

Sabrina Müller

# **Data Rich, Information Poor?**

**Chancen und Grenzen vergleichender  
Leistungsmessung an der Förderschule mit  
dem Förderschwerpunkt Lernen**

# Empirische Erziehungswissenschaft

herausgegeben von

Rolf Becker, Sigrid Blömeke, Wilfried Bos,  
Hartmut Ditton, Cornelia Gräsel, Eckhard Klieme,  
Rainer Lehmann, Thomas Rauschenbach,  
Hans-Günther Roßbach, Knut Schwippert,  
Ludwig Stecher, Christian Tarnai, Rudolf Tippelt,  
Rainer Watermann, Horst Weishaupt

Band 53



Waxmann 2014  
Münster • New York

Sabrina Müller

# Data Rich, Information Poor?

Chancen und Grenzen vergleichender Leistungsmessung  
an der Förderschule mit dem Förderschwerpunkt Lernen



Waxmann 2014  
Münster • New York

Diese Arbeit wurde von der Fakultät Erziehungswissenschaft und Soziologie der Technischen Universität Dortmund 2012 als Dissertation angenommen.

### **Bibliografische Informationen der Deutschen Nationalbibliothek**

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

### **Empirische Erziehungswissenschaft, Band 53**

ISSN 1862-2127

Print-ISBN 978-3-8309-3166-9

E-Book-ISBN 978-3-8309-8166-4

© Waxmann Verlag GmbH, 2014

[www.waxmann.com](http://www.waxmann.com)

[info@waxmann.com](mailto:info@waxmann.com)

Umschlaggestaltung: Pleßmann Design, Ascheberg

Druck: Systemdruck Köln

Gedruckt auf alterungsbeständigem Papier, säurefrei gemäß ISO 9706

Printed in Germany

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.

Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

## Danksagung

Zum Gelingen dieser Arbeit haben eine ganze Reihe von Menschen durch vielfältige Formen der Unterstützung, des Zuspruchs, durch gute Beratung, gemeinsames Denken, Geduld, Kreativität und fruchtbare Diskussionen beigetragen. Allen voran möchte ich Wilfried Bos und Knut Schwippert für die zahlreichen konstruktiven Beratungsgespräche danken, die immer von großer Wertschätzung geprägt waren. Eine besondere Rolle im Entstehungsprozess der Dissertation kommt außerdem Rainer Peek zu, der mich nicht nur mit seinem Enthusiasmus und seiner Leidenschaft für das wissenschaftliche Arbeiten geprägt und für das Feld der Bildungsforschung begeistert, sondern mich darüber hinaus mit seiner Freude am wissenschaftlichen Denken angesteckt hat. Knut Schwippert danke ich ganz besonders für seine spontane Bereitschaft, nach dem Tod von Rainer Peek die Betreuung meiner Arbeit zu übernehmen.

Ebenso danke ich der Research School ‚Education and Capabilities‘ dafür, dass sie nicht nur die Durchführung meines Projektes finanziell ermöglicht und in jeder Hinsicht unterstützt hat, sondern darüber hinaus mit der Bereitstellung räumlicher, zeitlicher und personeller Ressourcen optimale Rahmenbedingungen bereitgestellt hat, um ein komplexes und spannendes Projekt in einem Zeitraum von nur drei Jahren umsetzen zu können. Mein besonderer Dank richtet sich an die Speaker der Research School, Uta Quasthoff und Hans-Uwe Otto, außerdem an Ruth Springer, Petra Strähle, Sabine Schäfer, Anne Meyer-Hannes und Ursula Siebel-Militzer. Ohne das Engagement so vieler Personen auf allen Ebenen wäre das Promotionsprojekt nicht umsetzbar gewesen! Ein ganz herzlicher Dank gebührt meinen Kolleginnen und Kollegen aus dem IFS und der Research School, insbesondere Tobias C. Stubbe, Magdalena Buddeberg, Claudia Dohe, Daniel Kasper und Rolf Strietholt, die mir bei der Datenaufbereitung und Datenauswertung mit Rat und Tat zur Seite standen, sowie Marjan Bijani, Verena Feldhaus, Martin Goy, Vivien Heller, Jolanta Kavale, Lola Ping, Claudia Müller, Giang Pham und Anke Walzebug für zahlreiche interessante Gespräche und ermutigende Worte. Mario Venemann, Verena Harpering, Janine Hannemann, Sarah Rempe, Julia Mink und Jennifer Pfannenstern danke ich für die zuverlässige Unterstützung bei der Dateneingabe.

Des Weiteren danke ich Elisabeth Moser Opitz, Okka Fresemann, Reiner Lehmann, Ellen Brodesser, Michaela Greisbach, Ansgar Schwarz und Hans-Jürgen Kuhn für die Bereitstellung von (zum Teil) unveröffentlichtem Material ihrer eigenen Forschung.

Den Hauptakteuren meiner Forschung, den Lehrerinnen und Lehrern der PARS-Förderschulen und ihren Schulleiterinnen und Schulleitern, gilt an dieser Stelle mein ganz besonderer Dank. Ohne das Interesse und Vertrauen, das sie mir entgegengebracht haben, und ihre Zeit, meine vielen Fragen in Form von Fragebögen, Interviews und Fokusgruppengesprächen zu beantworten, wäre mein Projekt niemals möglich gewesen. Sie haben mir spannende Einblicke in ihre pädagogische Praxis, ihre Wünsche und Bedürfnisse gegeben, die ich hoffe, mit meiner Arbeit stärker in die wissenschaftliche und bildungspolitische Diskussion einbringen zu können.

Last but not least bedanke ich mich für alle Arten mentaler Unterstützung, ausdauernden Korrekturlesens und geduldigen Zuhörens ganz besonders bei Charlotte Bünte, Hannah Dannenberg, Michael Grosche, Roswitha Hoffmann, Max Hornig, Marion Hornig, Norbert Müller, Claudia Müller, Nico Müller und Jürgen Wilbert. Ich kann nur hoffen, dass ich ausreichend Gelegenheit haben werde, die Unterstützung, die ich von euch und Ihnen erfahren habe, an anderer Stelle weitergeben zu können.

## Zusammenfassung

Die vorliegende empirische Arbeit beschäftigt sich mit der Frage, inwiefern extern administrierte, vergleichende Leistungsmessung für Förderschulen – genauer gesagt für die Förderschule mit dem Förderschwerpunkt Lernen – ein Instrument zur datengestützten Schul- und Unterrichtsentwicklung sein kann. Diese Frage gewinnt vor dem Hintergrund der bildungspolitischen Veränderungen seit dem PISA-Schock im Jahr 2001 insbesondere deswegen an Relevanz, weil die Förderschule Lernen durch ihre speziellen organisatorischen, curricularen und personellen Merkmale in besonderer Weise von einem Instrument wie Vergleichsarbeiten profitieren müsste. Der Nutzen, den Schulen aus den Ergebnissen vergleichender Leistungsmessung ziehen, ist jedoch höchst unterschiedlich und hängt von einer Reihe von Faktoren auf unterschiedlichen Ebenen ab. Neben Merkmalen einzelner Lehrpersonen und schulorganisatorischen Faktoren (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) spielen Merkmale der Rückmeldeberichte, die sich als Medium zur Übermittlung solcher Informationen etabliert haben, eine entscheidende Rolle (Schneewind, 2007a, 2007b).

Primäres Ziel der Rückmeldung von Ergebnissen aus vergleichender Leistungsmessung ist die Unterstützung der Lehrerinnen und Lehrer in Prozessen der Schul- und Unterrichtsentwicklung durch die Bereitstellung empirischer Daten zur Überprüfung und Reflexion eigener mentaler Modelle und Handlungen (vgl. Rolff, 2002). Durch die Rückmeldung von Schülerleistungen in vergleichender Perspektive sollen in den Schulen Anlässe geschaffen werden, vorhandenes Handlungswissen und unterrichtsbezogene Routinen kritisch zu hinterfragen und gegebenenfalls zu verändern. Eine ausgeprägte Orientierung an den Bedürfnissen, Wünschen und Kompetenzen der Adressaten von Ergebnismeldungen ist vermutlich eine wichtige Grundlage für die Nutzbarkeit der zurückgemeldeten Informationen und ihre pädagogische Anschlussfähigkeit.

Mit dieser explorativen Untersuchung, die in einem sequentiellen Design qualitative und quantitative Methoden miteinander kombiniert, liegen die ersten empirischen Befunde zum Einsatz von Vergleichsarbeiten an der Förderschule mit dem Förderschwerpunkt Lernen und der adressatengerechte Gestaltung von Klassenrückmeldungen vor. Es muss jedoch darauf hingewiesen werden, dass die Anzahl der Teilnehmerinnen und Teilnehmer der Online-Befragungen mit  $N = 30$  im ersten und  $N = 13$  im zweiten Untersuchungsjahr sehr klein waren. Bedingt durch die kleine Stichprobengröße und den Umstand, dass die Teilnahme an der Leistungsstudie *PARS-F* und ebenso die Teilnahme an den Online-Befragungen im Rahmen der Dissertationsstudie auf freiwilliger Basis stattfanden, sind der Generalisierbarkeit der Ergebnisse enge Grenzen gesetzt.

