenen Testreihen zu besseren Mathematikkenntnissen, Englischleistungen, Lese- oder Orthographieleistungen auf Schüler/innenebene erfordert eine Spezifizierung der vermittelnden Prozesse zwischen der Makrosteuerung bis auf die Ebene der Kompetenzsteigerungen bei Schülerinnen und Schülern. Dazu bedarf es theoretischer Konzepte, die dies darstellbar machen. Sie müssten dann Hypothesen ermöglichen, wie Unterschiede in der Makrosteuerung zu Unterschieden auf der Mikroebene des Lehrens und Lernens führen. Das zentrale *methodische* Problem ergibt sich aus der Forderung, für solche Vermittlungsprozesse *empirische Belege* zu erbringen.

3 Theoretische und methodische Beiträge von Governance-Konzepten zur Wirksamkeitsanalyse

Bei der Suche nach theoretischen Konzepten, die bei der Aufklärung von Wirkungsketten zwischen der Makrosteuerung des Bildungswesens und Effekten auf der Schüler/inne-

nebene hilfreich sein könnten, sind Bildungsforscher/innen auf jene des *Educational Governance* gestoßen (s. bs. Altrichter et al. 2007; Altrichter und Maag Merki 2010; Berkemeyer 2010). Sie weckten die Hoffnung, dass durch sie die Zusammenhänge zwischen mehreren Steuerungsebenen des Bildungswesens dargestellt werden könnten, also auch jene zwischen der auf Makroebene angesiedelten „Neuen Steuerung“ und dem Mikrogeschehen des Unterrichts.

Governance-Konzepte (Benz 2007) stammen aus der Politikwissenschaft, erheben aber den Anspruch, Modelle zu sein, die hilfreich sind, um Regulierungsformen und das Zusammenspiel von Institutionen, Akteuren und ihren Handlungsstrategien in *allen* Sektoren staatlicher Verantwortung zu beschreiben. Ob die Aufklärung von *Prozessen im Rahmen von Institutionen*, die durch die Governance-Konzepte beansprucht wird, auch auf den Institutionsbereich Bildungswesen übertragen werden kann, gilt es hier zu überprüfen. Das Ziel besteht darin, theoretische Konzepte für die mögliche Wirksamkeit von Makrosteuerungen zu finden, etwa jene von Standards, Tests und Evaluationsverfahren (s. z. B. Heinrich 2007; Kussau und Brüsemeister 2007; Langer 2008).

In der politologischen Diskussion werden seit geraumer Zeit die Grenzen staatlicher Steuerung diskutiert. Staatliche Gesetzgebung und gesetzeskonformer Vollzug sind vielfach keine ausreichenden Erklärungen für gesellschaftliche und institutionelle Veränderungen mehr. Dies gilt für die *Genese* von Entscheidungen, die durch die formalisierten Verfahren allein oft nicht verstehbar sind, und für deren *Ergebnisse*, die oft nicht mehr durch schlichte Anweisungen durchsetzbar sind. Politik gilt bei prononcierten Vertreter/innen der Governance-Konzepte als kultureller Prozess der interdependenten, netzwerkintensiven und interpretativen Wirklichkeitskonstruktion (Bevir und Rhodes 2010; Rhodes 2010; Rhodes et al. 2009). Verstärkt gilt dies für die Wirkungsanalyse von politischen Maßnahmen und deren Umsetzung in Verfahren der Wirklichkeitsgestaltung.

Die moderne Governanceforschung beschäftigt sich mit den interdependenten Interaktionen von Akteuren auf verschiedenen Handlungsebenen des politischen Entscheidungsprozesses, etwa jene im Umkreis der Europäischen Union. Deren prozessualer Gestalt, Reflexion und Wertung sowie deren Folgewirkungen widmet sie ihre beobachtende Aufmerksamkeit. Die Koppelung dieser Ebenen wird dabei besonders beachtet. Somit wird der institutionelle Rahmen, in dem sich diese Prozesse abspielen, nicht ausgeblendet, sondern als Orientierungsfolie, „was eigentlich formell ablaufen müsste“, und als Regulierungsgröße vorausgesetzt. Zusätzlich liegt aber die Betonung darauf, dass die politischen Prozesse und deren Umsetzungen von Akteuren (*situated agents*) erfolgt, die intelligent wahrnehmen und präferenzengelenkt handeln. Diesen Prozessen wären danach auch institutionelle Strategien wie jene der Neuen Steuerung im Bildungswesen ausgesetzt.

Governance ist somit ein Ansatz, um die Frage zu beantworten, wie sozialer Wandel angesichts sich verändernder staatlicher Regulierungen und angesichts sich verändernder sozialer Wirklichkeiten geschieht und wie dieser über neue emergente Regulierungen von intelligenten Akteuren mitgestaltet wird. Die staatliche Regulierung wird damit insofern „verflüssigt“, als sie mit neuen Strukturen, mit neuen Akteuren (z. B. den non-governmental organizations), neuen Handlungsinstrumenten (z. B. Öffentlichkeitsarbeit und Monitoring) und neuen Legitimationsprozessen (etwa durch Partizipation) zu arbeiten beginnt.

