

## 9 Bildungsstandards und externe Überprüfung von Schülerkompetenzen: Mögliche Beiträge externer Messungen zur Erreichung der Qualitätsziele der Schule

Herbert Altrichter & Anna Kanape-Willingshofer

### 1 Fragestellung und Vorgangsweise

Seit Mitte der 1990er Jahre befindet sich die Governance-Struktur – die Art und Weise, wie Beiträge verschiedener Akteurinnen und Akteure zur Tätigkeit und Leistungserbringung im Schulwesen koordiniert werden – auch in den deutschsprachigen Ländern in Veränderung. Die Entwicklung in Österreich, die im Nationalen Bildungsbericht 2009 eingehender analysiert wurde (vgl. Eder & Altrichter, 2009), ging dabei von einer Erweiterung der autonomen Entscheidungsmöglichkeiten an den Einzelschulen bis zur Forcierung des Konzepts einer „evidenzbasierten, outputorientierten Steuerung“ aus. Als Kern dieser Steuerungsstrategie gilt die Formulierung von verbindlichen Bildungsstandards und deren Überprüfung durch nationale Tests, die in Österreich durch eine Novellierung von § 17 des Schulunterrichtsgesetzes (SchUG BGBl. I Nr. 117/2008) gesetzlich verankert wurden.

Neue Steuerungsstrategie

Dieses neue Steuerungsmodell ist mit vielen Hoffnungen auf eine Rationalisierung und Dynamisierung der Entwicklung der Schulen verbunden, die letztlich zu einer nachhaltigen Verbesserung ihrer Leistungen führen sollen. Das Modell und die mit ihm verbundenen Erwartungen klingen in vielen Einzelaspekten plausibel, doch sind sie nicht im strengen Sinn empirisch geprüft (vgl. van Ackeren, 2007, S. 207; Morris, 2011). Die Forderung, großflächige Systemveränderungen vor ihrer Implementierung einer empirischen Prüfung zu unterziehen, wäre aber kaum zu erfüllen: Innovationen, die das ganze System betreffen, sind schwer experimentell oder „im Kleinen“ zu simulieren. Erfahrungen in anderen Systemen sind aufgrund unterschiedlicher Konstellationen von Kontextfaktoren nicht voll übertragbar. Eine Antizipation der Wirkungen von neuen Arrangements der Systemsteuerung unter Realbedingungen (also im Sinn von *effectiveness*; vgl. Wortman, 1983, S. 230) ist also nur annäherungsweise möglich. Aufgabe dieses Kapitels ist es in diesem Sinn, die Frage zu erörtern, *welche Beiträge die in Österreich in Umsetzung befindliche Version der Steuerungsstrategie „Bildungsstandards und externe Überprüfungen von Schülerkompetenzen“ zur Erreichung verschiedener Qualitätsziele der Schule leisten kann.*

Wirkung auf Qualitätsziele der Schule?

Dabei gehen wir in folgenden Schritten vor: Nach der Definition grundlegender Begriffe analysieren wir in Abschnitt 2 relevante Gesetze und Verordnungen, um die spezifische Ausformung dieser Steuerungsstrategie in Österreich herauszuarbeiten. Die Ergebnisse werden als ein hypothetisches *Wirkungsmodell* dargestellt, in dem erwartete Wirkungen und vermittelnde Prozesse (*Mechanismen*) ausformuliert werden. In Abschnitt 3 werden empirische Untersuchungen zu den Teilprozessen des Wirkungsmodells besprochen; auf dieser Basis wird diskutiert, wie weit angenommen werden kann, dass die neue Steuerungsstrategie tatsächlich die erwarteten Prozesse stimuliert und zu den erhofften Ergebnissen führt. Abschnitt 4 fasst die auf diese Weise erzielten Antworten zusammen und formuliert Optionen für Systementwicklung und Forschung.

Kapitelüberblick

E

## 2 Konzept der Steuerungsstrategien und bildungspolitische Erwartungen

### 2.1 Definitionen grundlegender Konzepte

Definition von Bildungsstandards

Die in den deutschsprachigen Schulsystemen seit der Jahrtausendwende propagierten *Bildungsstandards* stellen eine spezifische Version von „Leistungsstandards“ dar, die die vom Schulsystem erwarteten Qualifikationsleistungen formulieren. Für die bildungspolitische Entwicklung und wissenschaftliche Diskussion in den deutschsprachigen Ländern wurde eine Expertise von Klieme et al. (2007) prägend, die das Konzept folgendermaßen definiert: „Bildungsstandards [...] 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“ (S. 19; Hervorhebungen im Original).

Typischerweise decken Bildungsstandards nicht die gesamte Breite des schulischen Lernens oder eines speziellen Lernbereichs ab, sondern „konzentrieren sich auf einen [bedeutsamen und zentralen] Kernbereich, der für alle Lernenden verbindlich ist“ (Maag Merki, 2010, S. 147). Als Neuerung gegenüber traditionellen Zielformulierungen in Lehrplänen sollen die in den Standards formulierten Kompetenzanforderungen „in Kompetenzmodellen systematisch über verschiedene Kompetenzstufen geordnet [werden], die Aspekte, Abstufungen und Entwicklungsverläufe von Kompetenzen darstellen“ (Maag Merki, 2010, S. 147).

Ideen hinter externen Überprüfungen ...

Wenn im Kontext aktueller bildungspolitischer Diskussionen von *externer Überprüfung von Schülerkompetenzen* gesprochen wird, so können damit zwei unterschiedliche, nicht notwendig miteinander verbundene Ideen angesprochen sein:

(1) Verstärkte Externalisierung der schulischen Leistungsbeurteilung

(1) Einesteils kann damit die Forderung nach einer *stärkeren Externalisierung der schulischen Leistungsbeurteilung*, die sich letztlich in Zeugnissen, Zertifikaten und Berechtigungen niederschlägt, angesprochen sein. Die Leistungsbeurteilungsverordnung (LBVO idF BGBl. II Nr 35/1997) lässt weite Interpretationsspielräume, sodass im österreichischen Schulsystem eine *schulinterne Form der Beurteilung* vorherrscht, bei der nach in der Einzelschule (real meist: von einzelnen Lehrpersonen) produzierten Maßstäben beurteilt wird, die nicht mit jenen der Nachbarschule vergleichbar sein müssen. Die mangelnde Messqualität der einzel-schulischen Leistungsmessung sowie die mangelnde Vergleichbarkeit der Maßstäbe zwischen verschiedenen Schulen und Lehrpersonen sind – verstärkt nach dem Vorliegen landesweit vergleichbarer externer Leistungstests (vgl. Eder, Neuweg & Thonhauser, 2009; Haider & Schreiner, 2006, S. 233) – kritisiert worden.

(2) Überprüfung von Schülerkompetenzen durch externe Testungen

(2) Eine externe Überprüfung von Schülerkompetenzen kann aber auch durch *standardbezogene Tests* (in der deutschsprachigen Diskussion oft als „Lernstandserhebungen“ bezeichnet) erfolgen, deren Ergebnisse ursprünglich nicht für individuelle Leistungsbeurteilung konzipiert waren. Diese untersuchen periodisch die Erreichung der in den Bildungsstandards formulierten Schülerkompetenzen und melden diese Informationen im Sinne eines Systemmonitorings an die Öffentlichkeit und Entscheidungsträger/innen auf verschiedenen Ebenen des Schulsystems zurück (vgl. Prenzel & Seidel, 2010). Von letzteren wird erwartet, dass sie diese Informationen interpretieren, aus ihnen Entscheidungen ableiten und Maßnahmen zur Qualitätsentwicklung im System in Gang bringen können.

Zentrales Gelenkstück evidenzbasierter Steuerung

Bildungsstandards und externe Überprüfungen von Schülerkompetenzen im zweiten Sinn erhalten ihre gegenwärtige bildungspolitische Aktualität nicht als Einzelmaßnahmen. Sie werden als das „zentrale Gelenkstück“ (Klieme et al., 2007, S. 9) innerhalb der Anstrengungen

zur „Sicherung und Steigerung der Qualität schulischer Arbeit“ (Klieme et al., 2007, S. 9), als Signal- und zentrale Wirkungselemente eines neuen Steuerungsmodus für das Bildungssystem – der *evidenzbasierten Steuerung* – angesehen.

Dieses neue Steuerungsmodell soll ein älteres und als veraltet empfundenen ablösen, das durch hierarchisch-bürokratische Koordinationsformen und Fokussierung auf Input-Fragen (statt auf Systemergebnisse; vgl. Klieme et al., 2007, S. 11) charakterisiert war. Durch stärkere Zielorientierung, rationalere (weil „evidenzbasierte“) Entscheidungen, engere Wirkungskontrolle und zeitnahe Umsteuerungsmöglichkeiten versprechen sich die Proponentinnen und Proponenten dieses Steuerungsmodells einen gezielteren Einsatz der knappen Mittel und eine Qualitätssteigerung des Bildungssystems.

## 2.2 Die Bildungsstandard-Politik in Österreich

In Österreich wurden Bildungsstandards seit den 1990er Jahren entwickelt und als eine Maßnahme zur Qualitätsentwicklung thematisiert (Altrichter & Posch, 1999; Eder, Posch, Schratz, Specht & Thonhauser, 2002). Seit 2000 sind Bildungsstandards in den Regierungsübereinkommen vertreten (vgl. Specht, 2006). 2003 forderte eine ministeriell eingesetzte *Zukunftskommission* die Einführung von Standardmessungen (vgl. Zukunftskommission, 2003). Eine „Steuergruppe“ des Bildungsministeriums aus Vertreterinnen und Vertretern von Bildungspolitik, Administration und Schulaufsicht formulierte schließlich die österreichische Standard-Strategie (vgl. Lucyshyn, 2004) mit folgenden *Zielen* aus:

- Bessere Leistungen der Schüler/innen durch regelmäßige Rechenschaftslegung (Lucyshyn 2004, S. 1),
- Gleichwertigkeit der schulischen Ausbildung und der Abschlüsse (Lucyshyn, 2004, S. 2),
- Sicherung der Durchlässigkeit des Systems (Lucyshyn, 2004, S. 2),
- Beiträge zur Systementwicklung durch „Hinweise auf notwendige Weiterentwicklungen“ (Lucyshyn, 2004, S. 12).

Im Jahr 2008 erfolgte die *gesetzliche Verankerung* der Bildungsstandard-Politik durch die Novellierung von § 17 des Schulunterrichtsgesetzes (BGBl. I Nr. 117/2008). Dieses Dokument und weitere begleitende Regelungen werden im Folgenden analysiert. Die österreichische Standardstrategie entspricht in zentralen Begriffsbestimmungen und Strategieelementen weithin den Vorstellungen in den Gutachten von Klieme et al. (2007) sowie der Zukunftskommission (vgl. Zukunftskommission, 2003, 2005). Es gibt jedoch auch einige bedeutsame Abweichungen:

- *Kompetenztestung am Ende von Bildungsgängen*: Die Überprüfung der Schülerleistungen erfolgt „an zentralen Schnittstellen des Schulsystems“ (Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens [BIFIE], 2012a) und nicht schon früher, beispielsweise in der Mitte von Bildungsgängen. Klieme et al. (2007) hatten argumentiert, dass sich das „nationale Bildungsmonitoring nicht auf die Abschlussjahrgänge beziehen“ (S. 138) sollte, da dadurch die erhoffte entwicklungsstimulierende Funktion beeinträchtigt und in eine selektionsunterstützende Funktion verwandelt werden könnte.
- *Formulierung von „Regelstandards“*: Bildungsstandards können „eine Stufe festlegen, unter die kein Lernender zurückfallen soll („Mindeststandard“), eine mittlere Niveaustufe, die im Durchschnitt erreicht werden soll („Regelstandard“), oder ein Ideal („Maximalstandard““ (Klieme et al., 2007, S. 27). Die deutsche Expertengruppe hatte wie auch die österreichische Zukunftskommission (vgl. Zukunftskommission, 2005, S. 37) Mindeststandards empfohlen, denn die „Kompetenzmodelle und die zugehörigen Aufgaben (Operationalisierungen) beschreiben dann eindeutig, welche Leistungen ein/e Schüler/in erbringen muss, um die Mindestanforderungen zu erfüllen“ (Klieme et al., 2007, S. 27).

Hintergründe und Ziele der neuen Strategie

Gesetzliche Grundlagen der Bildungsstandard-Politik

Zeitpunkt der Standardüberprüfungen im Schulsystem?

Mindest-, Regel- oder Maximalstandards?

Dadurch würden „gerade die Leistungsschwächeren nicht zurückgelassen“ (Klieme et al., 2007, S. 27).

Die Bildungspolitik hat sich jedoch in Österreich wie auch in Deutschland entschlossen, Bildungsstandards als Regelstandards festzulegen (Lucyshyn, 2006, S. 5). Dies wurde mit den Erfordernissen eines differenzierten Schulsystems begründet (Lucyshyn, 2006, S. 6). Dadurch verlieren Standards jedoch ihre klare Signalfunktion: Regelstandards legen weder der Öffentlichkeit noch leistungsschwächeren Schüler/innen klar, was man können muss, um als erfolgreich zu gelten. Sie führen den Leistungsschwächeren aber vor Augen, wie weit sie von einem Ideal bzw. dem Durchschnitt entfernt sind (vgl. Klieme et al., 2007).

Verknüpfung der  
Standardüberprüfung mit  
Leistungsbeurteilung?

- *Trennung der Standardüberprüfung von der Leistungsbeurteilung:* In der SchUG-Novelle von 2009 (BGBl. II Nr 1/2009) wird explizit festgehalten, dass die Bestimmungen über die Leistungsfeststellungen und -beurteilungen von den gesetzlich verordneten Bildungsstandards und den Ergebnissen der Standardüberprüfung unberührt bleiben, was auch der Empfehlung von Klieme et al. (2007, S. 10) entspricht. Eine *Objektivierung der Vergabe von Berechtigungen* im Zuge der externen Standardüberprüfung, wie von der Zukunftskommission (Zukunftskommission, 2005, S. 34) gefordert, ist in Österreich nicht vorgesehen (Ausnahme: neue Reife- und Diplomprüfung) und wäre derzeit aufgrund des längeren Zeitraums zwischen Prüfung und Rückmeldung auch technisch nicht möglich (vgl. BIFIE, 2012a).