Für die Bildungsadministration sind die Befunde der vorliegenden Untersuchung insofern relevant, als dass sie zeigen, dass die vergleichende Leistungsmessung auch im Bereich der sonderpädagogischen Förderung ein enormes Potenzial hat und bei den Lehrkräften grundsätzlich auf eine breite Akzeptanz stößt. Darüber hinaus weisen die Ergebnisse darauf hin, dass die Relevanz der zurückgemeldeten Informationen für die Adressaten einen entscheidenden Einfluss auf die Intensität von Rezeption und Reflexion der Ergebnisse hat.

## Abstract

This empirical study deals with the question of how externally administered, comparative performance measurement can be a tool for data-driven school and teaching development for special schools – specifically for the German special needs school with a program focusing on learning development (Förderschule Lernen, FS L). In the context of educational policy changes beginning with TIMSS and PISA this question gained some relevance, since institutions offering special education should benefit from a tool such as comparative studies in a particular way because of specific organizational, curricular and personal features. However, the benefits, school draw from the results of comparative performance measurement are highly variable and depend on a number of factors at different levels. Besides characteristics of individual teachers and patterns of school organization (Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) the design of feedback reports, which have been implemented as a medium for the transmission of such information, play a crucial role (Schneewind, 2007a, 2007b).

The primary objective of feedback from comparative performance measurement is to support the teachers with regard to processes of education and instructional development by providing empirical data to review and reflect on their own mental models and actions (cf. Rolff, 2002). Through the feedback of students' performance in comparative perspective occasions should be created to challenge existing practical knowledge and teaching related routines critically and change if necessary. A strong focus on the needs, desires and skills of the addressees of performance feedback is probably an important basis for the usefulness of the feedback information and their educational connectivity.

This exploratory study that combines qualitative and quantitative methods in a sequential design, gives first empirical evidence on the use of comparative data-feedback and on factors for an applicable addressee-oriented design of feedback reports for German special schools with a program focusing on learning. It should be noted that the number of participants in the online survey with  $N = 30$  and  $N = 13$  in the first and in the second year of the study were quite small. Due to this small sample size and the fact that participation in the performance study PARS-F and both online surveys took place on a voluntary basis, the generalizability of the results are severely limited.

For the educational administration board, the findings of this study are relevant insofar as they show that the comparative performance measurement in the field of special education has enormous potential and meets a wide acceptance among special needs teachers. Moreover, the results suggest that the relevance of the feedback information for the addressee has a decisive influence on the intensity of reception and reflection of the results.

# Inhalt

1	Einleitung .....	13
1.1	Data Rich, Information Poor?.....	13
1.2	Ausgangsüberlegungen für die vorliegende Arbeit.....	14
1.3	Aufbau der Arbeit.....	15
2	Neue Steuerung – eine neue Perspektive für die Sonderpädagogik? .....	18
2.1	Die Rolle vergleichender Leistungsmessung in bildungspolitischen Steuerungsprozessen.....	18
2.1.1	Die Logik des neuen Steuerungsmodells.....	18
2.1.2	Neue Steuerung – neue Instrumente.....	21
2.1.3	Systematisierung von Schulentwicklung.....	30
2.1.4	Rezeption der Steuerungsdebatte in der Sonderpädagogik .....	37
2.2	Die Förderschule Lernen im Spiegel vergleichender Leistungsmessung .....	42
2.2.1	Sonderpädagogische Förderung in Deutschland .....	43
2.2.2	Schulleistungsstudien an der Förderschule Lernen .....	45
2.2.3	Charakteristika der Förderschule Lernen .....	47
2.2.4	Internationale Erfahrungen mit der Beteiligung von Förderschülern an vergleichender Leistungsmessung.....	52
2.3	Zwischenfazit: Potenziale vergleichender Leistungsmessung für die sonderpädagogische Förderung im Förderschwerpunkt Lernen .....	55
3	Rückmeldung aus vergleichender Leistungsmessung .....	60
3.1	Die Rolle von Feedback in der schulpolitischen Steuerung.....	60
3.1.1	Der erweiterte Feedbackkreislauf: eine besondere Herausforderung .....	62
3.1.2	Elemente von Schulrückmeldungen .....	64
3.1.3	Theoretische Überlegungen zur Nutzung von Datenfeedback in Schulen .....	65
3.2	Ergebnisse der Rezeptionsforschung.....	75
3.2.1	Standardisierte Rückmeldungen im Urteil von Lehrkräften .....	77
3.2.2	Rezeption von Ergebnissen aus Vergleichsarbeiten.....	79
3.2.3	Nutzung von Evaluationsergebnissen.....	81
3.2.4	Inanspruchnahme von Unterstützung .....	86
3.2.5	Einflussfaktoren der Rezeption und Nutzung von Rückmeldungen .....	87
3.3	Effekte von ErgebnISRückmeldungen .....	90
3.4	Ansprüche an ErgebnISRückmeldungen .....	91
3.4.1	Ansprüche an die Gestaltung von Rückmeldungen.....	91
3.4.2	Ansprüche an die Gestaltung von Rückmeldeprozessen.....	94
3.5	Die Förderschule Lernen als Adressatin von Rückmeldungen aus Leistungsvergleichsuntersuchungen.....	95
4	Fragestellung.....	97

4.1	Welchen Anforderungen muss eine Rückmeldung für die Förderschule Lernen gerecht werden? .....	98
4.2	Nutzung der Rückmeldungen durch die Lehrkräfte .....	100
4.3	Die Forschungsfragen auf einen Blick .....	103
5	Daten und Methoden .....	104
5.1	Die Panel Study at the Research School ‘Education and Capabilities’ in North Rhine Westphalia (PARS).....	104
5.1.1	Konzeption der Studie PARS an Förderschulen (PARS-F) .....	105
5.1.2	Ablauf der Datenerhebung in PARS-F .....	108
5.1.3	Leistungstests in PARS-F .....	109
5.1.4	Schülerbefragung in PARS-F .....	120
5.2	Anlage der Dissertationsstudie .....	122
5.2.1	Entwicklung des Studiendesigns .....	122
5.2.2	Theoretische Herleitung des Rückmeldedesigns im ersten Untersuchungsjahr .....	124
5.2.3	Integration qualitativer und quantitativer Methoden .....	129
5.2.4	Stichproben .....	131
5.3	Entwicklung der Rückmeldeberichte für das erste Untersuchungsjahr .....	136
5.3.1	Gestaltung des ersten Rückmeldeberichts im ersten Untersuchungsjahr .....	137
5.3.2	Gestaltung des zweiten Rückmeldeberichts im ersten Untersuchungsjahr .....	142
5.3.3	Gestaltung des dritten Rückmeldeberichts im ersten Untersuchungsjahr .....	145
5.4	Fokusgruppengespräche .....	148
5.4.1	Beschreibung der Methode .....	148
5.4.2	Diskussionsleitfaden .....	149
5.4.3	Datenaufbereitung und Datenauswertung .....	152
5.5	Online-Befragung am Ende des ersten Untersuchungsjahres .....	154
5.5.1	Organisation der Datenerhebung .....	154
5.5.2	Beschreibung der Variablen .....	155
5.6	Entwicklung der Rückmeldeberichte für das zweite Untersuchungsjahr... ..	164
5.6.1	Gestaltung des ersten Rückmeldeberichts im zweiten Untersuchungsjahr .....	164
5.6.2	Gestaltung des zweiten Rückmeldeberichts im zweiten Untersuchungsjahr .....	167
5.6.3	Gestaltung des dritten Rückmeldeberichts im zweiten Untersuchungsjahr .....	168
5.6.4	Kritische Überprüfung der Qualität der Rückmeldeberichte.....	169
5.7	Die Interviews.....	171
5.7.1	Methode .....	171
5.7.2	Interviewleitfaden .....	172
5.7.3	Auswahl der Interviewpartner .....	175
5.7.4	Auswertung der Interviews.....	180