Forschungsmethodisch sind damit auch neue Ansprüche verbunden. Die Aufmerksamkeit hat sich jetzt neben der Analyse *institutioneller Vorgaben* darauf zu richten, *wie die Akteure auf den verschiedenen Handlungsebenen interagieren*, wie sie sich aufeinander beziehen, welche Wirklichkeiten sie gemeinsam oder getrennt konstruieren, welche *Interessen* und *Handlungslogiken* sie jeweils folgen und welche „Neuerungen“ sie als „*findige Akteure*“ (Joas 1996) dabei entwickeln. Die alten Dichotomisierungen in „Bottom up“ Veränderung im Gegensatz zu „Top down“ Strategien werden dabei ebenso in Netzwerke und Interdependenzen verfeinert, wie die Dichotomie von zentral und lokal aufgelöst wird in das Zusammenspiel mehrerer Handlungsebenen. Eine wesentliche Komponente dieser neuen Strategien und ihrer Erforschung ist die Betonung seiner *subjektiven und mentalen Komponenten*. Subjektive Sichtweisen der Wirklichkeit und Interessen der beteiligten Akteure könnten auch die neuen Steuerungsprozesse begleiten. Sie könnten die offiziellen Zielsetzungen hemmen, fördern oder sie in möglicherweise ungewollte Richtungen drängen.

Lassen sich diese Konzepte nahtlos auf Bildungspolitik und Bildungsinstitutionen übertragen?

Einige Parallelen sind unübersehbar. Die Publikationen der letzten Jahre, stellvertretend für viele sei das „Handbuch Neue Steuerung im Schulsystem“ (Altrichter und Maag Merki 2010) genannt, belegen einen eminenten Fortschritt in den Konzepten, wie die Steuerung des Bildungswesens verstanden werden kann. Sie nähern sich denen der Governanceforschung an. Die klassische Dichotomie, die in der Wahrnehmung vieler Akteure im Bildungswesen immer noch besteht, also jene zwischen einem politisch-administrativen System (Rumpf 1966) und einer am Wohl des Kindes orientierten Lehrer/innenschaft, hat sich in Konzepte differenzierter Interdependenzen bei der Steuerung des Bildungswesens verfeinert. Auch die alleinige Konzentration auf die Schule oder den Unterricht als Handlungsebenen der Reform trat zugunsten von Konzepten in den Hintergrund, die alle Handlungsebenen berücksichtigen, jene von der Politik bis zu den Schülerinnen und Schülern. Die Mehrebenenstruktur der Analyse von Bildungsprozessen ist in der Analyse von Bildungssystemen fest etabliert (Fend 2001, 2008). Auch die Orientierung an Akteuren, an ihren Interessen und Wahrnehmungen im Rahmen institutioneller Strukturen ist Teil neuer Konzeptualisierungen institutionalisierten Lehrens und Lernens. Im Anschluss an die PISA-Ergebnisse in Deutschland hat sich zudem wie in der Governancetheorie die Wahrnehmung eines *politisch-administrativen Steuerungsdefizites* aufgedrängt. Die „alten“ Instrumente der Qualitätssicherung wie Schulaufsicht und Schulinspektion waren danach nicht einmal in der Lage, die Leistungsfähigkeit des Bildungswesens zu *sehen*, geschweige denn zu regulieren, so dass deren Dokumentation bei PISA als Schock empfunden wurde. Der Zugzwang, in den die Bildungspolitik dadurch geraten ist, hat dazu geführt, dass die bildungspolitisch gestaltbaren Instrumente überdacht wurden. Sie bestanden dann in den hier geschilderten Strukturen neuer Steuerung.

4 Epistemologische Grundlagen von School effectiveness und Governance

Wenn auf die hier geschilderte Weise school effectiveness-Forschung und Governance-Konzepte einander gegenüber gestellt werden, dann drängt sich die Einschätzung auf,

den jeweiligen Ansätzen könnten unterschiedliche Auffassungen über den Charakter des untersuchten Wirklichkeitsbereichs zugrunde liegen. In der school effectiveness-Forschung erscheint die Wirklichkeit als ein Geflecht von Variablen, die auf ihre Wirksamkeit hin überprüft werden. Welche Handlungen ihnen zugrunde liegen, wird lediglich indiziert, aber nicht substantiell expliziert. Sie kommen in den Operationalisierungen als Faktoren zum Vorschein, ohne dass umfassend geklärt wird, durch welche Handlungen in einem institutionellen Kontext sie reguliert und eine faktorielle Ausprägung erfahren. „Tests“ oder „Nicht-Tests“ sind schlicht zwei dichotome Ausprägungen in einem Wirkungsgefüge, wenn es um deren prädiktive Kraft bei Leistungsvergleichen geht.

Systematische, die Wirklichkeit indizierende Forschung ist natürlich auf solche Operationalisierungen und mathematischen Abbildungen angewiesen. Sie bedarf aber zusätzlich theoretischer Konzepte, wie Handeln reguliert ist und wie die dadurch erzeugte Wirklichkeit abgebildet werden kann. Die Governance-Forschung hat dafür zwei Antworten bereit: Handeln in Institutionen ist *durch Regeln geleitet* und durch *kulturelle, akteurbezogene Interpretationen* inspiriert. Sie lässt damit die *Spezifika des gestalteten humanen Bereichs* nicht implizit, wie dies zu tun eine rein quantitativ indizierende Forschung in Versuchung ist, sondern macht sie auf dem Hintergrund ontologischer Grundannahmen über den Charakter des Sozialen explizit. Bei wichtigen Vertreter/innen der Governanceforschung, etwa bei Bevir und Rhodes (2010) ist dies der Fall. In der Abgrenzung der *kulturellen und sozialen Phänomene* von jenen der *Naturwissenschaften* wird dies schon bei Max Weber (1947) artikuliert. Danach gilt es, soziale Handlungsmodelle explizit zu machen und den Gesellschaftsbereich von den intentionalen und interpretativen humanen Strukturen her zu rekonstruieren.