### 2.3 Bildungspolitische Erwartungen und Wirkungsmodell(e)

Was soll sich durch die Implementation der neuen Steuerungsinstrumente „Bildungsstandards und standardbezogene Tests“ im österreichischen Schulsystem ändern? Welche Wirkungen sollen sie auf welchen Wegen entfalten?

Wir wollen im Folgenden ein *Wirkungsmodell* der Steuerungsstrategie „Bildungsstandards und periodische Standardüberprüfung“ rekonstruieren. Das Vorgehen folgt den Vorschlägen von Leeuw (2003; vgl. auch Ehren, Leeuw & Scheerens, 2005; Ehren, Altrichter, McNamara & O’Hara, in Druck; ähnlich Resnick, Besterfield-Sacre, Mehalik, Sherer & Halverson, 2007). Dabei wird die „Programmtheorie“ eines politisch-praktischen Programms durch eine schrittweise Analyse der im Folgenden genannten rechtlichen Dokumente erarbeitet. Altrichter und Maag Merki (2010, S. 34) haben für die dabei entstehenden Analyseergebnisse die Bezeichnung „Wirkungsmodell“ vorgeschlagen, weil in einem solchen Modell rekonstruiert wird,

- ... durch welche vermittelnden Prozesse (hier als „Wirkungsmechanismen“ bezeichnet; vgl. Balog & Cyba, 2004; Langer, 2008) ...
- ... die *Elemente* eines politisch-praktischen Programms (in diesem Fall: die Steuerungsstrategie „Bildungsstandards und externe Leistungsüberprüfungen“ und ihre verschiedenen Begleitmaßnahmen) ...
- ... zu den erwarteten *Wirkungen* dieses Programms führen sollen.

Gab es bisher keine Modelle, die die Vermittlungsprozesse beim Wirksamwerden der neuen Steuerungsinstrumente plausibel machen könnten? In der deutschsprachigen Forschung (z. B. Altrichter, 2010, S. 228–231; Schulze, 2012) wird meist auf zwei Konzeptionen zurückgegriffen: Helmkes (2004) „Rahmenmodell zur pädagogischen Nutzung von Vergleichsarbeiten“ thematisiert unterrichtsbezogene Feedbackprozesse aus der Perspektive der einzelnen Lehrperson. Das „Rahmenmodell für die Rückmeldung von Leistungsdaten“ (Vischer & Coe, 2003; adaptiert von Bosen & von der Gathen, 2004) nimmt die Perspektive der Einzelschule ein. Beide konzentrieren sich auf die Verarbeitungsprozesse von Leistungsdaten und nehmen die Erwartungsbildung durch die Steuerungsmaßnahmen nicht in ihre

Konzeptualisierung auf. Auch berücksichtigen sie (verständlicherweise) die Spezifika der österreichischen Standardpolitik nicht und werden dem Mehrebenencharakter des Schulsystems (vgl. Web-Dokument 9.1) durch ihre Konzentration auf jeweils eine spezifische Ebene (Lehrer/in bzw. Einzelschule) nicht gerecht.

### 2.3.1 Elemente des neuen Steuerungsmodells

Das neue Steuerungsmodell (vgl. Abbildung 9.1), das durch die Standards-Politik eingeführt wird, besteht aus den zwei Hauptelementen (vgl. § 17 SchUG idGF):

(A) *Bildungsstandards* sind „konkret formulierte Lernergebnisse, die sich aus den Lehrplänen ableiten lassen. Sie legen jene Kompetenzen fest, die Schüler/innen bis zum Ende der 4. Schulstufe in Deutsch und Mathematik sowie bis zum Ende der 8. Schulstufe in Deutsch, Mathematik und Englisch nachhaltig erworben haben sollen“ (BIFIE, 2012a).

Bildungsstandards

(B) *Periodische Standardüberprüfungen* finden flächendeckend auf der 4. bzw. 8. Schulstufe statt und überprüfen für jene Lernbereiche, in denen Standards formuliert wurden, welche Kompetenzen die Schüler/innen erworben haben. Die Ergebnisse der Standardüberprüfung werden Schülerinnen/Schülern, Lehrpersonen, Schulleiterinnen/Schulleitern und der Schulaufsicht rückgemeldet, wobei alle Akteurinnen und Akteure neben Vergleichswerten die „sie betreffenden“ Resultate mittels Zugangscode online abrufen können, also Schüler/innen ihre individuellen Ergebnisse, Lehrpersonen die aggregierten Klassenergebnisse, nicht aber individuelle Schülerwerte usw. Bei den Rückmeldungen werden Geschlecht, Schultyp, Leistungsgruppe und Migrationshintergrund berücksichtigt, wodurch „faire Vergleiche“ angestrebt werden (Breit et al., 2012). Die erste Überprüfung (Mathematik, 8. Schulstufe) fand im Mai 2012 statt; die Ergebnisse sollen ab Dezember 2012 abrufbar sein.

Periodische  
Standardüberprüfung

Darüber hinaus werden in den Erläuterungen zur SchUG-Novelle (Vorblatt und Erläuterungen, 2008) einige Begleitmaßnahmen angekündigt: Lehrerbildung, Bereitstellung von zusätzlichen Materialien sowie verbindliche Umsetzung entsprechender Schulentwicklungskonzepte. Aus dem Webauftritt des BIFIE (2012b) ist zu entnehmen, dass folgende *Begleitmaßnahmen zur Implementierung der Bildungsstandards* gesetzt werden:

Begleitmaßnahmen

(C) *Information* „aller beteiligten Personengruppen und Systemebenen“ (BIFIE, 2012b)

(D) *Fortbildung* „für Multiplikatorinnen und Multiplikatoren, Führungskräfte (LSI [Landeschulinspektorinnen/-inspektoren], BSI [Bezirksschulinspektorinnen/-inspektoren], Schulleiter/innen) und Lehrende an Pädagogischen Hochschulen“ (BIFIE, 2012b). Zusätzlich können sog. *Rückmeldemoderatorinnen und -moderatoren* von Schulleiterinnen und Schulleitern als Unterstützung bei der Aufarbeitung der Ergebnisse aus den Standardüberprüfungen für zwei Termine von Pädagogischen Hochschulen angefordert werden. Durch die Initiative „Entwicklungsberatung in Schulen“ (EBIS, 2012) wird versucht, verschiedene schul- und unterrichtsbezogene Beratungsleistungen zu koordinieren und Qualitätsanforderungen dafür festzuschreiben.

(E) *Entwicklung und Verbreitung von Unterrichtsmaterialien*, und zwar „Publikation und fortlaufende Aktualisierung unterrichtsnaher Lehr-, Lern- und Begleitmaterialien (Praxishandbücher, Themenhefte, Aufgabensammlungen, Leitfäden, Broschüren, Videovignetten etc.) in gedruckter bzw. elektronischer Form“ (BIFIE, 2012b).

(F) *Entwicklung und Verbreitung von standardbezogenen Diagnoseinstrumenten* „zur Informellen Kompetenzmessung (IKM), die Lehrkräften Lernstandserhebungen auf Schüler- und Klassenebene ermöglichen und damit wesentlich zur gezielten Förderung der Lernenden im Unterricht beitragen“ (BIFIE, 2012b).

### 2.3.2 Wirkungshoffnungen

#### Wirkungshoffnungen in zentralen Rechtstexten

Welche Erwartungen in Hinblick auf *Wirkungen* werden mit Bildungsstandards und externen Überprüfungen von Schülerkompetenzen verbunden? In den zentralen Gesetzes- und Verordnungstexten sind die Ziele und damit die erhofften Wirkungen der neuen Steuerungsstrategie einestheils in sehr allgemeiner Weise durch Vorstellungen „höherer Qualität an Schulen“ charakterisiert, ohne dass Ergebnisziele im Sinne einer Outputorientierung angegeben wären. Konkretisierungen geschehen vor allem über die Angabe von Zwischenprozessen, die zu der erhofften Schulqualität führen sollen. Solche sind:

- Kompetenzorientierte Unterrichtsplanung und -gestaltung durch die Lehrpersonen (inkl. „bestmöglicher Diagnostik und individueller Förderung“, § 17 Abs. 1a SchUG idF BGBl. I Nr. 36/2012);
- „langfristige systematische Qualitätsentwicklung“, „Qualitätssicherung“, „Steuerung und Planung“ „in der Schule und im Bildungswesen insgesamt“ (Vorblatt und Erläuterungen, 2008);
- „Transparenz, Objektivität und Vergleichbarkeit“ (Vorblatt und Erläuterungen, 2008).

In den von uns geprüften zentralen rechtlichen Texten und Erläuterungen finden sich keine Stellen, die als Hinweise auf das Ziel Chancengleichheit oder -gerechtigkeit interpretiert werden könnten, außer man gibt sich mit „Transparenz, Objektivität und Vergleichbarkeit“ als Bedingungen von Chancengleichheit zufrieden.

Da die Zielgrößen in den speziell auf Bildungsstandards bezogenen rechtlichen Texten unterbestimmt sind, gehen wir aufgrund anderer gesetzlicher Bestimmungen (§ 2 SchOG idF BGBl. I Nr. 36/2012) und Theorien zur Schulqualität (Fend, 2006; Belfield & Levin, 2009) davon aus, dass sich die verbesserte Qualität schulischer Arbeit, als zentrales Ziel der Einführung von Bildungsstandards, zumindest in folgenden Zielgrößen niederschlagen sollte:

*Zielgröße 1 (Z1) – verbesserte Schülerkompetenzen/Leistungssteigerung des Systems:*

Ausgangspunkt und dynamisierendes Element für die Politik der Umsteuerung waren die enttäuschenden PISA-Ergebnisse. Das neue Steuerungsregime muss sich letztlich in verbesserten Schülerkompetenzen und damit in einer Leistungssteigerung des Schulsystems niederschlagen. Zentrale Aspekte wie „Leistungsförderung“ (Zukunftskommission, 2003, S. 35) und „Verbesserungen der Bildungserträge“ (Eder, 2009, S. 51) können auf verschiedenen Wegen geschehen: Einestheils könnte eine konsequente *Fokussierung des Unterrichts auf die Kompetenzentwicklung* der Schüler/innen zu einer Steigerung der (gemessenen) Leistungen auf der ganzen Breite führen und so die durchschnittliche Leistungsfähigkeit des Systems fördern. Auf der anderen Seite können *differenzierte Maßnahmen für spezielle Zielgruppen* einen Unterschied machen. Eder (2009, S. 51) sieht beispielsweise in der verstärkten und gezielten Förderung von leistungsschwächeren Schülerinnen und Schülern das größte Potenzial für nachhaltige Verbesserungen der Bildungserträge.

#### Zielgröße 1: verbesserte Schülerkompetenzen

#### Zielperspektive: Enkulturation und Integration

Im Sinne der Aufgaben der österreichischen Schule (siehe § 2 SchOG idF BGBl. I Nr. 36/2012) und beispielsweise von Fends (2006, S. 49–53) Schultheorie wird man unter den „Leistungen des Schulsystems“ nicht alleine seine qualifikatorischen Leistungen verstehen, wie sie sich in disziplinbezogenen kognitiven Leistungstests und der späteren Berufsbewährung von Schülerinnen und Schülern niederschlagen. Schulen haben auch eine „normativ zu gestaltende Aufgabe“ (Fend, 2006, S. 52) im Bereich der *Enkulturation* – der „Sinnvermittlung“ und der Förderung von „Fähigkeiten der Reflexion, der eigenständigen Urteilsbildung und der moralischen Entscheidungsfähigkeit“ – ebenso wie im Bereich der *Integration* unter den Ansprüchen von „Demokratisierung“ und „Inklusion aller Individuen in den menschlichen Verantwortungszusammenhang“ (Fend, 2006, S. 52; vgl. auch Belfield & Levin, 2009, S. 518).

*Zielgröße 2 (Z2) – Chancengleichheit und Bildungsgerechtigkeit im Schulwesen:*

Der durch internationale Schulleistungsvergleiche wieder aufgedeckte enge Zusammenhang zwischen sozialer Herkunft und Bildungskarrieren (z. B. Bacher, 2005; siehe auch das 5. Kapitel in diesem Band; Bruneforth, Weber & Bacher, 2012) stellt für die Steuerleute des Systems mindestens zwei Probleme dar: Einesteils ein *strategisches Problem*, weil die Verbesserung der Lernleistungen von Schülerinnen und Schülern aus unterprivilegierten Gruppen Potenzial für die Verbesserung der Gesamtleistung des Bildungssystems bietet; anderenteils ein *moralisches Problem*, weil dieser Zusammenhang ein „Ausdruck von Ungerechtigkeit [ist], dass ausgerechnet diejenigen am wenigsten vom Bildungssystem profitieren [...] die es am meisten benötigen“ (Stojanov, 2007, S. 29). Daher muss sich ein erneuertes Steuerungsregime auch der Förderung von Bildungsgerechtigkeit annehmen und tatsächlich werden Bildungsstandards häufig als ein Instrument dafür angesehen (vgl. Klieme et al., 2007, S. 54; Blömeke, Herzig & Tulodziecki, 2005, S. 151; Kornhaber, 2004, S. 97).

Zielgröße 2:  
Chancengleichheit und  
Bildungsgerechtigkeit

Als Ausdruck von Chancenungleichheit werden vor allem zwei Typen von *Disparitäten* gesehen:

Zwei Typen von  
Disparitäten ...

- einerseits Disparitäten zwischen den in objektivierten Tests gemessenen Leistungen und den schulisch vergebenen Berechtigungen, welche durch die „Formulierung von verbindlichen Kriterien (Standards) als Voraussetzung für die Zuerkennung von Berechtigungen, und ihre Messung durch objektivierte Leistungstests“ (Eder, 2009, S. 51) verringert werden könnten,
- andererseits Disparitäten der gemessenen Leistung zwischen verschiedenen sozialen Gruppen. Im Vordergrund stehen Unterschiede „zwischen Regionen, zwischen Kindern aus unterschiedlichen sozialen Schichten, zwischen in Deutschland Aufgewachsenen und Zugewanderten“ (Klieme et al., 2007, S. 11); auch geschlechtsbezogene Disparitäten wären hier zu berücksichtigen, spielen aber in den untersuchten zentralen Programmschriften eine untergeordnete Rolle.