5.8	Online-Befragung am Ende des zweiten Untersuchungsjahres .....	180
5.8.1	Organisation der Datenerhebung .....	181
5.8.2	Beschreibung der Variablen .....	181
5.9	Auswertungsstrategien für quantitative Daten .....	184
6	Ergebnisse .....	185
6.1	Relevante Einflussfaktoren .....	185
6.1.1	Erwartungen der Lehrkräfte an die Rückmeldeberichte .....	186
6.1.2	Aufgeschlossenheit gegenüber externer vergleichender Leistungsmessung .....	187
6.1.3	Einschätzung der Validität der Testinstrumente durch die Lehrkräfte .....	191
6.1.4	Einfluss des schul- und klassenspezifischen Kontextes .....	193
6.1.5	Zusammenfassung und Einordnung der Befunde zu den relevanten Einflussfaktoren .....	193
6.2	Forschungsfrage I: Welchen Anforderungen muss eine Rückmeldung für die Förderschule Lernen gerecht werden? .....	195
6.2.1	Geeignete Bezugsnormen für Datenrückmeldungen an die Förderschule Lernen .....	196
6.2.2	Methodische Aufbereitung der Ergebnisse .....	201
6.2.3	Relevante Hintergrundvariablen .....	202
6.2.4	Bewertung der Rückmeldeberichte durch die Lehrpersonen .....	205
6.2.5	Zusammenfassung und Einordnung der Befunde zur ersten Forschungsfrage .....	217
6.3	Forschungsfrage II: Nutzung der Rückmeldeberichte durch die Lehrpersonen .....	221
6.3.1	Intensität von Rezeption und Reflexion .....	222
6.3.2	Verwendungsprozesse und die Auswirkungen auf den Unterricht .....	232
6.3.3	Generelle Hemmnisse und schulformspezifische Einschränkungen .....	237
6.3.4	Chancen und Grenzen .....	240
6.3.5	Zusammenfassung und Einordnung der Befunde zur zweiten Forschungsfrage .....	248
7	Diskussion .....	252
7.1	Zusammenfassende Darstellung der Ergebnisse .....	252
7.2	Kritische Reflexion der eingesetzten Methoden .....	254
7.3	Implikationen für zukünftige Rezeptionsstudien .....	257
7.4	Implikationen der Studie für die Bildungsadministration .....	258
8	Verzeichnisse .....	261
8.1	Literaturverzeichnis .....	261
8.2	Abbildungsverzeichnis .....	280
8.3	Tabellenverzeichnis .....	283
8.4	Abkürzungsverzeichnis .....	286

*Wissenschaft wird immer eine Suche sein,  
niemals wirklich eine Entdeckung.*

*Es ist eine Reise,  
niemals wirklich eine Ankunft.*

Karl R. Popper  
(1902–1994)

# 1 Einleitung

## 1.1 Data Rich, Information Poor?

Der Anspruch an Lehrerinnen und Lehrer, im Rahmen ihrer beruflichen Tätigkeit mit extern administrierten Daten zu arbeiten, ist ein Novum im deutschen Schulsystem, das durch neue Formen der bildungspolitischen Steuerung Einzug in die Schulen gehalten hat. Als Ausgangspunkt dieser Entwicklung spielt die Beteiligung Deutschlands an TIMSS (*Third International Mathematics and Science Study*<sup>1</sup>) und PISA (*Programme for International Student Assessment*) in den Jahren 1995 und 2000 eine zentrale Rolle und der sich daran anschließende, den unerwartet schlechten Ergebnissen der 15-jährigen geschuldete, PISA-Schock. Ihm folgten weitreichende Änderungen auf der Ebene der bildungspolitischen Steuerung. Ob die Teilnahme an TIMSS und PISA und deren unerfreuliche Ergebnisse tatsächlich der Stein des Anstoßes für die empirische Wende waren (vgl. Altrichter, 2010; Bos & Schwippert, 2002; Goy, van Ackeren & Schwippert, 2008; Peek, 2006a; Voss & Blatt, 2010; für eine ausführliche Besprechung der Folgen von PISA vgl. Payk, 2009) oder – wie Oelkers und Reusser (2008b) betonen – weniger Ausgangspunkt, sondern „öffentlichkeitswirksamer[n] Höhepunkt bereits anhaltender Reformdiskussionen“ (S. 433), muss offen bleiben. Klar ist jedoch, dass TIMSS und PISA den angestoßenen Reformdiskussionen in Deutschland enormen Auftrieb verschafften und das „Vertrauen von Politik und Öffentlichkeit in ihre Bildungssysteme“ maßgeblich erschütterten (Altrichter, 2010, S. 171; vgl. auch Altrichter & Heinrich, 2007). Seither werden Reformdiskussionen zunehmend als „Steuerungsdiskurs“ geführt (Altrichter & Maag Merki, 2010b, S. 17). Die Notwendigkeit von Standardsetzung und Standardüberprüfung durch Bildungsstandards und vergleichende Leistungsmessung ist dabei zu einem zentralen Element der neuen Steuerungsphilosophie geworden, die auf systematische Qualitätssicherung und Qualitätsentwicklung an Schulen setzt (Ditton, 2000, 2007; Oelkers & Reusser, 2008b) und sich stark an den Prinzipien des New Public Management orientiert. In diesem Zusammenhang haben eine Vielzahl an Maßnahmen zur Leistungsüberprüfung Einzug in deutsche Schulen gehalten. Neben einer Beteiligung Deutschlands an den bereits genannten großen internationalen Schulleistungstudien, einer gemeinsamen Bildungsberichterstattung der Länder und der zentralen Überprüfung der Erreichung der Bildungsstandards im Ländervergleich, finden in allen Bundesländern zu mindestens zwei Zeitpunkten der Schullaufbahn aller Schülerinnen und Schüler flächendeckende Lernstandserhebungen statt (vgl. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2006).

---

1 Seit dem vierten Zyklus im Jahr 2003 steht das Akronym TIMSS für *Trends in International Mathematics and Science Study*.

Auf der Grundlage der gewonnenen (Leistungs-)Daten sind die Einzelschulen angehalten, den eigenen Unterricht zu reflektieren und gezielt weiterzuentwickeln (Helmke & Hosenfeld, 2005).

Während in den ersten Jahren nach der empirischen Wende Fragen der Datenerhebung und der Erreichung einer breiten Akzeptanz der neuen Maßnahmen in den Schulen eine hohe Aufmerksamkeit erhielten, gewinnt die Frage der Kommunikation und der Nutzung der erhobenen Daten seit einiger Zeit zunehmend an Bedeutung. Aus der Vielfalt der erwähnten Formen von vergleichender Leistungsmessung liegen inzwischen umfangreiche Daten vor, die jedoch – wie Rezeptionsstudien zeigen – von Schulen und von der Bildungsadministration nur unzureichend genutzt werden (vgl. z. B. Bensen & von der Gathen, 2004; Dederling, 2010; Schneewind & Kuper, 2009; van Ackeren & Klein, 2009). Helmke (2004) betont:

Auch bei hoher Bereitschaft zur Auseinandersetzung mit den zurückgemeldeten Ergebnissen bleibt das Ausmaß der tatsächlich eingeleiteten Verbesserungsmaßnahmen aber relativ bescheiden. (Helmke, 2004, S. 102)

Problematisch ist unter anderem, dass den Personen, die die Daten nutzen sollen, allzu häufig das notwendige sozialwissenschaftliche und methodische Hintergrundwissen fehlt (vgl. z. B. Müller, 2009; Peek, 2006a; Schwippert, 2004). Bensen, Büchter und Peek (2006) betonen jedoch nachdrücklich:

Die Einbettung externer Evaluationen in lokale Schulentwicklungsprozesse wird mittel- und langfristig vermutlich nur dann gelingen, wenn Lehrkräfte kompetent mit den Ergebnissen von Leistungsstudien und zentralen Vergleichsarbeiten umgehen können, d. h. wenn sie sowohl ihre Potenziale nutzen als auch ihre Grenzen erkennen lernen. (Bensen et al., 2006, S. 131)

Die Rückmeldungen aus Leistungsvergleichsstudien werden sich dementsprechend daran messen lassen müssen, inwiefern sie Nutzungsprozesse in Schulen behindern oder unterstützen, d. h. inwiefern die durch die Bildungsforschung gewonnen Daten in pädagogisch anschlussfähige Informationen transformiert werden können.

## **1.2 Ausgangsüberlegungen für die vorliegende Arbeit**

Förderschulen mit dem Förderschwerpunkt Lernen<sup>2</sup> wurden bislang nicht systematisch in Maßnahmen vergleichender Leistungsmessung und Überlegungen zur nachhaltigen Qualitätsentwicklung in Schule und Unterricht einbezogen, obwohl gerade diese Schulform die schwierige Aufgabe hat, Schülerinnen und Schüler zu beschulen, die in der

---

2 Im Folgenden wird synonym die Bezeichnung ‚Förderschule Lernen‘ genutzt.

Regelschule bereits gescheitert sind. Ausgangspunkt der empirischen Arbeit ist daher die Frage, inwieweit Konzepte, Strategien und Verfahren einer dateninduzierten Schul- und Unterrichtsentwicklung, wie sie seit einigen Jahren im allgemeinbildenden Schulwesen Praxis sind (vor allem: Lernstandserhebungen), an Förderschulen mit dem Förderschwerpunkt Lernen zur Qualitätssicherung und -entwicklung auf Schul- und Unterrichtsebene beitragen können. Inwieweit sind *benchmarks* und Standards aus vergleichend angelegter externer Evaluation dazu geeignet, die in besonderer Weise von „Capability Deprivation“ (Otto, 2009, S. 106) betroffenen Schulen mit ihren ganz spezifischen personellen, organisatorischen und curricularen Voraussetzungen mit dem Ziel einer besseren Handlungsbefähigung Orientierungen für schulische und unterrichtliche Qualitätsentwicklung zu geben?