Dies gilt auch für das soziale und kulturelle Feld von Bildungsinstitutionen. Die Unterrichtsforschung und Bildungsforschung haben es mit einem Bereich zu tun, der durch menschliche Handlungen und Überzeugungen kreiert worden ist und durch den Rückgang auf diese nicht nur erklärbar, sondern auch verstehbar ist. Der Bildungsbereich hat somit eine verstehensorientierte, akteurorientierte und institutionsorientierte Dimension. Die wissenschaftliche Analyse des Kulturbereichs „Bildungswesen“ besteht deshalb darin zu sehen, welche *institutionelle Gestalt* es hat (Fend 2006b), wie diese *historisch entstanden* (Fend 2006a) ist und in welchen *Prozessen* es sich realisiert (Fend 2008). Bildungsforschung ist damit auch eine Empirie der Gestaltungsprozesse bei der „Herstellung“ von Lernergebnissen. Damit wird sie zu einer Empirie der Prozesse und der Ergebnisse.

Die Weiterentwicklung einer Theorie des Bildungswesens hat zu einer theoretischen und metatheoretischen Architektur geführt, in der Schulen als *institutionelle Akteure der Humangestaltung* bzw. Menschenbildung betrachtet werden. Das Handeln der Einzelindividuen steht dabei im Rahmen von gesellschaftlich *vereinbarten* und *durchgesetzten* Regelungen, es ist *normativ reguliertes Zusammenhandeln*, das von *Regelungen*, spezialisiertem *Personal* und *kulturellen Ressourcen* getragen ist. Bildungssysteme sind damit institutionelle Akteure, die im Auftrage externer Akteure handeln und über Lehren und Lernen als wünschenswert definierte psychische Dispositionen in der nachwachsenden Generation „erzeugen“. Sie leisten über die Kulturvermittlung gleichzeitig „Humangestaltung“, „Seelenarbeit“ und „Menschenbildung“ im Sinne der Förderung von Wissen, Kompetenzen, psychischen Ressourcen und Werten. Als institutionelle Akteure können sie agieren, wenn sie drei Kernstrukturen zeigen: Programme des Lehrens, institutionelle

Regelsysteme sowie Kompetenzen und „Techniken“, das Programm auszuführen. „School is about changing people, making human kind, fabricating human kind“, so beschrieb Popkewitz in einem Zürcher Vortrag im Jahre 2005 die Struktur des Bildungswesens.

Schulen sind danach institutionell regulierte, von internen und externen Akteuren gestaltete sowie interpretativ getragene Handlungszusammenhänge, um ein inhaltliches Kulturprogramm in der nachfolgenden Generation zu realisieren und um diese in einer komplexen, arbeitsteiligen Gesellschaft handlungsfähig zu machen. An diese Kernaufgaben sind mehrere andere – wie die der sozialen Allokation und der politischen Stabilisierung der Gesellschaft – mehr oder weniger konfliktreich angebunden.

Eine Standardsicherung ist Teil des institutionellen Regelsystems. Sie ist human geschaffen, in einen größeren institutionellen Rahmen eingebettet, wird von Akteuren interpretiert und umgesetzt und von anderen rezipiert und transformiert.

Von den institutionellen Regelungen her ist der schulische Handlungszusammenhang differenziert in mehrere Handlungs- und Verantwortungsebenen (Makro-, Meso- und Mikroebenen). Auf allen Ebenen sind Akteure am Werk, die intellegibel und interessen-geleitet ihre Aufgaben interpretieren und ressourcengestützt ausüben. Das Konzept der *Rekontextualisierung* verweist dabei darauf, dass in diesem hierarchischen, aber von oben nach unten und von unten nach oben durchlässigen System regelgeleitete Formen des Zusammenhandelns auf der jeweiligen Verantwortungsebene (z. B. in der Schule, im Klassenzimmer) entstehen, die je besondere Handlungsinstrumente und Kompetenzen erfordern (Fend 2008, S. 34 ff.).

Auch diese Merkmale von Bildungssystemen im Allgemeinen treffen auf Standardsicherungen zu. Sie werden als professionelle Instrumente geschaffen, gesetzgeberisch abgesichert und auf mehreren Ebenen jeweils umgestaltet und auf die Interessen und Möglichkeiten der Akteure hin rekontextualisiert.

Das Kerngeschehen im Bildungswesen spielt sich zwischen den operativen Akteuren – den Lehrenden und Lernenden – ab. Hinter den lehrenden Akteuren steht jedoch eine kontrollierende und unterstützende Institution mit einem großen Kulturprogramm, das mit lernenden Systemen synchronisiert werden soll. Das lernende System, seine individuelle Lernfähigkeit, seine neuralen Strukturen und emotionalen Regulierungsprozesse nimmt das schulische Programm unterschiedlich schnell und perfekt auf und wird damit ein immer kompetenterer Kulturteilnehmer. Angebotsquantität und Angebotsqualität sind dabei jeweils unterschiedlich adaptiv zur Nutzungskompetenz und zur Nutzungsmotivation. Lehren bedeutet damit, in unzähligen Prozessen der Ko-Konstruktion ein adaptives Verhältnis von Lernangeboten und Nutzungschancen herzustellen.³ Auf einer solchen operativen Ebene muss eine Standardsicherung ankommen, wenn sie wirksam werden soll. Wie dies geschehen kann, gilt es im Rahmen eines solchen Paradigmas zu erforschen.