Mit diesen beiden Disparitäten korrespondieren zwei Interpretationen von Chancengerechtigkeit, die nicht unbedingt mit den gleichen Strategien zu erreichen sind:

... gehen mit zwei  
Interpretationen von  
Chancengerechtigkeit  
einher

- *Zugangs- und Beurteilungsgerechtigkeit*: Einesteils wird als „gerecht“ empfunden, wenn alle Schüler/innen gleichen Zugang zu Bildung bekommen, und familiäre Hintergründe bei der Notengebung keine Rolle spielen (Stojanov, 2007).
- *Prozess- oder Entwicklungsgerechtigkeit*: Das zweite Konzept von Chancengleichheit fokussiert auf die durch Herkunft entstandenen Benachteiligungen. Es gibt dem Bildungsprozess selbst die Aufgabe, an einer „gleichberechtigte[n] Befähigung zur (Weiter-)Bildung [zu arbeiten und] gleiche Chancen zum sozialen Aufstieg und sozialer Partizipation“ (Stojanov, 2007, S. 35–36) zu eröffnen.

### 2.3.3 Wirkungsmechanismen

Es stellt sich die Frage, *auf welche Weise* Bildungsstandards zur „Sicherung und Steigerung der Qualität schulischer Arbeit“ (Klieme et al., 2007, S. 9) beitragen sollen? Wir sammeln und ordnen im Folgenden Aussagen zu den vermittelnden Prozessen zwischen Bildungsstandards bzw. Standardüberprüfungen und den angestrebten Wirkungen.

Wie wirken  
Bildungsstandards?

#### (1) Setzen und Kommunizieren von Erwartungen

Bildungsstandards dienen zu allererst zur „Orientierung der Schulen auf verbindliche Ziele“ (Klieme et al., 2007, S. 9; vgl. auch O’Day, 2004). Sie sollen ein klares und transparentes Referenzsystem für professionelles Handeln und ebenso für Lernen oder staatliche Entwicklungsentscheidungen bilden (Klieme et al., 2007, S. 47–48).

Setzen und Kommunizieren  
von Erwartungen

„Verähnlichung“ von  
Erwartungen

*(1a) „Verähnlichung von Erwartungen“*

Durch die Formulierung als Bildungsstandards soll ein Ausschnitt von allen potenziell vermittelbaren Kompetenzen betont und aufgewertet werden, was – im Erfolgsfall – zu einer Angleichung der real verfolgten Erwartungen im System führt. Die Standardisierung von Kompetenzen ist einerseits erforderlich, um diese messbar zu machen (Vorblatt und Erläuterungen, 2008); andererseits sollen verbindliche Standards „zur Verähnlichung der Anforderungen – und damit zur Urteilsgerechtigkeit“ (Eder et al., 2009, S. 254) beitragen. Zentral entwickelte und durchgeführte Leistungsfeststellungen könnten „einen erheblichen Beitrag zur Behebung regionaler und standortbezogener Unterschiede leisten“ (Eder et al., 2009, S. 265) und durch sachbezogene Rückmeldungen „Fairness und Chancengleichheit im Bildungssystem deutlich erhöhen“ (Eder et al., 2009, S. 265).

Akteurinnen und Akteure  
sollen diese Erwartungen  
berücksichtigen

*(2) Berücksichtigung dieser Erwartungen bei den Handlungsentscheidungen der Akteurinnen und Akteure auf allen Ebenen*

Der Mechanismus (1) „Setzen und Kommunizieren von Erwartungen“ „funktioniert“, wenn Akteurinnen und Akteure auf verschiedenen Ebenen diese Erwartungen bei ihren Entscheidungen und Handlungen berücksichtigen, so z. B. Lehrpersonen bei ihrer Unterrichtsplanung und -durchführung, Schulleitungen bei der Fortbildungsplanung und beim Setzen von Entwicklungsprioritäten, die regionale Schulaufsicht bei regionalen Initiativen zur Fortbildung und zur Stimulierung von Schulentwicklung usw. Dies gilt ebenso für Schüler/innen, die Bildungsstandards beim Lernen berücksichtigen sollen; auch Familien tun wahrscheinlich gut daran, wenn sie in ihrer familiären Interaktion (z. B. bei Motivation, Aufmerksamkeitslenkung, Schaffung zusätzlicher Lernmöglichkeiten) auf spezifische Anforderungen von Bildungsstandards achten.

... bei der Planung und  
Gestaltung von Unterricht

*(2a) Kompetenz- und ergebnisorientierte Unterrichtsplanung und -gestaltung*

Eine besondere Stellung bei diesen handlungsbezogenen Erwartungen nimmt in den österreichischen Rechtsgrundlagen ein, dass Lehrpersonen Bildungsstandards zu einem orientierenden Prinzip von Unterrichtsplanung und -gestaltung machen sollen (vgl. Verordnung über Bildungsstandards, § 3 Abs. 2 BGBl. II Nr 1/2009).

... bei Beobachtung und  
Diagnose

*(2b) Beobachtung und Diagnose für die individuelle Förderung*

Gemäß Verordnung des BMUKK sollen Bildungsstandards auch einen Bezugspunkt für die Beobachtung und Diagnose von Entwicklungsverläufen von Schülerinnen und Schülern geben, die eine Grundlage für „bestmögliche individuelle Förderung der Schülerinnen und Schüler“ (§ 3 Abs. 3 BGBl. II Nr 1/2009) darstellen.

Messung und  
Rückmeldung von  
Kompetenzen

*(3) Periodische Messung von Kompetenzen und (4) Rückmeldung von tatsächlich erreichten Kompetenzen*

Durch periodische Standardüberprüfungen wird die Informationsbasis für einen Ist-Soll-Vergleich gelegt (Vorblatt und Erläuterungen, 2008). Die Ergebnisse der Messungen auf den Kompetenzdimensionen der Bildungsstandards werden an alle Systemakteurinnen und -akteure in unterschiedlich aggregierter Weise zurückgemeldet (vgl. BIFIE, 2012a). Die Möglichkeit der Wirkung von standardbezogenen Tests hing auch für die Zukunftskommission (Zukunftskommission, 2005, S. 37–38) stark von der Verwertung der Rückmeldungen ab, weshalb sie die „Erarbeitung entwicklungsfördernder Rückmeldeformen an die Schulen“ (Zukunftskommission, 2005, S. 38) forderte.

Rückmeldungen sollen  
Ist-Soll-Diskrepanzen  
aufdecken und Handlungen  
auf verschiedenen Ebenen  
stimulieren

*(5) Entdeckung von Ist-Soll-Diskrepanzen und (6) Entwicklungshandlungen auf allen Ebenen*

Diese Rückmeldungen erlauben Akteurinnen und Akteuren auf allen Ebenen Diskrepanzen zwischen den angestrebten Standards und den erreichten Leistungen zu entdecken. Die Diskrepanzen sollen ebenenspezifische Entwicklungshandlungen motivieren und in die richtige Richtung lenken. Wichtig erscheint dem Gesetzgeber, dass die Ergebnisse in einer Form rückmeldet werden, „dass sie für die langfristige systematische Qualitätsentwicklung in den Schulen nutzbringend verwendet werden können“ (Vorblatt und Erläuterungen, 2008).



Aber auch die Akteurinnen und Akteure auf übergeordneten Ebenen sollen die Rückmeldungen „als Basis für Steuerungsentscheidungen im Bildungswesen“ (Vorblatt und Erläuterungen, 2008) verwenden. Im österreichischen Modell sind auch Schüler/innen und Eltern Adressatinnen/Adressaten der Rückmeldungen (vgl. Maier & Kuper, 2012, S. 89); diese sollen so gestaltet sein, „dass damit die Motivation und die Richtung des Lernens positiv beeinflusst wird“ (Vorblatt und Erläuterungen, 2008).

Das Element, das Entwicklungshandlungen motivieren soll, ist offenbar primär die *kognitive Einsicht in eine Ist-Soll-Diskrepanz*. „Ziel ist eine neue Form der ‚testdatenbasierten‘ Schulentwicklung, d. h., eine interne Datennutzung, weitgehend ohne externen, testbasierten Rechenschaftsdruck“ (Maier, 2010, S. 127). Damit ist der österreichische *Accountability*-Ansatz, wie insgesamt die deutschsprachigen, „weit entfernt von harten Sanktionen“ (Maier & Kuper, 2012, S. 90), wie wir sie aus den angelsächsischen Systemen kennen. Dennoch bestehen „prinzipiell die gleichen Funktionserwartungen: Sie sollen dazu dienen Qualität zu sichern und gleichzeitig Lehrkräfte zur Verbesserung des Unterrichts stimulieren. Verschieden sind jedoch die angestrebten Steuerungsmechanismen, die im Falle der ‚high-stakes-tests‘ auf administrativ induzierte Leistungskonkurrenz und im Falle der Vergleichsarbeiten auf die *informationelle Unterstützung* [Hervorhebung v. Verf.] professioneller Schulentwicklung setzen“ (Maier & Kuper, 2012, S. 90).

#### (6a) *Qualitätsentwicklung an Schulen*

Mit welchen „ebenspezifischen Entwicklungshandlungen“ wird gerechnet? In der Verordnung zu den Bildungsstandards (§ 3 Abs. 4 BGBl. II Nr 1/2009) wird die Schulebene betont, auf der Maßnahmen zur Qualitätsentwicklung von Schulleitungen und Lehrpersonen umgesetzt werden sollen. Man kann annehmen, dass durch die Rückmeldung auch der Prozess (2a) „Kompetenz- und ergebnisorientierte Unterrichtsplanung und -gestaltung“ weiter bestärkt und stimuliert werden soll. Für den Prozess (2b) „Beobachtung und Diagnose für die individuelle Förderung“ gilt dies nicht in gleicher Weise, weil durch die in Österreich verordnete Form der Rückmeldung den Lehrpersonen keine schülerindividuellen Daten zur Verfügung stehen und daher auch keine spezifische Stimulation dieses Prozesses zu erwarten ist. Dies „erschwert eine mittelbare Nutzbarmachung der Testergebnisse für die Weiterentwicklung des Unterrichts und die individuelle Förderung der Schüler/innen, und lässt damit ein erhebliches Potenzial ungenützt“ (Eder et al., 2009, S. 254).

Rückmeldungen sollen Qualitätsentwicklungen stimulieren ...

#### (7) *Evaluierung der Qualitätsentwicklung an Schulen*

Durch diese Informationen soll eine „Kultur kontinuierlicher Selbstevaluationsmaßnahmen und gemeinsamer Qualitätsentwicklung von Lehrerinnen und Lehrern sowie an Schulstandorten insgesamt gefördert werden“ (Vorblatt und Erläuterungen, 2008).

... und zu verstärkter Selbstevaluation an Schulen führen

Werfen wir noch einen speziellen Blick auf *Mechanismen, die zur Erhöhung der Chancengerechtigkeit* führen können. Wie oben erwähnt, werden von einer (1a) „Verähnlichung von Erwartungen“ Wirkungen in Richtung gerechterer Anforderungen und Beurteilungen erwartet (vgl. Eder et al., 2009, S. 254; Klieme et al., 2007, S. 54), mithin auf Ziele von Chancengleichheit des Typs 1, auf die Zugangs- und Beurteilungsgerechtigkeit. Damit Standardüberprüfungen auch Chancengerechtigkeit verbessern, dürften die Teilprozesse (5) „Entdeckung von Ist-Soll-Diskrepanzen“ und (6) „Entwicklungshandlungen“ von den Akteurinnen und Akteuren auf den verschiedenen Ebenen nicht nur leistungsbezogen interpretiert werden, sondern auch gerechtigkeitsbezogen, z. B. in folgender Weise:

#### (5a) *Entdeckung von Chancenungleichheit*

Standardüberprüfungen sollen die Entdeckung von Chancenungleichheit erlauben, die sich vor allem als soziale und regionale Disparitäten niederschlägt (vgl. Beer, 2007, S. 43; Blömeke et al., 2005, S. 151; Klieme et al., 2007, S. 27).

(6b) Förderung von Zielgruppen

Einmal entdeckt, soll Chancengleichheit durch eine verstärkte und gezielte Förderung von Zielgruppen (vgl. Eder et al., 2009, S. 51) verringert werden.

### 2.3.4 Zusammenführung in einem Wirkungsmodell

Die Zusammenführung der Analyseergebnisse ergibt – trotz der modelltypischen Reduzierungen – einigermaßen komplexe Zusammenhänge. Über die obige Analyse hinaus sind in Abbildung 9.1 einige hypothetische Teilprozesse auf verschiedenen Akteursebenen dargestellt.