Auf der Grundlage von Leistungsdaten aus der regionalen Schulleistungsstudie *PARS an Förderschulen* wurden in einem intensiven Austausch mit den beteiligten Lehrkräften Klassenrückmeldungen für die Förderschule Lernen entwickelt und evaluiert. Die Untersuchung bietet durch ihre sequentielle Anlage die Möglichkeit, Förderschulen mit dem Förderschwerpunkt Lernen aktiv in den Dialog über Standardsetzung und dateninduzierte Qualitätsentwicklung einzubinden und gemeinsam mit den Schulen im Forschungsprozess Strategien zu entwickeln, die dazu beitragen sollen, dass reichhaltige Daten aus extern administrierter vergleichender Leistungsmessung nicht als unzureichende Informationen ihren Weg in die Praxis finden.

### 1.3 Aufbau der Arbeit

Mit dem vorgenommenen Perspektivwechsel, der einen veränderten Blick auf Schule und Unterricht ermöglicht, sind auch neue Vorstellungen darüber entstanden, auf welche Weise und durch welche Maßnahmen die Bildungsadministration Einfluss auf Schulen nehmen und steuernd in deren Entwicklung eingreifen kann. Die Logik und die Instrumente dieser neuen Steuerungskonzepte werden im ersten Abschnitt von Kapitel 2 beschrieben. Anschließend werden in Kapitel 2.1.3 die theoretischen Grundlagen der Schulentwicklungsforschung dargestellt und unter Rückgriff auf handlungstheoretische Überlegungen in Beziehung zu den Maßnahmen output-orientierter Steuerungskonzepte gesetzt. Die dargestellte Diskussion ist bislang primär eine Diskussion, die sonderpädagogische Förderorte im Allgemeinen und die Förderschule Lernen im Speziellen kaum einbezieht. Im Rahmen der vorliegenden Arbeit wird daher die Rezeption der Steuerungsdebatte in der Sonderpädagogik aufgearbeitet und zusammenfassend in Kapitel 2.1.4 dargestellt. In diesem Abschnitt kann gezeigt werden, dass die Frage nach der Notwendigkeit vergleichender Schulleistungsmessung in der sonderpädagogischen Förderung äußerst kontrovers diskutiert wird.

Im Anschluss an eine Übersicht über die Strukturierung sonderpädagogischer Förderung in Deutschland (Kap. 2.2.1) und bereits durchgeführte Schulleistungsstudien an

Förderschulen (Kap. 2.2.2) wird in Kapitel 2.2.3 mit Verweis auf die vorliegende Literatur gezeigt, dass der Erwerb fachlicher Kompetenzen auch an der Förderschule Lernen ein relevantes Bildungsziel ist und Schulen dieser Schulform ebenso wie Regelschulen ausdrücklich leistungsorientiert arbeiten (sollten). Unter dieser Prämisse braucht man laut Weinert (2002c, S. 26) „über die Notwendigkeit ihrer methodisch soliden und pädagogisch begründeten Erfassung [der Schülerleistung] nicht mehr zu streiten“. Die spezifischen personellen, organisatorischen und curricularen Voraussetzungen, unter denen die Förderschule Lernen arbeitet, müssen jedoch auf angemessene Weise berücksichtigt werden. Unter Berücksichtigung dieser Charakteristika der Förderschule Lernen werden nach einem kurzen Exkurs über die internationale Erfahrung mit der Beteiligung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf an vergleichender Leistungsmessung Potenziale outputorientierter Steuerung für die Förderschule Lernen im Rahmen eines Zwischenfazit herausgearbeitet (Kap. 2.3). In diesem Kontext wird bereits auf die Bedeutung des Feedbacks von Leistungsergebnissen hingewiesen, das in der Regel in Form von Rückmeldeberichten realisiert wird.

Wie solche Rückmeldeberichte gestaltet werden, welche Wirkungen sie haben und welcher Voraussetzungen sie bedürfen, wird in Kapitel 3 ausführlich thematisiert. Während in Abschnitt 3.1 theoretische Überlegungen zur Rolle von Rückmeldungen in der schulpolitischen Steuerung angestellt werden, spiegeln die folgenden Abschnitte von Kapitel 3 den aktuellen Forschungsstand der Rezeptionsforschung wider. Diese junge Forschungstradition untersucht die Wirkungen von Feedback aus extern administrierten Datenerhebungen auf schulische und unterrichtliche Prozesse. Dabei wurden bislang ausschließlich Lehrkräfte und Schulleiterinnen und Schulleiter von Regelschulen als Adressaten von Feedback in den Blick genommen. Ergebnisse für die Rezeption von Leistungsdaten an Förderschulen liegen bislang nicht vor. Daher werden im Rahmen der Untersuchung die Ergebnisse der Rezeptionsforschung an Regelschulen ausführlich dargestellt (Kap. 3.2). Kapitel 3.3 beleuchtet die Effekte auf Prozesse der Schul- und Unterrichtsentwicklung, die in der Vergangenheit in Folge von Rückmeldungen aus vergleichender Leistungsmessung erzielt wurden, während in Kapitel 3.4 Ansprüche an Ergebnisrückmeldungen dargestellt werden, wobei zwischen der Gestaltung von Rückmeldungen und Rückmeldeprozessen unterschieden werden muss. Zuletzt wird der Bogen zurück zur Förderschule Lernen gespannt (Kap. 3.5), indem aus dem dargestellten Forschungsstand – die Spezifika der Förderschule Lernen berücksichtigend – Schlussfolgerungen für Rückmeldeprozesse an der Förderschule Lernen gezogen werden.

Nachdem in Kapitel 4 die für die Untersuchung grundlegenden Forschungsfragen formuliert werden, dient Kapitel 5 der Darstellung des Untersuchungsdesigns und der verwendeten Methoden. In diesem Abschnitt wird die Studie *PARS an Förderschulen* vorgestellt, sowie die Anlage der Dissertationsstudie, die Entwicklung der Klassenrückmeldungen und die verwendeten Methoden zur Erhebung der Daten.

Im Ergebnisteil der Arbeit (Kap. 6) werden die im vierten Abschnitt formulierten Forschungsfragen der Reihe nach beantwortet. Eine Zusammenfassung und Einordnung

der Ergebnisse wird in den Abschnitten 6.2.5 und 6.3.5 vorgenommen, bevor diese in Kapitel 7 diskutiert und aus verschiedenen Perspektiven beleuchtet werden.

## 2 Neue Steuerung – eine neue Perspektive für die Sonderpädagogik?

### 2.1 Die Rolle vergleichender Leistungsmessung in bildungspolitischen Steuerungsprozessen

Länder wie Kanada, England, die Niederlande, Schweden oder Finnland, in denen das Bildungssystem stärker durch die Vorgabe von Standards und klaren Vorstellungen über den Output und die detaillierte Beschreibung schulischer Inhalte gesteuert wird, schießen bei PISA im Vorteil zu sein (Klieme et al., 2003). Infolgedessen wurde die Umstellung der Steuerung des deutschen Bildungssystems proklamiert und in allen Bundesländern zielstrebig implementiert, weg von einer ausschließlich input-orientierten Steuerung, hin zu mehr Output-Orientierung. Seitdem ist die Bildungsforschung darum bemüht, der Bildungspolitik sowie den Akteuren vor Ort in den Schulen Daten zur Verfügung zu stellen, die systematisch den Ertrag von Schule und Unterricht in den Blick nehmen – vornehmlich operationalisiert durch fachliche Leistungen von Schülerinnen und Schülern. Damit verbunden ist die Erwartung, „Hinweise auf Bedarf und Ansatzpunkte für die Optimierung der Systeme zu erhalten“ (Stanat, 2008, S. 11). Welcher Logik die Maßnahmen der neuen evidenzbasierten Steuerung von Bildungspolitik und Schulentwicklung gehorchen und wie sich dies auf Lehr- und Lernprozesse in Schule und Unterricht auswirkt, soll im Folgenden beschrieben werden<sup>3</sup>.