In dieser Prozessdarstellung zeigt sich eine gute Kompatibilität der epistemologischen Grundannahmen der Governance-Konzepte und ihrem Verständnis von Standardsicherung mit den Besonderheiten der *Gestaltung von Schule als kultureller Praxis im Mehrebenensystem*. Governance-Konzepte werden dann zu einem Teilaspekt schultheoretischer Modelle, die in den epistemologischen Grundannahmen konvergieren.

Die methodischen Ansätze der empirischen school effectiveness-Forschung artikulieren diese epistemologischen Grundlagen häufig nicht explizit. Dadurch wurde ihnen

von der Seite qualitativer Sozialforschung gelegentlich unterstellt, sie huldigten einem naturwissenschaftsnahen Kausalitätsmodell und dächten über die Wirksamkeit neuer Unterrichtskonzepte oder Steuerungskonzepte des Bildungswesen wie über die Wirksamkeit einer neuen Medizin. Analog sollte danach in möglichst sauberen experimentellen Designs überprüft werden, wie die neue Steuerung „wirkt“.

Experimentelles Denken, Strategien der Operationalisierung und empirische Kausalanalysen sind als formale Handlungsstrategien der methodisch geleiteten Fehlerreduktion von Aussagen bei wissenschaftlichen domainübergreifenden Arbeiten essentiell. Sie sind Teil einer sauberen Methodik in allen quantitativ arbeitenden Wissenschaften. Zusätzlich gilt es jedoch, die Besonderheiten verschiedener Wirklichkeitsbereiche zu beachten. Die Erforschung der sozialen und kulturellen Wirklichkeit legt ein anderes Verständnis der Prozesse und Handlungsregulierung nahe als die Erforschung der Natur. Kausalitäten resultieren in soziokulturellen Domänen aus Handlungsmodellen, Institutionsmodellen, interpretativen Modellen und Akteurmodellen, die die Naturwissenschaften nicht kennen.

Die Suche nach empirischen Zusammenhängen, etwa jene nach der Wirksamkeit von Neuen Steuerungsformen, muss von solchen Modellen geleitet sein. Das gleiche gilt, wenn empirische Relationen gefunden werden. Ihre Interpretation muss auf die obigen Modelle zurückgreifen. Sie dürfen nicht als „blinde“ Relationen, die dann kausal interpretiert werden, stehen bleiben.

5 Lange Wirkungsketten im Kontext schultheoretischer Modelle

Die entscheidende Frage nach der Fruchtbarkeit der institutionstheoretischen, handlungstheoretischen und mehrebenentheoretischen Konzeption von Schule ist in diesem Beitrag die, ob sie hilfreich ist, um die erwähnte lange Wirkungskette von Regulierungen auf der Makroebene bis zur Rekontextualisierung des schulischen Angebotes auf der Nutzungsebene der Schüler/innen zu überbrücken (s. den Kontext von Standardisierung und ihrer Umsetzung Oelkers und Reusser 2008). Welchen Stellenwert haben in einem solchen Modell die Standardverpflichtungen und deren Überprüfung durch objektive Testverfahren für die Steigerung der Lernergebnisse? Die Beantwortung dieser Frage ist der Test dafür, wie produktiv schultheoretische Konzepte und die in ihnen eingelagerten governancetheoretischen Konzepte sind, um neue Hypothesen und Forschungsansätze zu entwickeln.

Sie können in der Tat, so die hier formulierte Hypothese, auf Wege der „Wirksamkeit“ verweisen, die einem vorschnell quantifizierenden Ansatz verborgen bleiben. Dazu verhelfen folgende Strategien.

1. Der obige Ansatz der Konzeptualisierung des Bildungswesens als institutioneller Akteur⁴ ermöglicht in einem ersten Schritt eine *bessere Beschreibung der Großintervention* „Standardsicherung“. Dabei wird schon sichtbar, dass ihre Wirkungen und Nebenwirkungen von der *Konfiguration von Instrumenten und rechtlichen Regelungen* bzw. von Investitionen in begleitende Maßnahmen der Unterstützung abhängen (s. Abb. 1).

Hinter diesem Standardunternehmen und deren Testung stehen nämlich zwei kulturelle „Kreationen“. Einmal erfordern sie *politische Entscheidungen und rechtlich-institutio-*

	RECHTLICHE REGELUNGEN DER DURCHFÜHRUNG UND FOLGEN			
		Low stakes	High stakes für Schüler/innen	High stakes für Lehrpersonen
INSTRUMENTE DER ERGEBNISMESSUNG	intern lehrerabhängig und curriculumnah	Formative Evaluation	Klassenarbeiten mit Noten und Berechtigungen	
	extern klassenübergreifend und curriculumnah	Vergleichsarbeiten Aufgabenstellungen durch Vertreter/innen der Lehrerschaft	Zentralabitur Aufgabenstellungen durch Vertreter/innen der Lehrerschaft	
	extern testbasiert und curriculumübergreifend	Standardtestungen	Standardisierte Reifeprüfung	Standardtestungen mit Klassenkennung