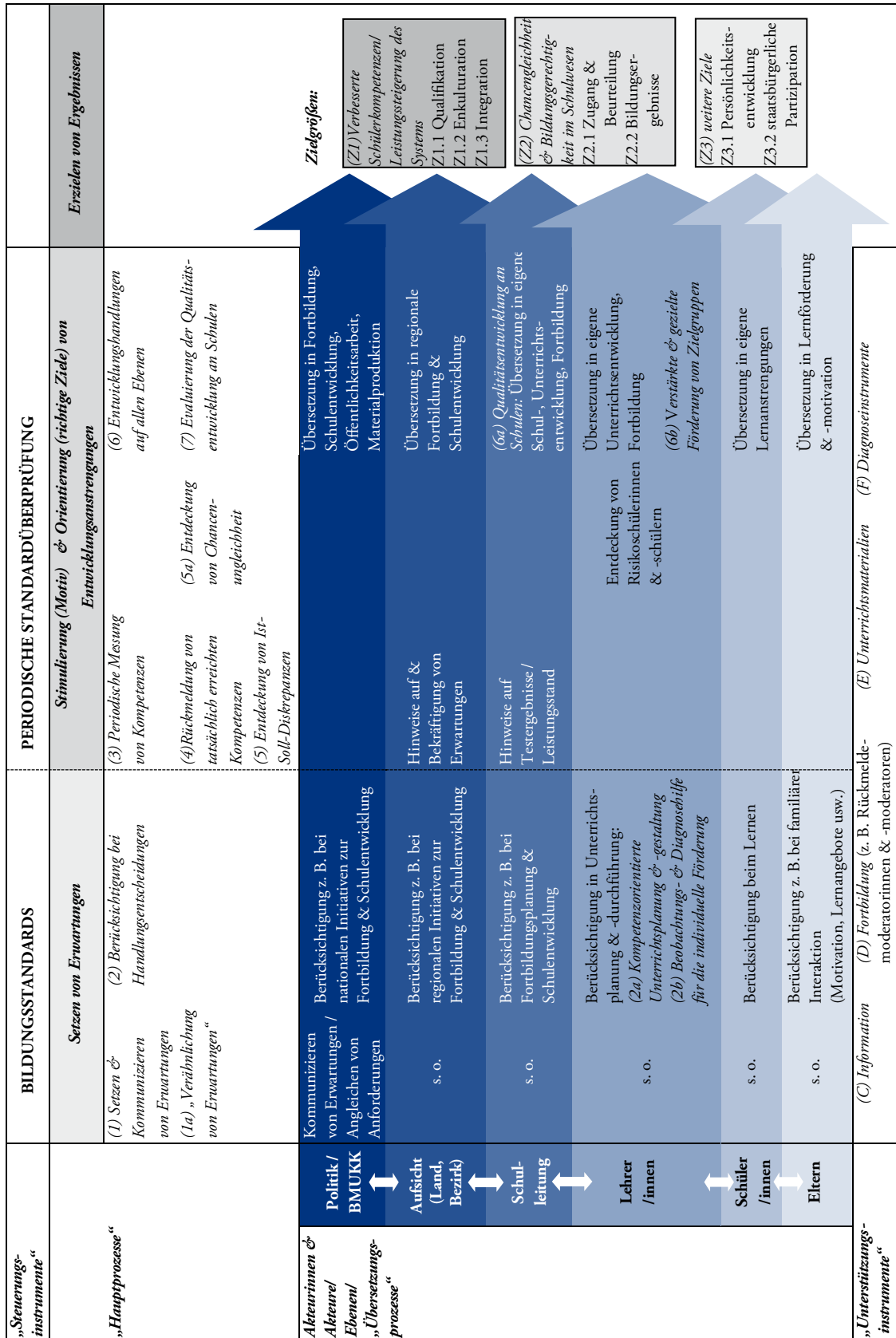
Berücksichtigung der Mehrebenenstruktur von Schule

Dabei zeigt sich, dass die *konkretesten Vorstellungen* bezüglich der Akteursebene *Lehrpersonen* bestehen, sowie dass manche der Vorstellungen pauschal an *die „Schulebene“* gerichtet sind – z. B. (6a) „Qualitätsentwicklung an Schulen“. Ein Charakteristikum von Mehrebenensystemen ist aber gerade das „Zusammenhandeln“ (Fend, 2006) von Akteurinnen und Akteuren verschiedener Ebenen. Dieses wird im Modell durch Pfeile zwischen den Akteursebenen dargestellt, wird jedoch in den Gesetzen und Verordnungen nicht explizit gemacht und offenbar der Implementierung überlassen.

Beispielhafte Darstellung des Wirkmodells

Der Gedanke des „Zusammenhandelns“ in Mehrebenensystemen soll für die Ebene der Einzelschule dargestellt werden: Schulleitungen, die durch die Öffentlichkeitsarbeit des Bundes, entsprechende Rundschreiben, aber auch durch Leiterdienstbesprechungen und andere Kommunikation vonseiten der vorgesetzten Behörde informiert werden, dass nach Bildungsstandards zu arbeiten wäre, müssten ihre Kommunikation bei Konferenzen ebenso wie bei informellen Kontakten an dieser Anforderung orientieren. Sie müssten Impulse an Fachgruppen geben und eventuell schulinterne Fortbildungen anregen und organisieren. Sie müssten sich vergewissern, dass Bildungsstandards einen angemessenen Stellenwert bei Unterrichtsplanung und -gestaltung haben sowie zu einer gewissen Angleichung der Anforderungen und Beurteilungsmaßstäbe führen. Lehrpersonen müssten an Schulleitungen Umsetzungen, Implikationen und Probleme von kompetenzorientiertem Unterricht kommunizieren und entsprechende Rahmenbedingungen begehren, z. B. Fortbildungswünsche äußern oder neue Strukturen (z. B. in Form der gemeinsamen Arbeit der Fachgruppen Deutsch und Englisch in einer Schule usw.) vorschlagen. Sie müssten Fortbildungsangebote in Anspruch nehmen oder schulinterne Fortbildungen organisieren. Auf der anderen Seite sollten Lehrpersonen – wenn ebenenübergreifendes „Zusammenhandeln“ erfolgen soll – die Schüler/innen auf Bildungsstandards und die Arbeit mit ihnen vorbereiten, ihnen helfen, ihre Lernplanung danach zu organisieren und ihnen auch entsprechende Rückmeldung geben.

Abb. 9.1: Rekonstruiertes Wirkungsmodell der Steuerungsinstrumente „Bildungsstandards und periodische Standardüberprüfung“



Anmerkung: Z1: Zielgröße 1, Z2: Zielgröße 2, Z3: Zielgröße 3.



### 3 Forschung zu Wirkungsmechanismen und Ergebnissen evidenzbasierter Steuerung

Steuerungssysteme verschiedener Länder unterscheiden sich ...

Viele industrialisierte Staaten haben in den letzten Jahren ähnliche Elemente evidenzbasierter Steuerungssysteme eingeführt oder verstärkt. Angesichts solcher Parallelität macht die Forderung Sinn, internationale Erfahrungen und Forschungsergebnisse für die eigene Entwicklung zu berücksichtigen. Auf der anderen Seite gibt es inmitten ähnlicher Globalstrategien doch nationale Unterschiede der einzelnen Politiken und der politisch-kulturellen Kontexte (vgl. Web-Dokument 9.3 und Web-Dokument 9.4). Dadurch setzen ziemlich ähnlich erscheinende Maßnahmen – z. B. die Einführung von schulübergreifenden Leistungstests – in verschiedenen Ländern unterschiedliche Prozesse in Gang, weil sie dort auf jeweils spezifische Ausgangssituationen und aktuelle Aufmerksamkeiten treffen, in die sie eingebettet werden (vgl. Ozga & Jones, 2006).

... hinsichtlich des Rechenschaftsdrucks

Die Rezeption internationaler Erfahrungen fällt leichter, wenn es Instrumente gäbe, die es erlaubten, die Nähe eines anderen Bildungssystems zum eigenen einzuschätzen (oder umgekehrt: vor typischen Inkompatibilitäten zu warnen). Ein solches Instrument existiert allerdings nicht (vgl. Ansätze bei Windzio et al., 2005). In Hinblick auf das hier diskutierte Thema besteht zumindest eine gewisse Einigkeit in der Forschergemeinde darüber, dass die Schärfe der Konsequenzen, die mit den Testergebnissen verbunden sind, einen Unterschied für die folgenden Prozesse und ihre Wirkungen macht: In sog. *Low-stakes-Systemen* sind keine oder wenig spürbare Konsequenzen mit den Leistungen in den standardbezogenen Tests verbunden. Die erreichten Ergebnisse hängen hier stark mit der intrinsischen Motivation der Schüler/innen, bei der Testung gut abzuschneiden, zusammen (Abdelfattah, 2010; Cole, Bergin & Whittaker, 2008). In *High-stakes-Systemen* ziehen Testleistungen gravierende Konsequenzen nach sich: Diese können sich einesteils auf Schüler/innen richten, indem Testleistungen einen wichtigen oder exklusiven Stellenwert beim Aufstieg in eine nächsthöhere Klasse bzw. für Abschlussprüfungen und Zertifikate haben. Konsequenzen für Schüler/innen können über Marktmechanismen indirekt den Accountability-Druck auf Lehrkräfte und Schulen weiterleiten, wenn Schulen, die besser auf Prüfungen vorbereiten, häufiger gewählt werden. Der Rechenschaftsdruck kann sich aber auch direkt auf Lehrpersonen und/oder Schulen richten, indem bestimmte Testleistungen oder jährlich steigende Anteile von Schülerinnen und Schülern, die die Mindeststandards erreichen (*Adequate Yearly Progress*), gefordert werden. Das Verfehlen solcher Leistungsziele zieht verschiedene Sanktionen nach sich, z. B. Veröffentlichung der Ergebnisse, finanzielle Einbußen, Verlust des bisherigen Rechtsstatus, Kündigung von Lehrpersonen oder Schulleitung, Schließung der Schule. Selbstverständlich lassen sich schüler- und schulbezogener Rechtfertigungsdruck auch kombinieren. In einem Mehrebenensystem können *stakes* auch unterschiedlich verteilt sein: Beispielsweise können Testleistungen für Schüler/innen keine Konsequenzen haben, für Lehrpersonen aber schon.

Die österreichische Version der Steuerungsstrategie Bildungsstandards und standardbezogene Tests ist sowohl auf institutioneller als auch auf Schülerebene eindeutig ein Low-stakes-System. Dies gilt insgesamt für die deutschsprachigen Schulsysteme, was Maier und Kuper (2012, S. 96) zur Einschätzung bringt, die angloamerikanische Forschung wäre „nicht unmittelbar übertragbar“.

Mikroprozesse in Steuerungssystemen dürfen nicht übersehen werden

Die bisher vorliegenden Forschungsergebnisse (vgl. Web-Dokument 9.5) sind uneinheitlich und erlauben nicht, dass ein System gegenüber dem anderen als global überlegen bezeichnet wird. Wir interpretieren die Uneinheitlichkeit auch als Hinweis darauf, dass wir mehr Augenmerk auf die Mikroprozesse lenken müssen, durch die Steuerungsimpulse im System verarbeitet und von Ebene zu Ebene weitergegeben („rekontextualisiert“) werden (vgl. Coburn & Turner, 2011). Es gibt offenbar nicht einige wenige „Megafaktoren“, die die Wirkungen von Schulsystemen entscheidend verändern können, sondern eine größere Zahl von individuel-

len, organisatorischen, systemischen und kulturellen Merkmalen, die in „Konfigurationen“ auf unterschiedliche Weise zusammenwirken (Fend, 2008, S. 114).

Im Folgenden wird aktuelle Forschungsliteratur in Hinblick auf die im Wirkungsmodell identifizierten Hauptprozesse der Steuerungsstrategie Bildungsstandards und standardbezogene Tests ausgewertet. Die Literaturrecherche erfolgte nach den Begriffen „Bildungsstandards“, „standardbezogene Tests“, „Setzen von Erwartungen“ und „Datenrückmeldung und -nutzung“, ihren Synonymen bzw. englischen Übersetzungen. In der Auswertung des dadurch entstandenen Literaturkorpus wurde zunächst untersucht, ob für den jeweiligen Prozess verwertbare Informationen aus österreichischen Begleituntersuchungen zur Implementierung von Bildungsstandards vorliegen. Sodann wurde Literatur über relevante Reformprozesse in deutschsprachigen Schulsystemen ausgewertet, von denen eine größere Ähnlichkeit zu österreichischen Kontextbedingungen angenommen werden konnte (vgl. Windzio et al., 2005, S. 11–16) als bei weiteren Studien aus dem englischsprachigen und skandinavischen Raum, die in einem dritten Schritt mit Schwergewicht auf Überblicksstudien berücksichtigt wurden.

### 3.1 Setzen von Erwartungen

Als ersten Hauptprozess haben wir (1) „Setzen und Kommunizieren von Erwartungen“ identifiziert. Durch die Formulierung und offizielle Propagierung von Bildungsstandards müsste Folgendes geschehen: Akteurinnen und Akteure auf allen Ebenen des Schulsystems müssten ihre Erwartungen hinsichtlich der Ergebnisse von Unterricht an den Kompetenzziele der Bildungsstandards ausrichten und damit vermutlich aneinander anpassen – (1a) „Verähnlichung von Erwartungen“ – sowie (2) bei ihren künftigen „Handlungsentscheidungen berücksichtigen“. Dies betreffe vor allem (2a) „kompetenzorientierte Unterrichtsplanung und -gestaltung“ sowie den Einsatz als (2b) „Beobachtungs- und Diagnosehilfe für die individuelle Förderung“.