#### 2.1.1 Die Logik des neuen Steuerungsmodells

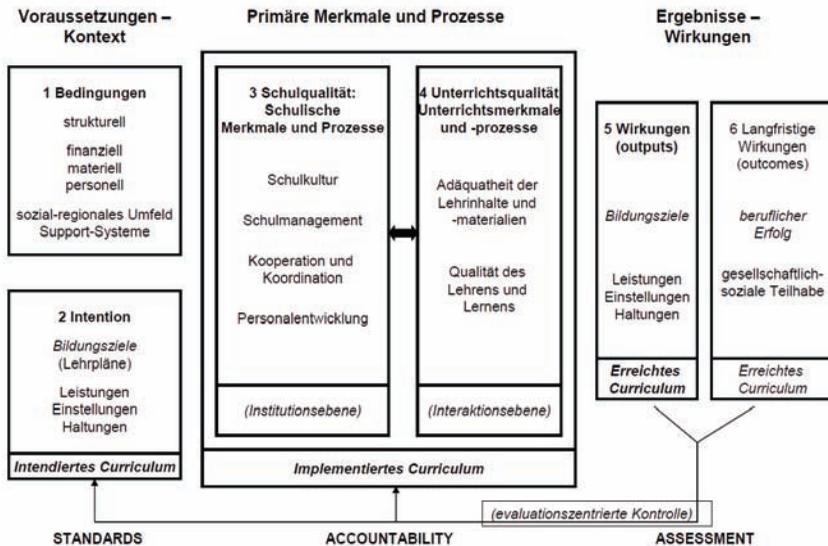
Zur Erklärung von Schülerleistungen sowie zur Beschreibung von Bildungssystemen hat sich in den vergangenen Jahrzehnten das ursprünglich aus der Ökonomie stammende *Input-Prozess-Output* Modell etabliert (Fend, 2008; Reynolds & Teddlie, 2000). Dieses Modell beschreibt Entwicklungen auf der Ebene eines gegebenen Inputs, der – beeinflusst durch Prozesse und den Gesamtkontext – ein Ergebnis (*output*) hervorbringt. Übertragen auf Schule und Unterricht entspricht der Output den (über)fachlichen Leistungen, Einstellungen und Haltungen der Schülerinnen und Schüler sowie den langfristigen Wirkungen schulischer Bildung (beruflicher Erfolg, gesellschaftliche Teilhabe),

---

3 Weitere zentrale Aspekte der neuen Steuerung des Bildungssystems wie Altrichter und Maag Merki (2010b) sie beschreiben (z. B. Aspekte der Erhöhung einzelschulischer Gestaltungsspielräume und Tendenzen der Verbetrieblichung der Einzelschule) können im Rahmen dieser Arbeit nur am Rande aufgegriffen werden. Interessierte Leserinnen und Leser seien auf das Handbuch von Altrichter und Maag Merki (2010a) verwiesen.

die einerseits von spezifischen individuellen Kontextbedingungen von Schule und Schülerschaft (*input*), andererseits von schulischen und unterrichtlichen Gestaltungsprozessen geprägt werden (Ditton, 2000, 2007; vgl. Abbildung 1).

Abbildung 1: Modell zur Qualitätsentwicklung in Schule und Unterricht



Quelle: Ditton, 2007, Abb. 1, modifiziert

Schülerleistungen werden in diesem Modell nicht mehr nur als individuelle Fertigkeiten gedeutet, sondern auch als Resultat des Unterrichts, der in der Verantwortung der Lehrkräfte liegt (vgl. Isaac, Halt, Hosenfeld, Helmke & Groß Ophoff, 2006). Der Einfluss dieser unterschiedlichen Determinanten schulischer Leistungen wurde vielfach empirisch bestätigt (vgl. z. B. DESI-Konsortium, 2008; PISA-Konsortium Deutschland, 2001, 2005; Weinert, 2002a).

Während input- und prozessorientierte Steuerungsmodelle – wie sie bislang in Deutschland vorherrschend waren (Döbert, Klieme & Sroka 2004) – Inputs und Prozesse als Bedingungen und Outputs als daraus folgende Konsequenzen adressieren, zeichnet sich output-orientierte Steuerung dadurch aus, „dass Outputs als Zwecke, Inputs und Prozesse hingegen als Mittel zur Erreichung dieser Zwecke adressiert werden“ (Diemer & Kuper, 2010, S. 260). Output-orientierte Steuerung unterscheidet sich von input- oder prozessorientierter Steuerung infolgedessen durch „die Art und Weise, wie Input-, Prozess- und Outputdimension jeweils ins Verhältnis zueinander gesetzt werden“ (Diemer & Kuper, 2010, S. 260). Im Kontext der neu implementierten Steuerungsstrategie, die vom Output ausgehend Input und Prozesse koordinieren will, trägt die

Schule durch ihr Wissen um Bildungsziele die Verantwortung für die Gestaltung der Prozesse. Durch die kontinuierliche Kontrolle des gewünschten Outputs hat sie die Möglichkeit, Wirkungen auf ihre Schülerinnen und Schüler gezielt und systematisch zu beeinflussen. Merkmale und Prozesse von Schule und Unterricht werden in diesem Kontext zuweilen auch als „Stellschrauben“ (Altrichter & Maag Merki, 2010b, S. 17) bezeichnet, über die die Einzelschule den Weg von vorgegebenen Standards zu überprüfbaren Leistungen beeinflussen kann. Oelkers und Reusser (2008a) fassen diese neue Sicht auf Steuerung von schulischen Bildungsprozessen folgendermaßen zusammen:

Bildungssysteme müssen wohl neu stärker evidenzbasiert vom Ertrag her betrachtet werden, aber auch vom Angebot und von den Prozessen her, wobei zwischen Angebotsqualität, Prozessqualität und Ergebnisqualität kein linearer Zusammenhang besteht. (. . .) Das Ergebnis des Unterrichts ist demnach abhängig von der Art und Weise sowie den gegebenen Möglichkeiten der Nutzung. Beide sind in mehreren Hinsichten variabel. (Oelkers & Reusser, 2008a, S. 2f)

Der Standardsetzung und der datenbasierten Standardüberprüfung kommt dabei eine zentrale Bedeutung zu (Peek, 2005) – sie sind zwei Seiten derselben Medaille, von denen die eine ihr Potenzial nur im Zusammenspiel mit der anderen entfalten kann. Standardsetzung ohne eine entsprechende Überprüfung der gesetzten Standards würde jeglicher Vorstellung von Steuerung zuwider laufen, da nur dann im engeren Sinne von Steuerung gesprochen werden kann, wenn Soll-Ist-Abweichungen Veränderungen im System bewirken (Ditton, 2007; Helmke, 2004). Während Bildungsstandards solche gemeinsamen Ziele formulieren, wird deren Erreichung durch Maßnahmen vergleichender Leistungsmessung überprüft. Dabei müssen Verfahren der Leistungsüberprüfung mit dem Ziel des Bildungs-Monitoring von Verfahren zur Schul- und Unterrichtsentwicklung unterschieden werden (vgl. Kap. 2.1.2). Der neuen Steuerungsstrategie liegt die Vorstellung zugrunde, „dass die Handlungsrationalität in Schulentwicklungsprozessen durch eine Verbesserung der Informationssituation gesteigert werden könnte“ (Heinrich, 2008, S. 31). Inwiefern mehr Information tatsächlich zu mehr handlungsrelevantem Wissen führt, wird an späterer Stelle zu diskutieren sein.

Weiterhin sind für die neue Steuerung im Bildungssystem neben einer starken (Re-)Zentralisierung durch Standardsetzung und Standardüberprüfung Maßnahmen der Deregulierung charakteristisch, infolge derer die Einzelschule einen größeren Spielraum im Umgang mit Ressourcen und zur eigenen Profilbildung erhält (Altrichter & Maag Merki, 2010b). Emmerich (2010) stellt infrage, inwiefern sich so ambivalente Steuerungstrends für eine zielgerichtete Beeinflussung von Schule überhaupt eignen. Peek (2003) hingegen argumentiert mit Verweis auf Dichanz und Tulodziecki (1995), dass die Deregulierung eine neue Dynamik in Schulen bringe:

Durch ein mehr an Selbstgestaltung und Selbstreflexion orientiertes Vorgehen sollen die Eigenkräfte der Schule mobilisiert, Motivation und Engagement gefördert und damit die Grundlagen für eine größere Identifikation mit der eigenen Schule gefestigt werden (vgl. Dichanz & Tulodziecki 1995). (Peek, 2003, S. 16)

Maßnahmen der Deregulierung berücksichtigen, dass Qualitätsentwicklung in Schule und Unterricht nicht einfach top-down – also von oben herab durch die Bildungsadministration – verordnet werden kann, sondern in der Schule von den Akteuren vor Ort getragen werden muss. Die Gewährung größerer Spielräume für solche Entwicklungen birgt gleichzeitig die Gefahr, dass sich Schulen auseinanderentwickeln (Rolff, 1998). Maßnahmen der Zentralisierung schaffen die gemeinsame Basis für die Entwicklung aller Schulen durch verbindliche vorgegebene Ziele (Standards) und deren Überprüfung (vergleichende Leistungsmessung).

Als logische Konsequenz output-orientierter Steuerung ist eine neue Anforderung an Schulen entstanden: Die Nutzung von Informationen aus interner und externer Evaluation. Anhand der erzielten Ergebnisse vor Ort (beispielsweise in Vergleichsarbeiten) sind Schulen angehalten, Rückschlüsse auf ihre eigenen Stärken und Schwächen zu ziehen und an diesen zu arbeiten (vgl. Peek & Dobbelstein, 2003; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2006).

## 2.1.2 Neue Steuerung – neue Instrumente

Output-orientierte Steuerung unterscheidet sich von input-orientierter Steuerung nicht grundlegend durch die verwendeten Instrumente (Diemer & Kuper, 2010). Die Kultusministerkonferenz (KMK) hat dennoch als ein bedeutsames Element der neuen Steuerungsstrategie die Einführung von Bildungsstandards beschlossen, also fachbezogenen Leistungserwartungen, die an bestimmte Übergänge im Bildungsverlauf gebunden sind (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a) und damit einhergehend die zentrale Überprüfung von Schülerleistungen. Im Folgenden wird die Konstruktion von Bildungsstandards und Vergleichsuntersuchungen in Deutschland zu erläutern sein, sowie die Frage, welche Hoffnungen auf und Erwartungen an Möglichkeiten der Steuerung mit diesen Instrumenten verbunden werden.