Abb. 1: Kombination von rechtlichen Regelungen und Instrumenten der Standardmessung

nelle Regelungen darüber, wer was unter welchen Bedingungen mit welchen Folgen prüfen darf. Dafür gibt es mehrere Alternativen. Testungen können eingesetzt werden, damit die Lehrperson über den Lernfortschritt von Schülerinnen und Schülern informiert ist. Die Ergebnisse können dann nur der Unterrichtsplanung dienen und keine Konsequenzen für Noten oder Abschlüsse haben. Vergleichsarbeiten von Leistungen über Klassen hinweg können wiederum sehr unterschiedlichen Zwecken dienen. Sie enthalten ein geringes persönliches Risiko (low stakes), wenn sie nur einem systembezogenen Monitoring dienen. Die politische Führung eines Landes kann davon aber betroffen sein, wenn sich wiederholt zeigt, dass es die Politik nicht schafft, Verbesserungen in der Leistungsfähigkeit des Bildungswesens zu erzielen. Für den einzelnen Schüler und die einzelne Schülerin sind solche Testungen risikolos. Dies ist dann nicht der Fall, wenn Tests zu Noten führen oder Teil von Abschlussprüfungen sind und damit die Zukunft der Schüler/innen in der Gestalt von Berechtigungen betroffen ist. Für Lehrpersonen sind dann hohe Konsequenzen (high stakes) mit Testungen verbunden, wenn sie Hinweise auf die Leistungen ihrer Schulklasse geben, in der sie lange unterrichtet haben und die somit ihre Kompetenz als Lehrende auf den Prüfstand stellen.

Je nach Testungsart (intern oder extern, standardisiert oder lehrerabhängig) und je nach den damit verbundenen Konsequenzen gehen die betroffenen Akteure unterschiedlich mit den Ergebnissen um. Wenn man die Wirksamkeit einer standardorientierten und testbasierten Steuerung untersuchen will, ist es somit ganz entscheidend, diese Intervention präzise zu beschreiben, um Hypothesen zu bilden, wie sie bei unterschiedlicher Ausgestaltung „wirkt“, d. h. in Schüler/innenleistungen umgesetzt wird.

2. Das Mehrebenenkonzept der Handlungsverantwortungen und der auf jeder Ebene erfolgenden Rekontextualisierung durch interessengeleitete, verstehende und problemlösende Akteure ermöglicht nicht nur, sondern fordert zwingend dazu auf, die *Verarbeitungsformen von durch Politik und Verwaltung geschaffenen Regulierungsprozessen*, also von Vorgaben der Makroebene, auf der schulischen und unterrichtlichen Ebenen zu untersuchen (s. z. B. Freudenthaler und Specht 2006; Grabensberger et al. 2008; Grillitsch 2010). Zu wissen, wie Lehrpersonen die Standardsetzungen und Testverfahren rekontextualisieren, ob und wie sie diese in ihre Unterrichtsgestaltung einfließen lassen, gehört zum Kern einer Wirksamkeitsanalyse.

Dass dies von erfolgreichen Forschungsanträgen im Rahmen des eingangs angesprochenen Schwerpunktprogramms „Systemsteuerung“ mehr oder weniger explizit berücksichtigt wird, zeigt folgende Passage über die vermutete Bedeutung einer standardisierten Testungsstrategie: „Die Verantwortung der Schulen und Lehrkräfte für Bildungsqualität wird mit standardbasierten Tests gestärkt; gleichzeitig wird deutlich, dass die Schulen diese Verantwortung im Rahmen institutioneller Regelstrukturen wahrnehmen, die ihrerseits einem Wandel unterliegen. So verlagert sich die Funktion der bislang hauptsächlich formalbürokratisch positionierten und agierenden Schulaufsichten in Richtung auf unterstützende, beratende aber auch kontrollierende Tätigkeiten. Damit ändert sich das Gefüge der institutionellen Regelung von Qualitätsentwicklung im Schulsystem grundlegend – insbesondere auch in Hinblick auf die Koordination zwischen den verantwortlichen Akteuren in der Schule (Lehrkräfte; Schulleitung) und der Schulaufsicht. Die Verlagerung von Verantwortung und Entscheidungskompetenz auf die Schule im Rahmen der Schulautonomie gilt als eine komplementäre Entwicklung zur testbasierten Schulreform.“