In den österreichischen Begleituntersuchungen zur Implementierung der Bildungsstandards finden sich einige Hinweise zur Diskussion dieser Annahmen. Bei diesen Studien handelt es sich jeweils um Befragungen mit standardisierten Fragebögen von Lehrpersonen und Schulleiterinnen und Schulleitern, deren Schulen in den jeweiligen Jahren an der Pilotierung (bzw. bei Grillitsch, 2010, an der Baseline-Testung) der Bildungsstandards teilgenommen hatten. Die Daten der unterschiedlichen Jahre sind aufgrund der wechselnden Größe der Stichproben, unterschiedlicher Rücklaufquoten und leicht variierender Fragestellungen nicht in allen Einzelheiten vergleichbar, doch lassen sich mit der gebotenen Vorsicht einige Interpretationen für unsere Leitfragen zur Diskussion stellen:

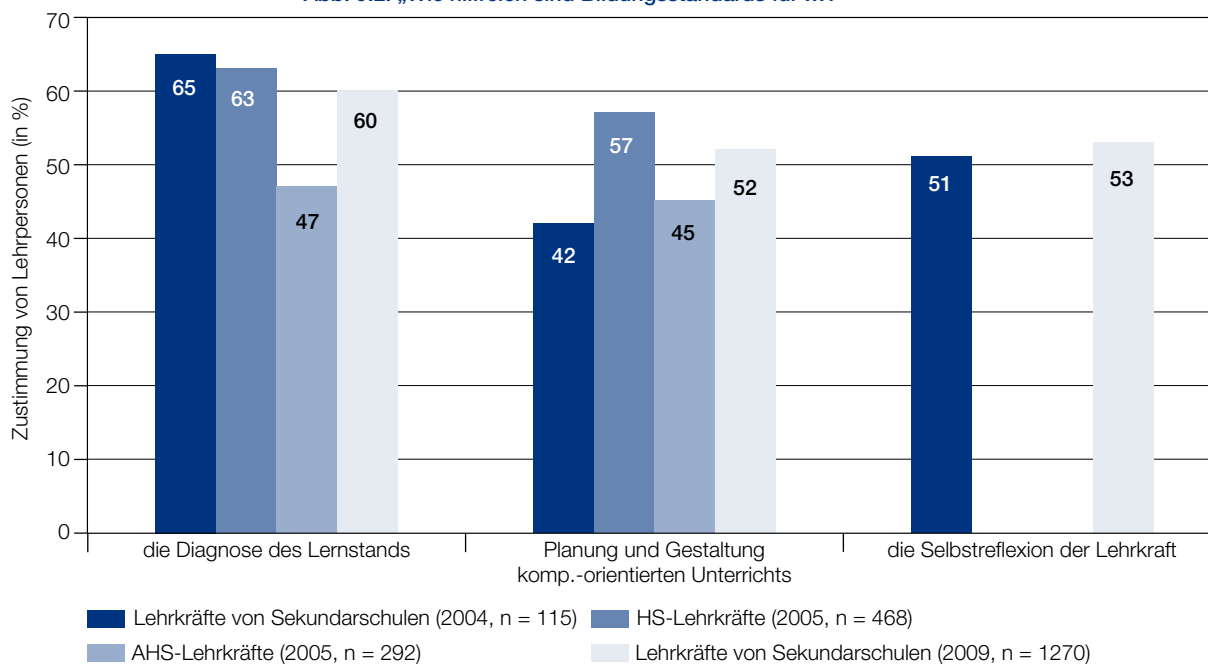
(1) Etwa die Hälfte der befragten Lehrpersonen, die in den Pilotprojekten Bildungsstandards kennen gelernt hatten, sahen durchaus hilfreiche Nutzungsmöglichkeiten von Standards für Unterrichtsplanung, Lernstandsdiagnose und Selbstreflexion der Lehrkräfte (Abbildung 9.2). Allerdings scheint die Zahl der Zweifler im Verlauf der beobachteten fünf Jahre kaum abzunehmen – mit Ausnahme der sehr bedeutsamen Nutzung für die Unterrichtsplanung. Die Zahl jener, die Bildungsstandards zusätzlichen Orientierungswert *gegenüber Lehrplänen* zubilligten, ist deutlich geringer (15–30 %; vgl. Freudenthaler & Specht, 2005, S. 50; Grillitsch, 2010, S. 81; ähnlich Böttcher & Dicke, 2008, S. 146).

(2) Der potenzielle Orientierungswert von Bildungsstandards wurde nur von einer kleinen Zahl an Lehrpersonen auch tatsächlich konsequent für die eigene Unterrichtsplanung genutzt. Die Gruppe jener, die angab, sie versuche es „hin und wieder“, ist schwer einzuschätzen (Abbildung 9.3).

Bildungsstandards werden  
durchaus als nützlich  
eingeschätzt ...

... aber wenige  
Lehrpersonen haben sie  
bislang konsequent genutzt

Abb. 9.2: „Wie hilfreich sind Bildungsstandards für ...?“



Anmerkungen: Antworten österreichischer Lehrpersonen verschiedener Stichproben.

HS: Hauptschule; AHS: Gymnasiale Langform; Sekundarschulen: HS und AHS.

Die Antwortalternative „Selbstreflexion der Lehrkraft“ ist in Freudenthaler & Specht (2006, S. 16) nicht dokumentiert.

2004 und 2005 wurden als Antwortmöglichkeiten „ja“ und „nein“ vorgegeben; die angegebenen Prozentwerte beziehen sich auf die Antwortalternative „ja“. 2009 enthielt die Skala sieben Antwortmöglichkeiten von „sehr hilfreich“ bis „gar nicht hilfreich“. Die angegebenen Prozentwerte beziehen sich jeweils auf die Summe der zustimmenden Antwortalternativen 1–3.

Quelle: Freudenthaler & Specht 2005, S. 42, 2006, S. 15–16; Grillitsch, 2010, S. 77.

(3) Dazu passt, dass die Nützlichkeit von Bildungsstandards für die eigene Unterrichtsarbeit eher skeptisch eingeschätzt wurde; ganz wenige Lehrpersonen (zwischen 10 % und 15 %) erwarteten sich eine Erleichterung ihrer Tätigkeit (vgl. Freudenthaler & Specht, 2005, S. 51; Grillitsch, 2010, S. 85).

#### Nutzung von Bildungsstandards in Deutschland

Aus der deutschsprachigen Forschung liegen darüber hinaus nur wenige Studien vor, die sich mit der Rezeption und Nutzung von Bildungsstandards für Unterrichtszwecke, unabhängig von ihrer Rolle bei standardbezogenen Tests, beschäftigen. Begleituntersuchungen zur Einführung der Bildungsstandards in deutschen Bundesländern zeigten nach Böttcher und Dicke (2008) eine Diskrepanz zwischen der allgemeinen Einsicht von Lehrkräften für den Sinn von Bildungsstandards und ihrer tatsächlichen Nutzung für den Unterricht (vgl. auch K. Maag Merki, elektronische Mitteilung, 08. 04. 2012<sup>1</sup>). Geringe Information (vgl. Pant et al., 2008) und geringe Unterstützung bei der Umsetzung von Bildungsstandards wurden hierfür als Erklärungen genannt (Böttcher & Dicke, 2008, S. 148–149).

Die qualitativen Studien von Asbrand, Heller und Zeitler (2012; vgl. auch Zeitler, Heller & Asbrand, 2012) über den unterrichtsbezogenen Umgang von Lehrkräften mit Bildungsstandards kamen zu folgenden Schlüssen: Die Implementierung von Bildungsstandards erwarte Unterrichtsentwicklung in Richtung kompetenzorientierten Lernens. Dafür stellen die bestehenden „impliziten Muster der Unterrichtsplanung“ (Asbrand et al., 2012, S. 39) sowie die

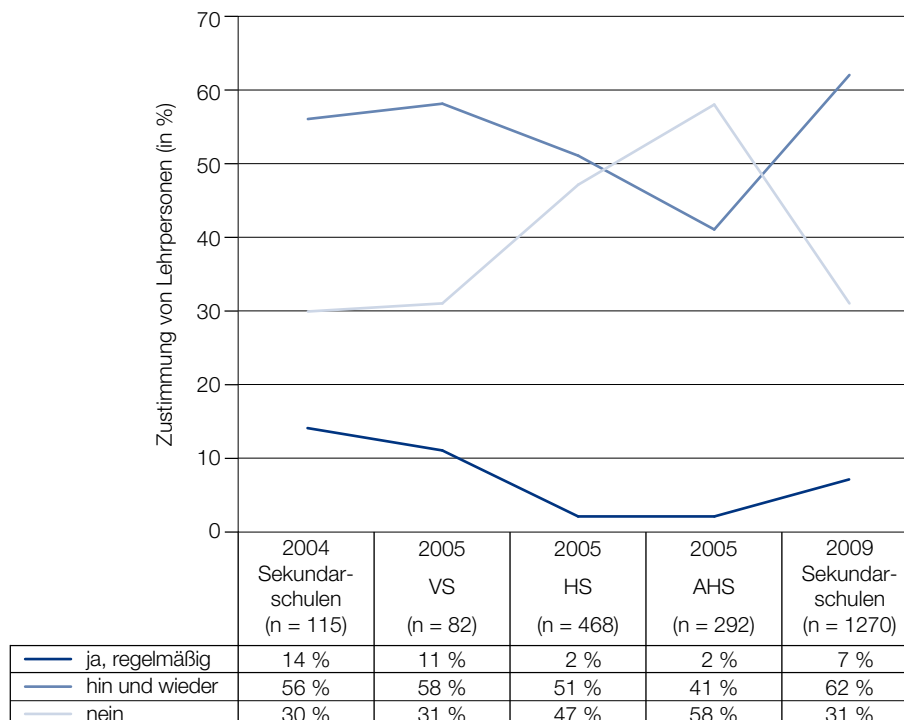
<sup>1</sup> Inhalt der Mitteilung ist die Kurzexpertise „Educational (performance) standards and external testing of student competences: Potential contributions of external measurement for achieving the quality goals of schooling“ für die Schweiz.

Vorerfahrungen im Kollegium mit Unterrichtsentwicklung bedeutsame Bedingungen dar. Dies mache eine „Differenzierung der Fortbildungsszenarien“ notwendig, die stärker die Unterschiedlichkeit der Lehrererfahrungen berücksichtige.

Das Angebot von unterstützenden Materialien, wie z. B. kompetenzorientierten Aufgaben, und von punktuellen Fortbildungsveranstaltungen über neue Analyse- und Planungsinstrumente bewirke allein keinen Perspektivenwechsel. Für viele Ausgangslagen wären „längerfristig begleitete Entwicklungsprozesse“ notwendig, bei denen „ein Ort der Reflexion über die unterrichtliche Praxis geschaffen wird. Hierfür bedarf es entsprechender zeitlicher und personeller Ressourcen und insbesondere der Begleitung durch qualifizierte und unterstützende Beraterinnen und Berater“ (Asbrand et al., 2012, S. 40).

Längerfristig begleitete Entwicklungsprozesse sind notwendig

**Abb. 9.3: Verwendung von Standards: „Verwenden Sie die Standards für Ihre Unterrichtsplanung bzw. -gestaltung?“**



Anmerkungen: Antworten österreichischer Lehrpersonen verschiedener Stichproben.

VS: Lehrkräfte der Volksschule (Primarschule); HS: Lehrkräfte der Hauptschule; AHS: Lehrkräfte der gymnasialen Langform.

Antwortkategorien 2004: „ja, intensiv und regelmäßig“, „ich versuche es hin und wieder“, „nie“. In der Untersuchung 2004 lautete die Frage „Verwenden Sie derzeit die Standards für Ihre laufende Unterrichtsplanung?“ (Freudenthaler & Specht 2005, S. 31). In der Untersuchung 2005 lautete die Frage: „Haben Sie die Standards – abgesehen von der Erprobung von Beispielaufgaben – für die Unterrichtsplanung im laufenden Schuljahr verwendet?“ (Freudenthaler & Specht 2006, S. 17). Die spezifizierte Fragestellung könnte eine Erklärung für die niedrigeren Werte in 2005 sein.

Quelle: Freudenthaler & Specht, 2005, S. 31; 2006, S. 17; Grillitsch, 2010, S. 91–92.

Die von uns geprüfte *englischsprachige Literatur* bezieht sich meist auf Länder, in denen High-stakes-Testsysteme vorherrschen. Möglicherweise ist in solchen Ländern die Wahrnehmung von Standards so eng mit der Testung und der Rückmeldung von Leistungsergebnissen verknüpft, dass wir keine Forschungsarbeiten darüber gefunden haben, die Bildungsstandards unabhängig von Tests und deren Resultaten als Instrument der Lenkung von Erwartungen und der curricularen Steuerung thematisieren.

### 3.2 Stimulierung und Orientierung von Entwicklungsanstrengungen

Wirkung von Standardmessungen und deren Rückmeldungen

Als zweiten Hauptprozess haben wir identifiziert: Durch die (3) *Periodische Messung von Kompetenzen* und die (4) *Rückmeldung von tatsächlich erreichten Kompetenzen* sollen Akteurinnen und Akteure auf verschiedenen Systemebenen in die Lage versetzt werden (5) *Ist-Soll-Diskrepanzen* auf ihrer Ebene zu entdecken (dazu gehört auch die (5a) *Entdeckung von Chancenungleichheit*). Diese Entdeckungen sollen (6) *Entwicklungshandlungen auf allen Ebenen*, insbesondere aber (6a) *Qualitätsentwicklung an Schulen* und (6b) *verstärkte und gezielte Förderung von Zielgruppen* stimulieren und auf die richtigen Entwicklungsziele orientieren. Daran soll eine (7) *Evaluierung der Qualitätsentwicklung an Schulen* angeschlossen werden. Man kann schließlich erwarten, dass durch die Rückmeldung von Ergebnissen die (1) *Ergebnis-Erwartungen* noch einmal kommuniziert und mit Bedeutung versehen werden, was zu einer weiteren (1a) *Verähnlichung von Erwartungen* beitragen könnte.

Meinungen und Einstellungen der Lehrpersonen

In den österreichischen Begleituntersuchungen zur Implementierung der Bildungsstandards wurden Lehrkräfte und Schulleitungen nach der ersten Rückmeldung der Ergebnisse der standardbezogenen Tests von Grabensberger, Freudenthaler und Specht (2008) befragt. Dabei fühlten sich die Befragten gut über die Funktionen der Testungen und Ergebnisrückmeldungen informiert und hatten „sich mehr oder weniger intensiv mit den Ergebnissen ihrer Klassen, mit deren Stärken und Schwächen auseinandergesetzt und auch über die Ursachen der Befunde nachgedacht“ (Grabensberger et al., 2008, S. 70).

Das rückgemeldete Leistungsniveau kam für mehr als die Hälfte der Lehrkräfte nicht unerwartet. Mehr als 50 % der Lehrkräfte berichteten über eine deutliche Verbesserung des allgemeinen diagnostischen Wissens (nicht jedoch bzgl. der Lerndiagnose *einzelner* Schüler/innen) und über Erkenntniszuwächse in Hinblick auf das relative Leistungsniveau sowie die Stärken und Schwächen der eigenen Klasse „im Vergleich zu anderen Klassen derselben Leistungsgruppe“. Mehr als die Hälfte der Lehrpersonen sah den Nutzen der Rückmeldungen darin, „dass sie der Lehrperson konkrete Daten und Fakten für die Selbstreflexion der eigenen Praxis liefern“ (Grabensberger et al., 2008, S. 71).