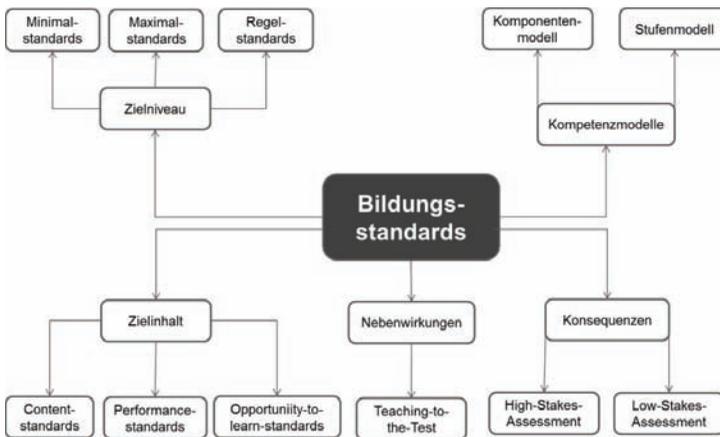
### *Standardsetzung durch Bildungsstandards*

Die Bildungsstandards bilden die Grundlage des neuen Steuerungsmodells (Avenarius et al., 2003; Sander, 2009). Sie sind an Unterrichtsfächer gebunden und beschreiben nicht mehr nur stoffliche Inhalte von Unterrichtsfächern, sondern darüber hinaus prozessorientierte Kompetenzen, die Schülerinnen und Schüler an den Übergängen des Bil-

dungssystems erreicht haben sollen. Seit dem im Jahr 2003 gefällten Beschluss, Bildungsstandards einzuführen, wurden schulformspezifische Standards für die diversen Übergänge im Bildungssystem erarbeitet. Mit der Fertigstellung der Bildungsstandards für die gymnasiale Oberstufe werden für das gesamte allgemeinbildende deutsche Schulsystem abschlussbezogene Standards vorliegen (Köller, 2008), deren Aufgabe es ist, die bis dahin geltenden Lehrpläne abzulösen und langfristig zu ersetzen.

Das Konzept der Bildungsstandards war zum Zeitpunkt der Erarbeitung in Deutschland nicht neu, sondern schon aus Ländern wie England und den USA, in modifizierter Form auch aus den Niederlanden und Skandinavien bekannt (Klieme, 2005). Bildungsstandards sind jedoch nicht gleich Bildungsstandards. Es gibt bedeutsame Unterschiede im Hinblick auf Zielniveau und Zielinhalt als auch bezüglich der Konsequenzen, die ein Nicht-Erreichen der Standards nach sich zieht (vgl. Abbildung 2).

Abbildung 2: Systematik der Bildungsstandards



Quelle: Maag Merki, 2005

Bildungsstandards können sich im Hinblick auf ihr Zielniveau unterscheiden – je nachdem, ob sie als Minimal-, Maximal- oder Regelstandards formuliert sind und damit das basale, das ideale oder das durchschnittliche Erwartungsniveau festgelegt werden soll. Dabei können Lernziele und Inhalte schulischen Lernens (*content standards*) genauso im Fokus stehen wie Kompetenzen (*performance standards*) oder Standards, die sich auf den Input beziehen – z. B. auf die Ausstattung von Schulen, Gestaltung von Unterricht etc. (*opportunity-to-learn standards*) (Maag Merki, 2005; Ravitch, 1995). Insbesondere dann, wenn sich Bildungsstandards auf zu erreichende Kompetenzen beziehen – wie dies zurzeit in Deutschland nahezu ausschließlich der Fall ist (Peek, 2011) – bedarf es differenzierter Kompetenzmodelle, mit denen sowohl Dimensionen einer Kompetenz (Komponentenmodell) dargestellt als auch Entwicklungsverläufe (Stufenmodell)

abgebildet werden können. Insbesondere Modelle zur Abbildung von Entwicklungsverläufen liegen bislang allerdings nicht vor, befinden sich aber in verschiedenen Disziplinen in der Entwicklung (vgl. Klieme, Leutner & Kenk, 2010). Bildungsstandards können sich des Weiteren im Hinblick auf die Konsequenzen unterscheiden, die bei ihrer (Nicht-)Erreichung eintreten (*High-Stakes Assessment* vs. *Low-Stakes Assessment*). Letztendlich lassen sich in Ländern, die Bildungsstandards einführen, häufig auch Nebenwirkungen beobachten, beispielsweise überzogene Reaktionen des Systems in der Form, dass die Lehrkräfte ihre Schülerinnen und Schüler gezielt auf die Tests zur Überprüfung der Bildungsstandards vorbereiten (Kronig, 2008).

Die Kultusminister heben hervor, dass mit der Einführung von Bildungsstandards nicht eine höhere Standardisierung von Unterricht erreicht werden soll, sondern Lehrkräfte sowie Schulleiterinnen und Schulleiter ganz im Gegenteil parallel zur Verpflichtung auf Ergebnisse, die den Standards entsprechen, größere Freiheiten in Fragen der Erreichung dieser Ergebnisse haben:

Bildungsstandards standardisieren nicht die schulischen Lehr- und Lernprozesse. Sie definieren eine normative Erwartung, auf die hin Schule erziehen und bilden soll. Die Wege dorthin, die genaue Einteilung der Lernzeit, der Umgang mit personellen Ressourcen sowie die Implementation von Standards und die notwendigen Unterstützungsmaßnahmen der Schulen bleiben den Ländern überlassen. Damit sind Standards, die die Schulen auf Ergebnisse verpflichten, die Voraussetzung für die Gewährung von mehr Eigenverantwortung der Schulen, z. B. im Bereich von Unterrichtsplanung, Personaleinsatz und -auswahl oder in der Gestaltung von Integrations- und Fördermaßnahmen. (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a, S. 11)

Konzeptionelle Grundlage für die Konstruktion der in Deutschland eingeführten Bildungsstandards war die Expertise zur Entwicklung nationaler Bildungsstandards, die von einer Arbeitsgruppe unter der Leitung von Eckhard Klieme erarbeitet wurde. Bildungsstandards, so schreiben die Autoren,

greifen allgemeine *Bildungsziele* auf. Sie benennen die *Kompetenzen*, welche die Schule ihren Schülerinnen und Schülern vermitteln muss, damit bestimmte zentrale Bildungsziele erreicht werden. Die Bildungsstandards legen fest, welche Kompetenzen die Kinder oder Jugendlichen bis zu einer bestimmten Jahrgangsstufe erworben haben sollen. Die Kompetenzen werden so konkret beschrieben, dass sie in Aufgabenstellungen umgesetzt und prinzipiell mit Hilfe von *Testverfahren* erfasst werden können. (Klieme et al., 2003, S. 19)

Klieme und Kollegen unterscheiden vier Ziele, die mit Hilfe von Standardsetzung und Standardsicherung verfolgt werden sollen, sich aber auf unterschiedliche Ebenen beziehen. Ein Ziel ist dabei die *Entwicklung und Überprüfung von Kompetenzmodellen*, die es möglich macht, Leistungen von Schülerinnen und Schülern vorher beschriebenen Kompetenzstufen zuzuordnen. Darüber hinaus sollen Standards eine größere Transparenz darüber gewährleisten, wo Schülerinnen und Schüler stehen, welche Fähigkeiten sie schon sicher beherrschen und welche Kompetenzen noch erworben werden sollten – für Eltern, Lehrkräfte und die Schülerinnen und Schüler selbst. Sie dienen damit auch der *individuellen Diagnostik*. Andererseits sollen mit Hilfe von Bildungsstandards auch Informationen über das gesamte Schulsystem gewonnen werden (*System-Monitoring*). Auf dieser Ebene geht es vornehmlich um Fragen und Daten, die insbesondere für politische Entscheidungsträger relevant sind. Die vierte und im Rahmen der vorliegenden Untersuchung relevante Ebene ist die Ebene der *Schulevaluation*; Bildungsstandards sollen den Schulen helfen, die Leistung ihrer Schülerinnen und Schüler in einen Gesamtzusammenhang einzuordnen, der über die soziale Bezugsnorm der eigenen Schule oder Klasse hinausgeht. Durch den Bezug auf in den Bildungsstandards festgelegte Kompetenzen, die Schülerinnen und Schüler zu bestimmten Zeitpunkten in ihrer Bildungsbiografie erworben haben sollten, wird den Schulen neben der sozialen auch eine kriteriale Bezugsnorm zur Verfügung gestellt (Watermann & Stanat, 2004). Schülerleistungen werden durch diese Form der Standardüberprüfung über die Grenze der eigenen Klasse oder Schule hinaus vergleichbar. Vorrangige Ziele dieses neuen Instrumentes sind:

- die Vergleichbarkeit schulischer Abschlüsse zu sichern – sowohl zwischen den Bundesländern als auch zwischen einzelnen Schulen,
- die Erhöhung der vertikalen Durchlässigkeit des Bildungssystems und
- die Wahrung der Bildungschancen für alle Schülerinnen und Schüler.