3. Im Rahmen der oben beschriebenen Rekontextualisierungstheorie ergibt sich die überraschende Hypothese, dass Regulierungen auf der Makroebene *direkt* auf die Mikroebene durchschlagen können, wenn es sich z. B. um high-stakes Testungen mit klassenübergreifenden Instrumenten handelt. Dann enthalten diese für Schüler/innen Regeln und Bedingungen dafür, wie sie erwünschte Abschlüsse erreichen können. Wenn die Standards extern formuliert und extern getestet werden, können Schüler/innen das Erreichen der Schulabschlussziele nicht dadurch erleichtern, dass sie Druck auf die Lehrperson, nicht zu streng zu sein, oder Druck auf die Mitschüler/innen ausüben, die Maßstäbe der Anstrengung nicht zu hoch zu setzen. Wenn die Leistungsstandards extern gesetzt und geprüft werden, dann verändern sich auch die Interpretationen und Handlungsstrategien von Lehrpersonen. Für sie hat dann eine hohe Erwartungshaltung und ein intensives Bemühen, dem Schüler oder der Schülerin möglichst viel beizubringen, damit er oder sie einen Abschluss schafft, einen anderen Stellenwert, als wenn sie selbst die Standards setzt und deren Erfüllung selber prüfen muss. Im ersten Fall kann sie glaubhaft als Coach im Schüler/inneninteresse handeln, im zweiten wird sie gleichzeitig zum Richter über die Schulkarriere einer Schülerin oder eines Schülers. Es wäre schließlich denkbar, dass selbst bei weniger konsequenzenreichen Standardtestungen die schlichte Formulierung, Überprüfung und Dokumentation von Standards, das *Qualitätsbewusstsein* in bisher nicht dagewesenem Maße vergrößert. Wenn dies durch entsprechende Kompetenzvergleiche noch regional oder lokal differenziert aussagekräftig gemacht würde und auf die Lernvoraussetzungen hin spezifiziert werden könnte, dann könnte das Bewusstsein geschärft werden, was *möglich* ist und anderswo Wirklichkeit geworden ist. Unübersehbar sind dies bislang nur Annahmen.

Aber nicht nur Hypothesen über *direkte* Wirkungspfade von der Makro- auf die Mikroebene können formuliert werden, sondern auch solche darüber, wie Akteure Maßnahmen neutralisieren und unwirksam machen können.⁵ Nicht zuletzt lassen sich im Rahmen des Rekontextualisierungsparadigmas auch *unbeabsichtigte Nebenwirkungen* systematisch erforschen. Lehrpersonen sind, wenn schlechte Testleistungen in ihrer Klasse gravierende Folgen für ihre eigene berufliche Situation haben (bei sogenanntem high-stake testing mit Folgen für Bezahlung, Anstellung und Beförderung), nicht schicksalhaft einer solchen Situation ausgeliefert. Sie haben das Interesse zu bestehen und gute Konsequenzen ihres

Berufshandelns zu erleben, sei dies ein schlichtes Lob, sei dies eine Gehaltserhöhung, sei dies eine Beförderung. Aus dem Zusammentreffen externer Testungen mit beruflichen Interessen von Lehrpersonen und Schulen können vielfältige Nebenwirkungen resultieren (Bellmann und Weiß 2009): Auf Unterrichtsebene besteht die Möglichkeit, dass Lehrpersonen die Unterrichtszeit auf jene Lernbereiche konzentrieren, die getestet werden. Gleichzeitig können sie sich bemühen, die besondere Art von Aufgaben zu üben, die bei standardisierten Tests üblich sind. Aber auch Hilfen während der Testdurchführung und nachträgliche Korrekturen von Testbögen sind denkbar. Neue Methoden mit unsicherem Erfolg werden Lehrpersonen, deren berufliches Schicksal von den Testergebnissen abhängt, seltener wagen einzusetzen.

Aber auch ganze Schulen, deren Ansehen von den Testleistungen der Schüler/innen-schaft beeinflusst ist, werden Strategien entwickeln, um in der Öffentlichkeit ein möglichst gutes Erscheinungsbild zu erzeugen. Die wichtigste ist die, möglichst leistungsstarke Schüler/innen zu rekrutieren und sich von leistungsschwachen zu entlasten. Bei der Testdurchführung sind ebenfalls einige „Tricks“ denkbar, z. B. leistungsschwache Schüler/innen mit mehr oder weniger guten Gründen von den Tests auszuschließen. Wenn Schullevaluationen anstehen, sind ebenfalls viele Strategien einsetzbar, um das Erscheinungsbild („window dressing“) zu optimieren. Schulfeste, Dekorationen, attraktive Gestaltung des Umfeldes für die Evaluator/innen, außergewöhnliche Initiativen usw. können das Image der Schule positiv beeinflussen, ohne dass damit reale und nachhaltige Wirkungen auf die Lernergebnisse erzielt werden.

Als nicht-intendierter Effekt dieser Steuerungsmaßnahmen könnte damit sogar nachhaltig die pädagogische Grundhaltung der Lehrer/innenschaft beschädigt werden. Ihre professionelle Konzentration auf den Unterricht könnte durch die dauernde Beschäftigung mit professionsfremden Aufgaben und die unterrichtsferne Verteilung von Ressourcen leiden. Letztlich wäre es sogar denkbar, dass Überlebensstrategien im Beruf wirksam werden, die sich an Äußerlichkeiten der Darstellung orientieren, von der Wahrnehmung von Mißtrauen der Öffentlichkeit geprägt sind und deshalb zu defensiven Reaktionen auf mögliche Gefahren hin führen. Nach Bellmann und Weiß (2009) gibt es Hinweise, dass im amerikanischen Raum solche Prozesse im Umkreis von high-stakes testing beobachtbar sind. Ob dies auch im deutschen Sprachraum der Fall ist oder sein wird, ist eine zu überprüfende Hypothese. Deren Formulierung und ihre Überprüfung sind von dem übergeordneten theoretischen und methodischen Konzept geprägt, dass Akteure „intelligent“ mit institutionellen Vorgaben umgehen, also von einem Konzept der Rekontextualisierung von Rahmenbedingungen ihres aktuellen beruflichen Handelns.