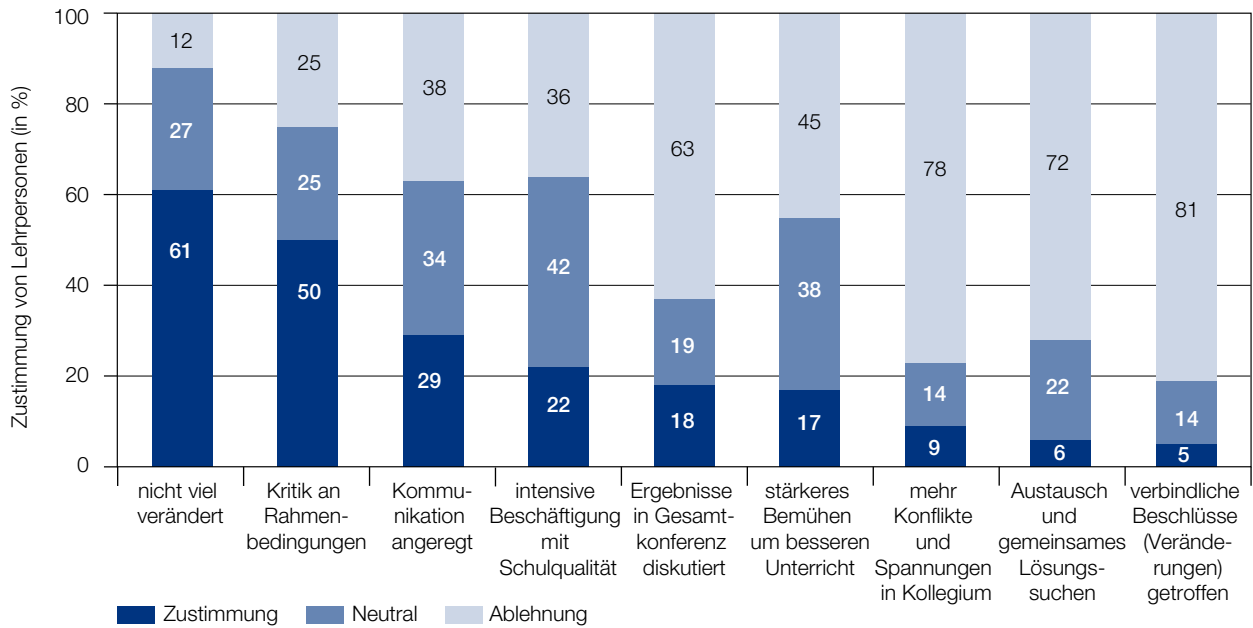
„Dagegen ist die unmittelbare Relevanz dieser Daten für die Einleitung von Veränderungen vergleichsweise deutlich geringer. Wie ‚eine bessere Planung und Gestaltung des Unterrichts im Sinne der Förderung von Grundkompetenzen‘ erfolgen könnte – darüber bezieht nur ein Drittel der Lehrpersonen Hinweise aus den Daten. Und noch einmal deutlich weniger wissen nun besser, wie sie mit Problemen der Leistungsbeurteilung umgehen könnten. Insgesamt also liegt der Nutzen für die Praktikerinnen und Praktiker eindeutig in der verbesserten Diagnose, weniger in der Handlungssteuerung“ (Grabensberger et al., 2008, S. 71).

Rückmeldungen haben bislang nur geringe Wirkung auf Unterrichts- und Schulentwicklung gezeigt

Zur Unterstützung der Datennutzung wurden sog. Rückmeldemoderatorinnen und -moderatoren eingesetzt. Die Gespräche mit diesen wurden von den Lehrkräften überwiegend als sehr nützlich eingeschätzt. „Etwas weniger ertragreich waren die moderierten Gespräche in Bezug auf die Frage, welche Konsequenzen die Schule aus den Ergebnissen ziehen soll“ (Grabensberger et al., 2008, S. 70). Abbildung 9.4 zeigt die Antworten auf eine Frage nach konkreten Reaktionen auf die Ergebnisrückmeldung an der jeweiligen Schule. „Die Wirkungen, die die Ergebnisrückmeldungen zu den Standard-Testungen auf die Schule als Organisation gezeigt haben – beispielsweise in der Form von gemeinsam getragenen Einsichten, Beschlüssen oder Schulentwicklungsmaßnahmen – halten sich in engen Grenzen“ (Grabensberger et al., 2008, S. 71). Zwar wurde über etwas vermehrte Kommunikation berichtet; diese hat aber nur in sehr wenigen Fällen dazu geführt, „dass etwa konkrete Beschlüsse oder Maßnahmen getroffen worden sind. Die Aussage mit der höchsten Zustimmungsrate bei den Lehrkräften ist, es habe sich ‚wenig verändert‘ (~ 60 %)“ (Grabensberger et al., 2008, S. 71).



**Abb. 9.4:** „Hat die Standardtestung bzw. Ergebnisrückmeldung an Ihrer Schule zu konkreten Reaktionen geführt?“



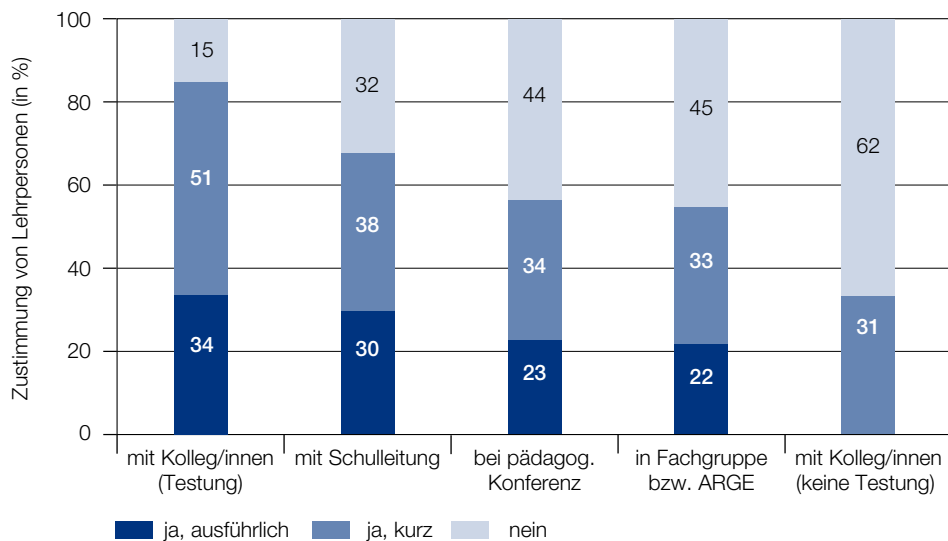
Anmerkungen: Antworten österreichischer Lehrpersonen (n = 487, Rücklaufquote 70,6 %).

Antworten auf 5-stufigen Skalen von (1) „trifft völlig zu“ bis (5) „trifft überhaupt nicht zu“.

Quelle: Grabensberger, Freudenthaler & Specht, 2008, S. 71–72.

Um den Fragebogen an Lehrkräfte „möglichst kurz zu halten“ (Amtmann, Grillitsch & Petrovic, 2011, S. 134), wurde diese Frage 2010 nicht wiederholt (ein Vergleich der Antworten der Schulleitungen aus den beiden Erhebungen ließe eine eher verminderte Anregungskraft der Ergebnisrückmeldung für Schulentwicklung, aber leicht vermehrte Beschlüsse vermuten). Stattdessen wurden die Lehrpersonen direkt gefragt, ob die rückgemeldeten Ergebnisse mit anderen Personen besprochen wurden (vgl. Abbildung 9.5) bzw. ob sie Veränderungen in ihrer Unterrichtsplanung und -gestaltung nach sich ziehen würden.

**Abb. 9.5:** „Haben Sie die Ergebnisse mit anderen Personen besprochen?“



Anmerkungen: Antworten österreichischer Lehrpersonen (n = 1205, Rücklaufquote 65,6 %)

Quelle: Amtmann, Grillitsch & Petrovic, 2001, S. 83).



Lehrpersonen tauschen sich über Ergebnisse aus

Die Resultate wurden von den Evaluatorinnen und Evaluatoren in beiden Fällen als ermutigend interpretiert: 5 % der Lehrpersonen bejahten Veränderungen des Unterrichts als Konsequenz der Rückmeldungen, 24 % antworteten mit „eher ja“. Ein Drittel der Lehrpersonen hat „ausführlich“ mit Kolleginnen und Kollegen in getesteten Fächern gesprochen, die Hälfte „kurz“. Mehr als ein Fünftel berichtete, „ausführlich“ in pädagogischen Konferenzen über die Rückmeldung gesprochen zu haben, ein weiteres Drittel „kurz“ (vgl. Abbildung 9.5). Interessanterweise gaben Lehrpersonen in der Befragung von 2010 keinen „allzu großen Unterstützungsbedarf“ (Amtmann et al., 2011, S. 80) bei der Interpretation der Ergebnisse bzw. bei der Umsetzung von Qualitätsentwicklungsmaßnahmen zu Protokoll.

Werden die Ergebnisse auch auf anderen Ebenen des Schulsystems genutzt?

Bei der Baseline-Testung 2009 wurde versucht, Informationen über die Nutzung der Rückmeldungen durch *Schüler/innen* zu erheben. Vermutlich aufgrund des langen Zeitraums zwischen Testung und Rückmeldung – die Testung erfolgte im Mai 2009, der Zugriff auf die Rückmeldungen war im Jänner 2010 möglich – haben nur 14 % der 10.843 getesteten Schüler/innen überhaupt die Ergebnisse abgerufen. Die geringe Rücklaufquote selbst dieser Gruppe machte die Befragungsergebnisse unbrauchbar (Amtmann et al., 2011, S. 141).

In der gleichen Studie wurde auch eine Interviewbefragung einer Stichprobe von 30 *Schulaufsichtsbeamtinnen und -beamten* durchgeführt. Diese verstanden die Ergebnisrückmeldung als Analyse- und Entwicklungsinstrument (Amtmann et al., 2011, S. 150). Die Schulaufsicht hatte große Hoffnungen, dass aus den Rückmeldungen Qualitätsentwicklungsmaßnahmen abgeleitet werden können, betonte aber die Notwendigkeit unterstützender Maßnahmen, wie beispielsweise einer Professionalisierung der Rückmeldemoderation und einer Institutionalisierung der Arbeit in Fachgruppen (Amtmann et al., 2011, S. 161–162). In den Interviews gaben sie an, selbst die Ergebnisse zu rezipieren, zu analysieren und das Gespräch mit den Schulleitungen zu suchen. „Größtenteils wird jedoch davon abgesehen, konkrete Handlungsmaßnahmen auf Basis der Resultate abzuleiten, da sie dies im Zuständigkeitsbereich der Schulleiter/innen sehen bzw. die Ergebnisse der Baseline-Testung nicht überbewerten wollen“ (Amtmann et al., 2011, S. 159).

E

Datennutzung in der deutschsprachigen Forschung

Aus der deutschsprachigen Bildungsforschung liegen zur Frage der Datennutzung für Schul- und Unterrichtsentwicklung zahlreiche Untersuchungen vor: Wir greifen hier auf einen aktualisierten Forschungsüberblick von Altrichter (2010; vgl. auch Bonsen & Gathen, 2004; Maier, 2006; Maier & Kuper, 2012; Schulze, 2012) zurück.

(1) Lehrkräfte und Schulleitungen berichten häufig über offene bzw. „moderat positive“ (Maier, 2006, S. 24) *Einstellungen* gegenüber externer Leistungsmessung, doch finden sich auch viele skeptische Stimmen (vgl. Bonsen, Büchter & Peek, 2006, S. 140 ff.; Schrader & Helmke, 2003). Wenn das Instrument in transparenter Weise eingeführt wird und bedrohliche Elemente, wie Veröffentlichung, vermieden werden, lassen sich offenbar aversive Reaktionen vermeiden (vgl. Maier 2006, S. 26). Ein relevanter Teil der Lehrpersonen äußert eine generelle Bereitschaft, *sich mit den Ergebnissen von Lernstandserhebungen auseinanderzusetzen und Konsequenzen für den Unterricht zu ziehen* (z. B. Bonsen et al., 2006, S. 143; Kühle, 2010; Peek, 2004). Die allgemeine Akzeptanz von Leistungsrückmeldung muss dabei nicht mit der *Einschätzung ihrer Unterrichtsrelevanz* einhergehen (vgl. Maier, 2008, S. 106–107).

(2) Auch wenn Lehrpersonen der Datenrückmeldung offen gegenüberstehen und sich damit auseinandersetzen, scheint es ihnen nicht leicht zu fallen, die Informationen über die Leistungsstände ihrer Schüler/innen zu verarbeiten und in konstruktive Entwicklungsinterventionen umzusetzen. Der Großteil der vorliegenden Studien hat enttäuschend *wenig Nutzung von Datenrückmeldung für Unterrichtsentwicklung* erhoben (z. B. Kühle & Peek, 2007; Peek, 2004; Steffens, 2009). Wenn Entwicklungsmaßnahmen durch Datenfeedback angestoßen werden, dann betreffen diese *selten eine weitergehende Innovation des Unterrichts*. Viel eher werden bisherige Inhalte und Aufgaben wiederholt, Impulse aus den Messungen (z. B. bestimmte Aufgabenformate) direkt übernommen oder geringfügige Veränderungen in der In-

teraktion vorgenommen (von nichttestförmigen Begleitmaterialien – vgl. Blum, Drücke-Noe, Hartung & Köller, 2006; Lersch, 2010 – wird hier eine Verbesserung erhofft). Einmal gezogene Konsequenzen werden in der Regel nicht von einer Klasse auf eine andere übertragen (vgl. Groß Ophoff, Koch, Helmke & Hosenfeld, 2007, S. 423).

(3) Lehrpersonen scheinen eher bereit, die Rückmeldungen aus externen Leistungsüberprüfungen für *Aufgaben der Leistungsbeurteilung und Lerndiagnose* zu verwenden, als *didaktische Weiterentwicklung* aus ihnen abzuleiten (vgl. Maier, 2009). Damit bleiben sie eng an der „Testförmigkeit“ der verwendeten Instrumente: Leistungstests werden im Schulalltag für Förder- und Selektionsdiagnose eingesetzt.