Ditton (2007) sieht in der Einführung von Bildungsstandards einen ersten und wichtigen Schritt zur Zielpräzisierung im Bildungssystem und erkennt darin ein Potenzial, „das es klug zu nutzen gilt“ (S. 92).

Entgegen der eindeutigen Stellungnahme durch die Arbeitsgruppe der Expertise hat die KMK allerdings entschieden, anstatt Mindeststandards in Deutschland Regelstandards einzuführen. Auch der Forderung, Bildungsstandards strikt von Selektions- und Laufbahnentscheidungen zu trennen, kamen die Kultusminister nicht nach. Sander (2009) befürchtet, dass diese Entscheidung „das Projekt Bildungsstandards um einen Gutteil seiner möglichen positiven Effekte bringen wird“ (Sander, 2009, S. 26; vgl. auch Lersch, 2006; Maier, 2009b). Insbesondere für schwache Schülerinnen und Schüler haben Regelstandards den Nachteil, dass sie

(...) implizit die Botschaft [enthalten], dass man eine Art Normalverteilung der Kompetenzen erwartet, bei der es im Vergleich zum Regelfall immer Gewinner

und Verlierer gibt. (. . .) Die für die Stützung leistungsschwächerer Schüler entscheidende Frage, was diese wissen und können müssen, um als erfolgreich gelten zu können, lässt sich mit Regelstandards nicht beantworten – jedenfalls nicht positiv. (Klieme et al., 2003, S. 28)

Sander (2009, S. 12) kritisiert auch, dass die Implementierung von Bildungsstandards durch „teils hektische Aktivitäten“ politischer Entscheidungsträger geprägt war und „keineswegs von einem einheitlichen Gesamtkonzept oder einem konsensuellen Masterplan getragen“ wurde. Trotz der neuen und anspruchsvollen Anforderungen an die Schulen wurden ihnen keine Entlastungen zuteil, die „in der Logik dieses Prozesses“ notwendig gewesen wären (Sander verweist auf Entlastung der Schulleitung von Unterrichtsverpflichtung, Bereitstellung von Personal für das Management der selbstständigen Schule und Abschaffung des Beamtenstatus).

### *Standardüberprüfung durch vergleichende Leistungsmessung*

Es ist schon angeklungen, dass die Überprüfung von Schülerleistungen im Rahmen des neuen Steuerungsmodells nicht mehr nur der individuellen Bewertung von Schülerinnen und Schülern dient, sondern auch der Beobachtung des Bildungssystems insgesamt sowie der Entwicklung von Einzelschulen. Zur Beobachtung der Entwicklung des Bildungssystems sollen die Daten aus internationalen Leistungsvergleichsstudien wie PISA, TIMSS und IGLU<sup>4</sup> und deren nationalen Erweiterungsstudien (PISA-E, TIMSS-E, IGLU-E) genutzt werden sowie die Ergebnisse aus der zentralen Überprüfung der Erreichung der Bildungsstandards im Ländervergleich<sup>5</sup> (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2006). Darüber hinaus sind in den letzten Jahren eine Reihe regionaler Studien mit dem Ziel des System-Monitorings durchgeführt worden, z. B. die Hamburger Vollerhebungen LAU<sup>6</sup> und KESS<sup>7</sup>, QuaSUM<sup>8</sup> in Brandenburg, ELEMENT<sup>9</sup> in Berlin oder MARKUS<sup>10</sup> in Rheinland-Pfalz. Die *National Educational Panel Study* (NEPS) ist an dieser Stelle als Großprojekt zu nennen, das in den kommenden zehn Jahren eine Fülle an Daten – nicht

---

4 Internationale Grundschul-Leseuntersuchung, vgl. z. B. Hornberg, Bos, Buddeberg, Potthoff und Stubbe (2007).

5 Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011 (Stanat et al. 2012).

6 Aspekte der Lernausgangslage und der Lernentwicklung an Hamburger Schulen (Lehmann, Peek & Gänsfuß, 1997).

7 Kompetenzen und Einstellungen von Schülerinnen und Schülern (Bos, 2006; Bos, Bensen et al. 2007).

8 Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (Lehmann, Peek, Gänsfuß, Lutkat, Mücke & Barth, 2000).

9 ELEMENT. Erhebung zum Lese- und Mathematikverständnis Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin (Lehmann & Lenkeit, 2008).

10 Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext (Helmke und Jäger, 2002).

nur zur Kompetenzentwicklung – liefern wird, sondern darüber hinaus zu Lernumwelten, Bildungsentscheidungen, Bildungsrenditen und Übergängen an allen Schnittstellen des Bildungssystems<sup>11</sup>. Der Ertrag von Bildungssystemen soll auf diese Weise sichtbar gemacht werden und Anhaltspunkte für weitere Maßnahmen der Planung und Steuerung liefern. Für Studien mit dieser Zielsetzung werden im Folgenden die Begriffe *Systemmonitoringstudien*, *Large-Scale Assessments* und *Leistungsvergleichsstudien* genutzt.

Davon müssen die unter dem Akronym VERA implementierten *Vergleichsarbeiten* abgegrenzt werden (vgl. Bonsel et al., 2006; Bonsel & von der Gathen, 2004; Hosenfeld, Groß & Bittins, 2006; Peek, 2006a; van Ackeren, 2004). Diese werden in den Fächern Deutsch und Mathematik am Ende von Klasse 3 (VERA 3) beziehungsweise in Deutsch, Mathematik und der ersten Fremdsprache am Ende von Klasse 8 (VERA 8) geschrieben. Nachdem verschiedene Bundesländer zunächst selbstständig ihre eigenen Vergleichsarbeiten erstellten, auswerteten und koordinierten, wurde im Jahr 2008 die zentrale Verantwortung für die Aufgabenkonstruktion dem Institut für Qualitätsentwicklung im Bildungswesen (IQB) übergeben<sup>12</sup>. Seit diesem Zeitpunkt werden Vergleichsarbeiten jährlich bundesweit in den betreffenden Klassenstufen durchgeführt. Lediglich das Bundesland Baden-Württemberg nimmt nicht an VERA 8 teil, sondern führt stattdessen weiterhin selbst konzipierte *Diagnose- und Vergleichsarbeiten* (DVA) in Klasse 9 durch. Darüber hinaus werden in einigen Bundesländern auch zentral koordinierte Vergleichsarbeiten in weiteren Klassenstufen eingesetzt, etwa in Hamburg, Hessen, Mecklenburg-Vorpommern, Sachsen, Schleswig-Holstein und Thüringen am Ende der sechsten Klasse (VERA 6). Vergleichsarbeiten werden in Hamburg, Hessen und Nordrhein-Westfalen (in NRW nur in Klasse 8) auch als *Lernstandserhebungen* (LSE) bezeichnet; in Thüringen und Sachsen ist die Bezeichnung *Kompetenztests* (KT) für die Erhebungen in allen Klassenstufen geläufig.

Vergleichsarbeiten sind als Vollerhebungen konzipiert und die Schulen aller Bundesländer sind verpflichtet, sich an ihnen zu beteiligen. Die Implementation von Bildungsstandards und ergebnisorientierten Unterrichtsentwicklungsmaßnahmen soll unter anderem mit Hilfe der Lernstandserhebungen erfolgen. Die im Rahmen der Lernstandserhebungen eingesetzten Testaufgaben haben im Gegensatz zu den Testaufgaben in Systemmonitoringstudien einen klaren Bezug zu den Bildungsstandards und die Nutzung ihrer Rückmeldeberichte soll Bestandteil einer fachlichen Evaluationskultur in den Schulen werden (Kühle & Peek, 2007). Die Datenerhebung wird von Lehrkräften selbst durchgeführt und die Daten werden über eine Datenmaske online an den externen Partner übermittelt, der die Daten auswertet und das Feedback erstellt (in der Regel Landesinstitute, Universitäten oder Schulministerien).

Die Hauptfunktion der Lernstandserhebungen liegt im Gegensatz zu Studien zum System-Monitoring auf der Ebene der Entwicklung von Schule und Unterricht. Primär,

---

11 Umfassende Informationen finden sich unter <https://www.neps-data.de>.

12 Informationen zur Konzeption und zur Durchführung der Vergleichsarbeiten werden auf der Website des IQB zur Verfügung gestellt: <http://www.iqb.hu-berlin.de/vera>.

so wird betont, sollen die Lernstandserhebungen Schulen befähigen, einen Diskurs „über erreichte und erreichbare Standards“ (Kühle & Peek, 2007, S. 429; Peek, 2011, S. 151) zu führen und damit einen Beitrag zur ergebnisorientierten Schul- und Unterrichtsentwicklung zu leisten (vgl. auch Hosenfeld et al., 2006; Peek, 2006a; Peek & Dobbstein, 2006b). Es sollen Erkenntnisse *über* die Einzelschule *für* die Einzelschule gewonnen werden (vgl. Schulevaluation bei Klieme et al., 2003).