6 Methodische Schlussfolgerungen

Aus diesem Konzept der Analyse neuer Steuerung ergibt sich konsequenterweise nicht nur die Legitimität, sondern die für anwendungsorientierte Erkenntnisinteressen zwingende Notwendigkeit von *Fallstudien* und von *interpretativen Verfahren*. Natürlich müssen auch diese Methoden Qualitätskriterien entsprechen, etwa im Design klug auf die Fragestellung angepasste kontrastierende Fallanalysen sein und im interpretativen Ver-

fahren kontextualisierte und vergleichende Rekonstruktionen von Mentalitäten und interesselgeleiteten Interpretationen enthalten.

Dabei darf aber auch eine Gefahr nicht übersehen werden, wenn sich Studien auf *Nutzungs- und Akzeptanzfragen* bzw. auf *Transformationsfragen* bildungspolitischer Maßnahmen konzentrieren und dabei lediglich Fallstudien durchgeführt werden. Sie könnten, wenn sie nur in Prozessanalysen verbleiben, den Bezug zu den erwünschten Veränderungen verlieren: den Bezug zur Verbesserung der *Lernergebnisse* und der überfachlichen Wirkungen des Bildungswesens. Um sie aus der Beschränkung auf Sonderkonstellationen, etwa aus bevorzugten Umständen eines Modellversuchs, herauszuführen und generalisierungsfähig zu machen, wären nach einer sensiblen Prozessanalyse wieder repräsentative Fragestellungen und korrespondierende Forschungen nötig. Dies gilt sowohl für eine Quantifizierung der Rekontextualisierung als auch für die Lernergebnisse. Somit sind dann wieder „harte“ Daten und repräsentative Studien zur Lernwirksamkeit und zu den erzieherischen Wirkungen erforderlich, deren Ergebnisse aber handlungstheoretisch zu rekonstruieren wären.

So müssen sich die Traditionen der school effectiveness-Forschung und die der handlungs-, akteur- und institutionsorientierten Prozessanalyse ergänzen, um zusammen am Rationalitätskonzept der evidence-based (oder zumindest evidence-inspired) policy im Bezug auf die Weiterentwicklung des Bildungswesens zu arbeiten. Dieses muss sich letztlich an den positiven Wirkungen bei Lernergebnissen und Persönlichkeitsentwicklungen von Schülerinnen und Schülern messen lassen. *Gestaltungsinstrumente* sind somit auch hier immer wieder an den durch sie *geschaffenen und beobachtbaren Wirklichkeiten* (Fend 2006b, S. 177) zu überprüfen. Dies erfordert zwingend einen Methodenpluralismus. Interpretative Prozessanalysen, die Aufschluss über „Mechanismen“ geben, sind ebenso unerlässlich wie sozialwissenschaftliche Wirkungsanalysen, die den Fokus auf die Ergebnisse legen, in der Interpretation der Ergebnisse sich aber von interpretativen und handlungsorientierten Prozessanalysen leiten lassen.

Anmerkung: Der Artikel fußt auf dem Manuskript eines Vortrags bei der Tagung „Strategien zur Erforschung der Governance im Bildungswesen“, die vom 25.–26. November 2010 an der Johannes Kepler Universität Linz stattfand. Weitere Beiträge zum Thema finden sich im Band „Educational Governance als Forschungsperspektive. Strategien, Methoden und Forschungsansätze in interdisziplinärer Perspektive“, der von Katharina Maag Merki, Roman Langer und Herbert Altrichter im Verlag für Sozialwissenschaften, Wiesbaden herausgegeben wird.

Anmerkungen

- 1 Fragen der Steigerung von Equity und Equality sind davon allerdings nicht berührt. Dazu sollen andere Instrumente verhelfen, etwa die Veränderung der Schulstrukturen, die Verlängerung gemeinsamen Lernens und Vermeidung von Leistungsgruppierungen, die Abschaffung der Klassenwiederholung, die Abschaffung der Querversetzung in Schulformen, die Stärkung der Elternrechte über freie Schulwahl usw.

Um die empirische Untersuchung und Untersuchbarkeit dieser letzten Fragen soll es im Folgenden nicht gehen. Es soll die neue Standardsicherung über die jeweiligen Verfahren der Kompetenzmessung für New Governance als Paradigma herangezogen werden. Sie ist gewissermaßen die große Intervention zur Steigerung der Lernergebnisse in Schulen.

- 2 Dass wissenschaftliche Studien mit diesen Problemen auch produktiv umgehen können, zeigt die Gesamtschulforschung der 1970er Jahre (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 1982; Fend 1982)
- 3 Siehe das Forschungsprogramm im LOEWE-Zentrum des DIPF: Center for Individual Development and Adaptive Education – www.dipf.de.
- 4 Das Paradigma des akteurzentrierten Institutionalismus (Scharpf 2000) ist dafür eine wichtige Leitperspektive.
- 5 Siehe dazu das Projekt „Rekonstruktion von Prozessen der Implementation von Bildungsstandards auf Schulebene“, das von Barbara Asbrand und Mitarbeiterinnen auf der Linzer Tagung „Strategien zur Erforschung von Governance im Bildungswesen“ (25.–26.11.2010) präsentiert wurde.