(4) Den vorliegenden Studien sind nur wenige Hinweise über die *Stimulierung von Schulentwicklung und andere ebenenübergreifende Wirkungen* der Ergebnisrückmeldung von standardbezogenen Tests zu entnehmen.

Auch aus der englischsprachigen Forschung liegen zu diesen Fragen zahlreiche Untersuchungen vor. Pedulla (2003) zeigte, dass in jenen Staaten der USA mit High-stakes-Systemen die Ergebnisse externer Leistungstests am häufigsten verwendet wurden; je größer die Konsequenzen, die mit den Ergebnissen der Standardüberprüfungen verbunden waren, desto stärker die Wirkungen auf Lehrerhandeln, und zwar:

Nutzung der  
Ergebnisse externer  
Leistungsmessungen in  
anderen Ländern

(1) *Verwendung auf Schulebene*: Obwohl Lehrer/innen im Allgemeinen die Verwendung von Testergebnissen für die Rechenschaftslegung als wenig angemessen empfanden (am wenigsten für Lehrer-, am ehesten für Schülerbeurteilung), fanden sich in Staaten mit High-stakes-Systemen etwas höhere Akzeptanzwerte. In den Low-stakes-Staaten gaben mehr Lehrer/innen (etwa 10 %) an, nicht mit den Ergebnisrückmeldungen an Schulen und in Bezirken vertraut zu sein.

(2) *Professionelle Weiterentwicklung*: Lehrer/innen in Low-stakes-Staaten erklärten häufiger, keine Ansprechperson für Standardüberprüfungen an ihrer Schule zu haben, und erachteten das Fortbildungsangebot zu externen Leistungsüberprüfungen als weniger adäquat.

(3) *Verwendung im Unterricht*: 40 % der Lehrer/innen in High-stakes-Systemen gaben an, dass die Testergebnisse ihrer Schüler/innen ihren täglichen Unterricht beeinflussen (im Gegensatz zu 10 % in Low-stakes-Staaten der USA). Die Leistungsresultate beeinflussen Entscheidungen im Klassenzimmer, wie Unterrichtsplanung (60 %) und die Auswahl von Unterrichtsmaterialien (50 %). Lehrer/innen in High-stakes-Staaten verwenden die Ergebnisse häufiger für Schülerfeedback, Bewertung des Lernfortschritts, Zuteilung zu Lerngruppen und Festlegung von Noten.

(4) *Einfluss auf Schulentwicklung*: Eine Übersicht zur Verwendung von Schuleffektivitätsdaten für Schulentwicklung findet sich bei Kelly und Downey (2010). Damit ganze Schulen externe Leistungsdaten für ihre Weiterentwicklung nutzen können, braucht es nach Wikeley, Stoll und Lodge (2002) gezielte externe Unterstützung bei der Interpretation und Ableitung von Maßnahmen, aber auch günstige interne Bedingungen (wie allgemeines Innovationsklima, Bewusstsein über Veränderungsnotwendigkeit, Leitungspersonen mit Visionen, die sich mit Entwicklungsoptionen, die sich aus Rückmeldeergebnissen ergeben, verbinden lassen, sowie Lehrpersonen mit Evaluations- und Unterrichtsentwicklungsexpertise).

Einige Studien stellen die Nutzbarkeit von High-stakes-Tests für Schulentwicklung grundsätzlich in Frage, weil sie deren Validität als Indikatoren für Schulqualität bezweifeln: Ingram, Louis und Schroeder (2004) zeigten an Schulen, die von den Schulbehörden als Best-Practice-Beispiele angesehen wurden, dass sich deren professionelle Maßstäbe deutlich von den Testinformationen unterschieden und eher auf informellem und fallbezogenem Wissen basierten. Mintrop und Trujillo (2007) fanden keinen systematischen Zusammen-

hang zwischen üblichen Kriterien der Schul- und Organisationsqualität und den Ergebnissen der externen Leistungsüberprüfung in Kalifornien.

Nichtintendierte  
Effekte von externen  
Leistungsmessungen

Neben den erwünschten Auswirkungen führen High-stakes-Testungen auch zu *nichtintendierten Effekten*, die meist auf strategischem Verhalten von Schulen, Lehrerinnen/Lehrern und Schülerinnen/Schülern basieren, wodurch die Testergebnisse an Aussagekraft und Validität verlieren (vgl. Koretz, 2008). Dabei werden folgende zentrale Probleme genannt (vgl. Franklin & Snow-Gerono, 2007; Koretz, 2008; Maag Merki, 2010; Maier & Kuper, 2012, S. 90–91; Morris, 2011; Nichols & Berliner, 2005; vgl. ausführlicher in Web-Dokument 9.7):

(1) *Konzentration auf testmethodische und -strategische Kompetenzen*: Mehr Unterrichtszeit wird verwendet, um Teststrategien (z. B. Umgang mit Multiple-Choice-Aufgaben) zu vermitteln und typische Aufgabenformate bzw. Aufgabenbeispiele aus vergangenen Tests zu üben. Lehrgelenkte Unterrichtsmethoden und testähnliche Aufgabenformate werden bevorzugt. Dies führt zu einer Beschränkung der Lerninhalte sowie zu vermehrtem Auswendiglernen und weniger problemlösendem und verstehendem Lernen.

(2) Eine *Verengung des Curriculums* findet sowohl durch Umschichtung von Unterrichtszeit zwischen Fächern (z. B. von Musik oder Geschichte zu getesteten Fächern wie Mathematik oder Englisch) als auch innerhalb von Fächern (verstärkte Konzentration auf testrelevante Fähigkeiten und Inhalte, z. B. mehr Lesen auch im Geschichteunterricht) statt.

(3) Der *Ausschluss von Schülerinnen und Schülern aus Lerngelegenheiten* kann in Form der Zuweisung zu nicht getesteten Schülergruppen (z. B. Schülerinnen und Schülern mit Beeinträchtigungen) bzw. zu anderen Schulen, durch Nichtteilnahme von schlechteren Schülerinnen und Schülern oder Abgang von der Schule erfolgen (vgl. Education, Audiovisual & Culture Executive Agency [EACEA], 2009, für die Niederlande). Schüler/innen im Grenzbereich, die die Accountability-relevanten Leistungsziele noch erreichen können, erhalten stärkere Aufmerksamkeit bei der Prüfungsvorbereitung als „sichere“ und „hoffnungslose“ Fälle. Die Verwendung von Wachstumsmodellen, bei denen der Leistungszuwachs und nicht die gezeigte Leistung von vorrangiger Bedeutung ist, soll diesen Effekt verringern.

(4) *Schummeln auf Lehrerseite* erstreckt sich von ‚kleinen‘ Hilfestellungen während der Testdurchführung, wie Hinweis auf ausgelassene Items, Betonung beim Vorlesen der Instruktion etc. bis zur Manipulation der Testdaten in Form von nachträglichen Ergänzungen oder Änderungen.

### 3.3 Ergebnisse

Führen Bildungsstandards und externe Leistungsüberprüfung zu den erhofften Wirkungen in Hinblick auf Zielgröße 1 „Verbesserte Schülerkompetenzen/Leistungssteigerung des Systems“ und auf Zielgröße 2 „Chancengleichheit und Bildungsgerechtigkeit im Schulwesen“? Uns sind aus den deutschsprachigen Schulsystemen keine empirischen Untersuchungen bekannt, die beanspruchen würden, fundierte Aussagen zu dieser Frage zu machen. Werfen wir daher einen Blick auf andere Schulsysteme.

Befunde zu  
Leistungssteigerungen in  
High-stakes-Systemen

Dabei zeigt sich, dass Wirkungen von Testregimes auf Leistungen fast nur in High-stakes-Kontexten untersucht wurden (vgl. Morris, 2011, S. 35). Für US-amerikanische Schulen haben Figlio und Loeb (2011, S. 402–410) eine Zusammenfassung von Studien zur Wirkung von Standardtestungen auf die Schülerleistungen vorgelegt und Verbesserungen der Mathematikleistungen sowie, wenn auch weniger eindeutig, für Lesen festgestellt. Goldhaber und Hannaway (2004) konnten nach einer Standardüberprüfung Verbesserungen der Schülerleistungen in einer schlechtabschneidenden Schule in Florida erkennen. Diese waren jedoch nicht direkt und eindeutig auf die Nutzung des Datenfeedbacks durch Lehrpersonen und Schulleitung zurückzuführen, da die Schule gleichzeitig zusätzliche finanzielle Unterstützung

erhielt, die in Klassengröße, neue Materialien, zusätzliches Personal sowie Nachmittagsbetreuungsprogramme investiert wurde (vgl. auch Stecher, 2002, S. 86).

Nach Amrein und Berliner (2003) haben High-stakes-Staaten eine insgesamt rückläufige Leistungsentwicklung, während Rosenshine (2003) moderate Leistungseffekte für die 4. Schulstufe (*National Assessment of Educational Progress* [NAEP]-4) in Mathematik und substanzielle Effekte bei NAEP-8-Mathematik und NAEP-4-Lesen feststellte. Maier (2010, S. 142) kommt nach einer Analyse aktueller amerikanischer Arbeiten zum Schluss, dass bei High-stakes-Testkonsequenzen für Schüler/innen und Schulen die Leistungen, aber auch die soziale Selektivität steigen. Attali, Neeman und Schlosser (2011) fanden differenzielle Leistungseffekte von Low- vs. High-stakes-Kontexten nach Geschlecht und ethnischen Gruppen. Maier und Kuper (2012, S. 91–92) vermuten, dass eine „Kombination aus hohem Rechenschaftsdruck auf Schulebene und zentralen Leistungsmessungen [...] zu Schülerleistungssteigerungen führen“ kann.

Franklin und Snow-Gerono (2007) beschreiben eine Reihe von nichtintendierten Nebenefekten der Einführung von neuen standardisierten High-stake-Tests auf Lehrer/innen, wie zunehmender Stress und Druck sowie abnehmende Arbeitszufriedenheit (vgl. Herman, Dreyfus & Golan, 1990). High-stakes-Testungen wurden auch mit einer Reihe von körperlichen Begleiterscheinungen bei den Schülerinnen und Schülern zur Zeit der Testungen in Zusammenhang gebracht, z. B. mit mehr Übelkeit, Kopfweg, Bauchweh, Weinen etc. (Smith, Edelsky, Draper, Rottenberg & Cherland, 1987, S. 284). Nach Figlio und Loeb (2011) führen High-stakes-Messungen auch zu Veränderungen auf dem Arbeitsmarkt: Lehrer/innen meiden schwächere Schulen oder versuchen diese möglichst bald wieder zu verlassen.

Welche *Auswirkungen auf Chancengleichheit und Bildungsgerechtigkeit* sind zu erwarten? Die Einführung externer standardisierter Tests bietet offensichtlich Möglichkeiten, *Chancenungleichheit des Typs 1*, also die Nichtübereinstimmung von gemessenen Leistungen und vergebenen Noten und Zertifikaten zu entdecken. So werden in den Niederlanden nationale Leistungsüberprüfungen mit den Ergebnissen der jeweiligen schulstandortbezogenen Leistungsmessungen verglichen, um Chancenungleichheit zwischen einzelnen Schulen zu entdecken und sicherzustellen, dass Schulen Abschlüsse nicht zu „leicht“ gestalten (M. Ehren, elektronische Mitteilung, 23. 02. 2012<sup>2</sup>).

Verbesserte  
Chancengleichheit  
bei Noten und  
Zertifikatsvergabe?

E

Kuusela (2006) berichtet, wie Daten aus externen Leistungsüberprüfungen zur Aufdeckung von Chancenungleichheit im finnischen Bildungssystem genutzt werden: So konnte gezeigt werden, dass der Zusammenhang von in den Einzelschulen vergebenen Noten mit den nationalen Leistungsüberprüfungen in verschiedenen Schulen sehr unterschiedlich war. Engere Zusammenhänge gingen meist mit einem hohen allgemeinen Leistungsniveau der Schule einher. Außerdem wurden systematische Ungleichheiten in der Beurteilung sichtbar: In Mathematik müssen Jungen mehr Leistungen in den standardisierten Tests für die gleiche Note erbringen als Mädchen, während das Verhältnis in Muttersprache und Literatur umgekehrt ist (vgl. auch Hannover & Kessels, 2011, sowie die österreichischen Untersuchungen zu den Medizin-Auswahlverfahren von Spiel, Schober & Litzenberger, 2008).

Eine wirksame Maßnahme zur Verringerung von Segregationseffekten scheint die Überarbeitung der Schweizer Übertrittsverfahren in die Sekundarstufe II zu sein (vgl. Maag Merki, 2012). Baeriswyl, Wandeler, Trautwein und Oswald (2006) konnten zeigen, dass durch ein neues Übergangsmodell von der Grundschule zur Sekundarstufe I, welches sich unter anderem auf standardisierte Leistungsbeurteilungen stützt, der Einfluss des sozioökonomischen Hintergrunds nur mehr sehr schwach auf die Übertrittsempfehlung der Lehrpersonen wirkt. Muller und Schiller (2000) fanden anhand von US-Daten heraus, dass häufiges Low-stakes-

2 Inhalt der Mitteilung ist die Kurzexpertise „Short expertise on ‚Educational (performance) standards and external testing of student competences: Potential contributions of external measurement for achieving the quality goals of schooling‘ – the Netherlands“.

Testen zwar zu keiner Leistungssteigerung führt, aber die soziale Selektivität des Lehrerurteils senkt.

Verbesserte  
Chancengleichheit von  
Schülerinnen/Schülern  
unterschiedlicher Herkunft?

*Chancengleichheit des Typs 2* strebt an, aus Herkunft entstandene Benachteiligungen im Verlaufe der Schulzeit zu kompensieren. Hanushek und Raymond (2005) haben gezeigt, dass der durchschnittliche Leistungszuwachs für die Gruppe der weißen Schüler/innen in den USA stärker ausgeprägt ist als für Schüler/innen aus ethnischen Minderheiten. Figlio und Loeb (2011) verweisen sowohl auf nordamerikanische Studien, in denen Accountability-Systeme einen starken Leistungszuwachs für Schüler/innen aus Minderheiten mit sich bringen, als auch solche, bei denen kein Einfluss externer Standardüberprüfungen auf die Verringerung von Leistungsunterschieden zwischen ethnischen Gruppen zu finden war (vgl. auch Tymms & Merrell, 2007). Auch die Frage, ob schwache Schulen von Rechenschaftslegung stärker profitieren als bessere, lässt sich laut Figlio und Loeb (2011) anhand der bisherigen Untersuchungen nicht eindeutig beantworten.