Lernstandserhebungen / Vergleichsarbeiten sind Ausdruck einer an Standards orientierten neuen Bildungsplanung und Schulgestaltung. Sie sollen durch ihre evaluative Funktion zu einer systematischen Wahrnehmung unterrichtlicher Wirkungen führen und – darauf aufbauend – Unterrichtsentwicklung und Fördermaßnahmen unterstützen. (EMSE-Netzwerk, 2008, S. 3)

Als zentrale Ziele von Lernstandserhebungen nennen Bonsen et al. (2006) – neben der schon genannten Möglichkeit der Standardüberprüfung in vergleichender Perspektive zur Vergewisserung der Einzelschule über die eigene Wirksamkeit – die Unterstützung der Einzelschule bei der Feststellung des Lern- und Förderbedarfs in den überprüften fachlichen Bereichen, die Stärkung der diagnostischen Kompetenz von Lehrkräften und die Setzung von Impulsen für die didaktische und methodische Weiterentwicklung des Unterrichts. Darüber hinaus betonen sie die Funktion, den Prozess der Implementation der Bildungsstandards zu unterstützen.

In der oben zitierten Stellungnahme des Netzwerkes „Empiriegestützte Schulentwicklung“<sup>13</sup> drückt sich in aller Deutlichkeit aus, dass es sich bei den Vergleichsarbeiten bzw. Lernstandserhebungen weder um ein diagnostisches Instrument auf Individual-ebene handelt, das mit psychologisch-diagnostischen Testverfahren vergleichbar ist oder als deren Ersatz angesehen werden kann, noch um ein Instrument, das der Kontrolle der Schule durch die Schulaufsicht dient. Es ist jedoch gerade bei den Lernstandserhebungen eine beträchtliche Diskrepanz zwischen theoretischen und tatsächlichen Zielen und Funktionen zu beobachten (Pant, Dezember 2010). Dem Instrument werden in der Praxis Funktionen auf der Ebene des Systemmonitoring, der Rechenschaftslegung, der Zertifizierung und Selektion sowie der Entwicklung der Einzelschule zugeschrieben (van Ackeren, 2004, S. 38). Dadurch entsteht ein Spannungsfeld von formativer und summativer Evaluation (Maier, 2009b, 2010), das die Gefahr einer Funktionsüberfrachtung birgt (Kühle & Peek, 2007; Maier, 2010; Posch 2009). Dieser Umstand wird auch international diskutiert, z. B. in England:

However, the task of assessing pupils summatively for external purposes is clearly different from the task of assessing on-going work to monitor and improve progress. Some argue that these two roles are so different that they should be kept apart. We do not see how this can be done, given that teachers must have

---

13 <http://www.emse-netzwerk.de>.

some share in responsibility for the former and must take the leading responsibility for the latter. (Black & William, 2001, S. 11f)

Dass Daten, die nicht stichprobenbasiert gewonnen werden, sondern aus einer Vollerhebung hervorgehen, für die Beobachtung der Entwicklung auf Systemebene extrem wertvoll sind, versteht sich von selbst. Doch das Hauptaugenmerk der Lernstandserhebungen sollte auf der Entwicklungsfunktion für die Einzelschule liegen, denn deren Entwicklung hat Priorität gegenüber der Systemkoordination (Rolf, 1998) und während es verschiedene Instrumente gibt, die Daten für die Bildungsadministration liefern, sind die Lernstandserhebungen das einzige Instrument, das Schulen Leistungsdaten zur Verfügung stellt, die zur schulinternen Reflexion von Unterricht genutzt werden können. Auch Kühle und Peek (2007) betonen ausdrücklich die Schlüsselposition der Transparenz bezüglich der Intentionalität von Lernstandserhebungen:

Es ist zu befürchten, dass eine dem Projekt unreflektiert zugeschriebene Multifunktionalität (neben der Unterrichtsentwicklung vor allem individuelle Förderung, aber auch Systemmonitoring) zu starken Verunsicherungen im Umgang mit den Ergebnissen führt. Es bedarf hier einer Verstärkung fachlich fundierter Beschreibungen über Potenziale und Grenzen des Projektes (vgl. dazu Peek, 2007). (Kühle & Peek, 2007, S. 445)

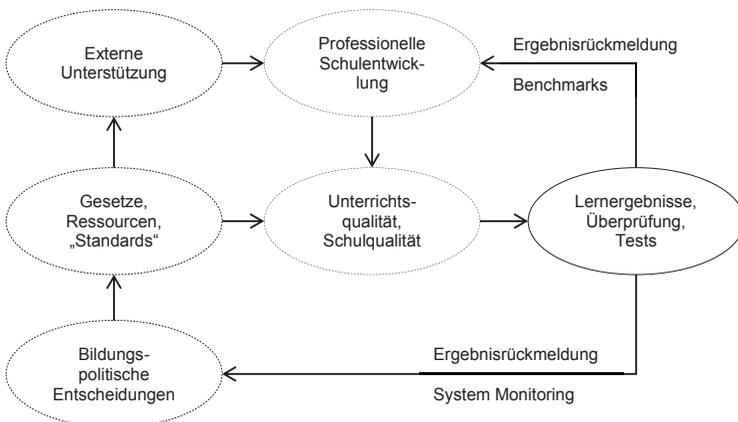
Die Erkenntnisse von Stockmann (2006) unterstützen das Argument von Kühle und Peek: Auch er weist darauf hin, dass Evaluation verschiedene Funktionen haben kann. Neben der *Gewinnung von Erkenntnis* und der *Ausübung von Kontrolle* sind dies die *Schaffung von Transparenz und Dialogmöglichkeiten, um Entwicklungen voranzutreiben* und *Legitimation von durchgeführten Maßnahmen*. Obwohl diese Funktionen alle eng miteinander verknüpft sind, sollte nur eine von Ihnen bei jeder Evaluation schwerpunktmäßig verfolgt werden, denn „die Festlegung auf eine prioritäre Funktion steuert die Herangehensweise und bestimmt das Design und die Durchführung einer Evaluation“ (Stockmann, 2006, S. 69). Das Design der Vergleichsarbeiten unterscheidet sich dementsprechend von dem Design von Systemmonitoringstudien und Tests zur Individualdiagnostik.

Verschiedene Länder, allen voran Großbritannien und die USA, betonen den Kontrollcharakter von Vergleichsarbeiten besonders stark, indem sie mit dem schlechten Abschneiden von Schulen bei Leistungstests zur Überprüfung von Standards sogar dramatische Konsequenzen für die Schule und/oder Lehrkräfte wie Budget- oder Gehaltskürzungen verknüpfen (Maier, 2009b; Nichols & Berliner, 2005; Roach & Frank, 2007). Die Etablierung solcher *High-Stakes Assessments* betont insbesondere die Kontrollfunktion der Leistungsüberprüfung. Haben die Tests hingegen keine negativen Konsequenzen für Schule, Lehrkräfte oder Schülerinnen und Schüler, spricht man von *Low-Stakes Assessments*. Vergleichsarbeiten in Deutschland können – trotz der berichteten Vermi-

sung von Kontroll- und Entwicklungsfunktion – eher der Gruppe der *Low-Stakes Assessments* zugeordnet werden. Keiner Gruppe von Akteuren drohen aufgrund schlechter Ergebnisse in den Vergleichstests ernsthafte negative Konsequenzen (z. B. finanzielle Einbußen, eine verstärkte Kontrolle durch die Schulaufsicht). Dies unterscheidet die Situation in Deutschland deutlich von der Situation in Großbritannien und den USA.

Peek (2006a) weist jedoch darauf hin, dass auch in Deutschland ein „Automatismus positiver Effekte“ (S. 1358) von Lernstandserhebungen nicht zu erwarten ist (vgl. auch Kohler & Schrader, 2004; Nachtigall & Jantowski, 2007; Rolff, 2002; Stamm, 2003). Für die Setzung positiver Impulse spielen die Qualität der Testaufgaben sowie die Rezeption und der Umgang mit den Ergebnissen in Schulen und Bildungspolitik eine entscheidende Rolle. Die zurückgemeldeten Ergebnisse aus extern administrierter vergleichender Leistungsmessung müssen also pädagogisch nutzbar sein, d. h. sie müssen in einer anschlussfähigen Form an die Lehrerinnen und Lehrer zurückgemeldet werden, damit der Transfer von wissenschaftlichem Wissen in pädagogische Handlungsstrategien gelingen kann. Um schließlich aus Ergebnissen zu lernen bedarf es einer Rückkopplung der Informationen, die nach der Überprüfung der Lernergebnisse über eine Ergebnismeldung stattfindet. Schulen benötigen die Kompetenz, Daten aus interner und externer Evaluation zu lesen, zu verstehen und auf ihrer Grundlage Handlungsbedarf zu identifizieren, damit Standardsetzung im Zusammenspiel mit vergleichender Leistungsmessung zu einem Entwicklungsinstrument für Schule und Unterricht werden kann. Um dies leisten zu können, bedürfen die meisten Schulen jedoch externer Unterstützung.

Abbildung 3: Idealmodell outputorientierter Steuerung



Quelle: Specht, 2006, Abb. 1