Literatur

- Altrichter, H., Brüsemeister, T., & Wissinger, J. (2007). *Educational Governance. Handlungskoordination und Steuerung im Bildungswesen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Altrichter, H., & Maag Merki, K. (2010). *Handbuch Neue Steuerung im Schulsystem*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Arbeitsgruppe Bildungsberichterstattung (Hrsg.). (2008). *Bildung in Deutschland 2008*. Bielefeld: W. Bertelsmann.
- Bellmann, J., & Weiß, M. (2009). Risiken und Nebenwirkungen Neuer Steuerung im Schulsystem. *Zeitschrift für Pädagogik*, 55(2), 286–308.
- Benz, A. (Hrsg.). (2007). *Handbuch Governance: Theoretische Grundlagen und empirische Anwendungsfelder*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Berkemeyer, N. (2010). *Die Steuerung des Schulsystems*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bevir, M., & Rhodes, R. A. W. (2010). *The state as cultural practice*. Oxford: Oxford University Press.
- Böttcher, W., Bos, W., Döbert, H., & Holtappels, H. G. (Hrsg.). (2008). *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive*. Münster: Waxmann.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (Hrsg.). (1982). *Modellversuche mit Gesamtschulen. Auswertungsbericht der Projektgruppe Gesamtschule*. Bühl: Konkordia.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Creemers, B., Scheerens, J., & Reynolds, A. J. (2000). Theory development in school effectiveness research. In C. Teddlie & A. J. Reynolds (Hrsg.), *The international handbook of school effectiveness research* (S. 283–298). London: Falmer Press.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London: Routledge.
- Dubs, R. (1996). *Schule, Schulentwicklung und New Public Management*. St. Gallen: Institut für Wirtschaftspädagogik an der Universität St. Gallen.
- Fend, H. (1982). *Gesamtschule im Vergleich*. Weinheim: Beltz.
- Fend, H. (2001). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung* (2. Aufl.). Weinheim: Juventa.
- Fend, H. (2006a). *Geschichte des Bildungswesens. Der Sonderweg im europäischen Kulturraum*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fend, H. (2006b). *Neue Theorie der Schule. Eine Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Freudenthaler, H. H., & Specht, W. (2006). *Bildungsstandards: Der Implementationsprozess aus der Sicht der Praxis*. Graz: Zentrum für Schulentwicklung.
- Grabensberger, E., Freudenthaler, H. H., & Specht, W. (2008). *Bildungsstandards: Testungen und Ergebnisrückmeldungen auf der achten Schulstufe aus der Sicht der Praxis*. Graz: bifie.
- Grillitsch, M. (2010). *Bildungsstandards auf dem Weg in die Praxis*. Graz: bifie.
- Heinrich, M. (2007). *Governance in der Schulentwicklung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität*. Seelze: Kallmeyer/Klett.
- Joas, H. (1996). *Die Kreativität des Handelns*. Frankfurt a. M.: Suhrkamp.
- Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projektes ‚Pythagoras‘. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule* (S. 127–146). Münster: Waxmann.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The pythagoras study – Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 137–160). Münster: Waxmann.
- Kussau, J., & Brüsemeister, T. (2007). *Governance, Schule und Politik*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Langer, R. (Hrsg.). (2008). *Warum tun die das? Governanceanalysen zum Steuerungshandeln in der Schulentwicklung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Luyten, H. (2003). The size of school effects. *School Effectiveness and School Improvement*, 14(1), 13–511.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch der Neuen Steuerung* (S. 145–170). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Oelkers, J., & Reusser, K. (2008). *Expertise: Qualität entwickeln – Standards sichern – mit Differenz umgehen*. Bonn: Bundesminister für Bildung und Forschung.
- Reusser, K. (1998). Denkstrukturen und Wissenserwerb in der Ontogenese. In F. Klix & H. Spada (Hrsg.), *Enzyklopädie der Psychologie* (Bd. 6, S. 115–166). Göttingen: Hogrefe Verlag für Psychologie.
- Rhodes, R. A. W. (2010). *Everyday life at the top*. Oxford: Oxford University Press.
- Rhodes, R. A. W., Wanna, J., & Weller, P. (2009). *Comparing Westminster*. Oxford: Oxford University Press.
- Rumpf, H. (1966). *Die administrative Verstörung der Schule*. Essen: Neue deutsche Verlagsanstalt.
- Scharpf, F. W. (2000). *Interaktionsformen. Akteurzentrierter Institutionalismus in der Politikforschung*. Opladen: Leske + Budrich.
- Schedler, K. (1995). *Ansätze einer wirkungsorientierten Verwaltungsführung*. Bern: Paul Haupt.
- Scheerens, J., & Bosker, T. (1997). *The foundations of educational effectiveness*. New York: Pergamon.
- Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2006). *Review and meta-analysis of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organization and Management.
- Sørensen, A. B., & Morgan, S. L. (2000). School effects. Theoretical and methodological issues. In M. T. Hallinan (Hrsg.), *Handbook of the sociology of education* (S. 137–162). New York: Kluwer.
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.

-
- van der Wolf, J. C. (1988). School effects: The use of planned intervention to study causal processes. In M. Rugger (Hrsg.), *Studies in psychosocial risk: The power of longitudinal data* (S.221–226). Cambridge: Cambridge University Press.
- Weber, M. (1947). *Wirtschaft und Gesellschaft* (3. Aufl.). Tübingen: Mohr Siebeck.
- Wößmann, L. (2002). *How central exams affect educational achievement: International evidence from TIMSS and TIMSS-Repeat*. Paper presented at the Takind Account of Accountability: Assessing Politics and Policy, John F. Kennedy School of Government, Harvard University.