Information über  
Schülerleistungen führt  
nicht per se zu verbesserter  
Chancengleichheit

Wir lesen die bisher vorliegenden Untersuchungen so, dass die Verbesserung der Informationsbasis über Schülerleistungen per se kaum zu Prozessen führen wird, die die Aufmerksamkeit für benachteiligte Gruppen und deren gezielte Förderung entscheidend verstärken werden. Auch gezielte High-stakes-Incentives innerhalb der Logik der Evidenzbasierung können mit Schwierigkeiten verbunden sein: Das US-amerikanische *No Child Left Behind*-Programm wollte durch die Forderung nach „Adequate Yearly Progress“ (also nach einer steigenden Anzahl von Schülerinnen und Schülern, die Mindeststandards erreichen) für benachteiligte soziale Gruppen deren spezifische Förderung sicherstellen, erhöhte dadurch aber auch die Hürden für Schulen, die mit verschiedenen sozialen Benachteiligungen umzugehen hatten (vgl. Mintrop & Sunderman, 2012, S. 11–12).

Die positiven Beispiele (vgl. oben; Qualität in Multikulturellen Schulen [QUIMS], 2012; vgl. Web-Dokument 9.8) zeigen aber, dass Informationen aus externen Leistungsüberprüfungen nützlich sein können, wenn sie gezielt analysiert und als Ansatzpunkte für *spezielle und zusätzliche Förderungsprogramme* verwendet werden.

## 4 Zusammenfassung und Schlussfolgerungen

### 4.1 Zusammenfassende Einschätzung

Die Leitfragestellung dieses Beitrags – *Welche Beiträge können Bildungsstandards und externe Überprüfungen von Schülerkompetenzen zur Erreichung verschiedener Qualitätsziele der Schule leisten?* – kann für die gegenwärtige Ausformung der evidenzbasierten Steuerung in Österreich nicht abschließend beantwortet werden, weil aus Österreich (und sehr ähnlichen Schulsystemen) keine entsprechenden Untersuchungen vorliegen und der Übertragbarkeit von angloamerikanischen und skandinavischen Ergebnissen enge Grenzen gesetzt sind. Wohl aber kann man begründete Schlüsse aus der internationalen Forschungslage ziehen. Diese lauten zusammengefasst:

Einschätzung der  
Nützlichkeit ≠ tatsächliche  
Umsetzung

In Hinblick auf den ersten Prozess „*Setzen von Erwartungen*“ lässt sich, nach Untersuchung von Vermittlungsprozessen im Schulsystem bezüglich Wirkungen evidenzbasierter Steuerung (vgl. Abschnitt 3.1), festhalten: Nach den Selbstberichten der in den österreichischen Pilotierungen beteiligten Lehrpersonen (also einer speziellen, Neuerungen vermutlich eher aufgeschlossenen Gruppe) können Bildungsstandards nützlich für die Diagnose des Lernstands sowie für die Planung und Gestaltung eines kompetenzorientierten Unterrichts sein; real werden sie bisher nur von einer sehr kleinen Gruppe regelmäßig für kompetenzorientierte Unterrichtsplanung und -gestaltung verwendet. Wir fanden keine Hinweise in der empirischen Forschung, dass Bildungsstandards bisher deutliche Impulse zu einer „Verähnlichung

von Erwartungen“ geleistet haben sowie als „Beobachtungs- und Diagnosehilfe für die individuelle Förderung“ verwendet wurden.

Aus Fallstudien (Aiglsdorfer & Aigner, 2005; Hölzl & Rixinger, 2007) haben wir – nicht notwendigerweise verallgemeinerbare – Hinweise darauf, dass die Vernetzung der Entwicklungsprozesse über die verschiedenen Ebenen hinweg Probleme bereiten könnte: Sie deuten darauf hin, dass Bildungsstandards bei den Schülerinnen und Schülern bislang nur in der sehr eingeschränkten Bedeutung von Tests und Aufgabenbeispielen angekommen sind. Sie haben aber nicht gelernt, diese Kompetenzbeschreibungen für eigenverantwortliche Lernplanung und -gestaltung zu nutzen. Auch die Schulleiter/innen in diesen Studien hatten zu diesem Zeitpunkt Bildungsstandards nicht als Führungsinstrument und als Orientierungspunkt für die Weiterentwicklung von Schule und Unterricht verstanden, sondern delegierten die Erledigung der im Laufe der Standarderprobung anfallenden Aufgaben an die Fachlehrpersonen der betroffenen Klassen.

Probleme bei der Vernetzung verschiedener Ebenen

Für den zweiten Hauptprozess „Stimulierung und Orientierung von Entwicklungsanstrengungen“ gilt nach unserer Einschätzung: Unter den Bedingungen deutschsprachiger Schulsysteme kann die periodische Messung von Kompetenzen und deren Rückmeldung nach und nach eine allgemeine positive Bewertung bei einer Mehrheit von Lehrpersonen finden. Durch die ErgebnISRückmeldung werden offenbar auch in moderatem Ausmaß neue Informationen – „Ist-Soll-Diskrepanzen“ – entdeckt. Allerdings gibt es noch wenige Daten darüber, um welche speziellen inhaltlichen Bereiche es sich dabei handelt und welche handlungsstimulierend sind; die „Entdeckung von Chancenungleichheit“ scheint jedenfalls selten explizit auf.

Ist-Soll-Diskrepanzen werden entdeckt

Die bisherigen Untersuchungen zeigen wenige und – wenn überhaupt – wenig innovative „Entwicklungshandlungen“ als Reaktion auf ErgebnISRückmeldungen, was sowohl für Unterrichtsentwicklung, für Qualitätsentwicklung auf Schulebene als auch für die Selbstevaluierung von Schulen gilt. Für den zentralen Wirkungsmechanismus der deutschsprachigen Accountability-Modelle, die Stimulierung ergebnisorientierter Weiterentwicklung des Unterrichts durch Rückmeldung von Leistungsdaten, fehlen also empirische Belege (vgl. Maier, 2010, S. 143).

Ergebnisse führen zu wenigen Entwicklungshandlungen

Im Fokus der aggregierten Rückmeldung im österreichischen Modell (Lehrpersonen erhalten Klassen-, Schulleitungen Schuldaten usw.) stehen nicht die Individualdiagnose und differenzielle Entwicklungsanstrengungen für die gezielte Förderung von Zielgruppen. Für die Stärkung der Diagnosekompetenz der Lehrer/innen sind begleitende Maßnahmen in der Lehrerbildung sowie das Angebot von Instrumenten der informellen Kompetenzmessung vorgesehen. Auch gibt es bisher keine Hinweise darauf, dass Lehrkräfte und andere Akteure den Ergebnis-Erwartungen der Bildungsstandards mehr Bedeutung als früher zuschreiben oder dass sich ihre Ergebnis-Erwartungen „verähnlicht“ hätten.

Mehr Handlungsdruck in High-stakes-Steuerungssystemen

Die bisherigen Ergebnisse der deutschsprachigen Studien zur Datennutzung sind insofern überraschend, als viele Kritiker/innen der Bildungsstandards-Politik – oft in Extrapolation internationaler Forschungsergebnisse – sogar „übersteuernde“ Wirkungen befürchtet hatten (z. B. *Teaching-to-the-test*). Die High-stakes-Struktur der angloamerikanischen Bildungssysteme scheint aber bedeutend mehr Rezeptions- und Umsteuerungsdruck zu produzieren als die meist administrativ geprägte Implementation der neuen Instrumente in den deutschsprachigen Bildungssystemen. Studien aus High-stakes-Kontexten zeigen nämlich, dass externe Leistungsrückmeldungen sehr wohl die Lern- und Entwicklungsanstrengungen von Schülerinnen und Schülern, Lehrpersonen, Schulen und Bezirken/Teilschulsystemen beeinflussen können. Gleichzeitig sind diese Konsequenzen häufig mit unerwünschten Nebenwirkungen verbunden: Täuschungsstrategien nehmen zu; die Orientierung auf quantitative Steigerung der gemessenen Leistungen kann mit einer Verringerung der Breite und qualitativen Tiefe der Lernerfahrungen oder gar mit einem Ausschluss von schwachen Schülerinnen und Schülern von Lern- und Prüfungsgelegenheiten einhergehen.

Begleitmaßnahmen haben  
ungenutztes Potenzial

Studien aus Österreich und vergleichbaren Schulsystemen zeigen nach unserer Einschätzung, dass die Steuerungsimpulse, die von einem evidenzbasierten Steuerungssystem ausgehen, von Lehrpersonen und anderen Akteurinnen und Akteuren im Schulwesen bisher nicht automatisch und regelmäßig aufgegriffen und zu einer Weiterentwicklung von Unterricht und Schule im Sinne der proklamierten Ziele verwendet werden. Das heißt aber noch nicht, dass dies auch längerfristig unmöglich wäre. Die Untersuchungen von Grillitsch (2010, S. 99) sowie Asbrand et al. (2012) lassen im Gegenteil vermuten, dass sich durch gezielte Unterstützung der Lehrpersonen und Schulen die Nutzungshäufigkeit und -qualität von Bildungsstandards im Unterricht steigern ließen.

Dies ist Aufgabe einer Reihe von *Begleitmaßnahmen zur Implementierung von Bildungsstandards*, wie Information, Fortbildung sowie die Entwicklung und Verbreitung von Unterrichtsmaterialien und standardbezogenen Diagnoseinstrumenten (vgl. BIFIE 2012b). Es ist plausibel, dass diese die Prozesse „Setzen von Erwartungen“ und „Stimulierung und Orientierung von Entwicklungsanstrengungen“ fördern *können* (bspw. können die zusätzlichen Diagnoseinstrumente [IKM] einige der Nachteile der aggregierten Form der Rückmeldung der standardbezogenen Leistungsergebnisse entschärfen, die Entdeckung von sozialen Disparitäten und die Umsetzung individualisierter Lernförderung ermöglichen). Ob sie dies auch tatsächlich tun, sollte – auch weil hier massiv investiert werden müsste, um eine Breitenwirkung zu erzielen – begleitend und von unabhängigen Forschungsgruppen untersucht werden. Bisher wurden aus Österreich keine Studien über die aktuellen Implementierungsmaßnahmen und ihre differenzielle Wirkung veröffentlicht. Die von uns analysierten älteren Studien deuten jedenfalls an, dass das Potenzial dieser Maßnahmen noch nicht realisiert ist (vgl. z. B. Grabensberger et al., 2008, S. 70 über Rückmeldemoderatorinnen und -moderatoren).

Bildungsstandards  
als Aufforderung zum  
Perspektivenwechsel

Insgesamt entsteht der Eindruck, dass in den deutschsprachigen Schulsystemen gelegentlich überzogene Erwartungen an diese Steuerungsinstrumente vorliegen (van Ackeren, 2007, S. 207; ähnlich Kuper, 2005, S. 101–102). Sie stellen sehr wohl Innovationsimpulse dar, doch deren Realisierung ist offenbar schwieriger als vermutet: Bildungsstandards wären nicht die propagierte Neuerung, wenn sie nicht Umstellungen von einem Gutteil der Lehrkräfte erforderten. Bildungsstandards sind Aufforderungen zu einem „Perspektivenwechsel“ im Unterricht – zu einem stärker kompetenzorientierten Unterricht, der gekennzeichnet ist „von offenen Aufgaben, Lebensweltbezug sowie kompetenzorientierter Diagnose von Schülerleistungen als Ausgangspunkt für individuelle Förderung“ (vgl. Asbrand et al., 2012, S. 32–33).

Ob die spezielle Version von Bildungsstandards und externer Leistungsüberprüfungen, wie sie in Österreich gewählt wurde, zu *verbesserten Schülerkompetenzen* führt (vgl. Abschnitt 3.3), kann derzeit nicht durch empirische Untersuchungen belegt werden, und muss – da empirische Belege für die vorgelagerten Prozesse der Erwartungsbildung und der Stimulierung von Entwicklungshandlungen fehlen – gegenwärtig bezweifelt werden. Untersuchungen aus Schulsystemen, die mit den Ergebnissen externer Prüfungen gravierende Konsequenzen (High-stakes) für Schüler/innen, Lehrer/innen und Schulen verbinden, zeigen ebenfalls keine eindeutige Befundlage, jedoch finden sich dort auch Fälle von Leistungsverbesserungen; allerdings werden auch häufig nichtintendierte Nebeneffekte berichtet.

Zugangs- und  
Beurteilungsgerechtigkeit

In Hinblick auf das Ziel *erhöhter Chancengleichheit* finden sich international einige Beispiele dafür, dass Chancengleichheit vom Typ 1 – Zugangs- und Beurteilungsgerechtigkeit – durch externe Prüfungen verbessert werden kann, weil diese offenbar weniger stark durch den sozioökonomischen Hintergrund der Prüflinge beeinflusst sind als Lehrerurteile. Für ihre Verwendung für Übertrittssituationen gibt es positive Beispiele; häufigere Verwendung von (auch Low-stakes-)Tests kann zu Anpassungsprozessen bei Lehrerurteilen führen, die einen sozialen Bias verringern. Da die Ergebnisse externer standardbezogener Tests in Österreich keinen Einfluss auf die Leistungsbeurteilung haben dürfen, wird dieser Wirkungsmechanis-



mus ausgeschlossen; die aggregierte Aufbereitung von Daten wiederum erschwert deren Nutzung im Sinn von Chancengleichheit des Typs 2.

Diese Prozess- oder Entwicklungsgerechtigkeit ist dann gegeben, wenn das Schulwesen durch Herkunft entstandene Benachteiligungen nicht weiter verstärkt, sondern so weit ausgleicht, dass unterschiedliche Herkunftsgruppen in gleicher Weise aus dem schulischen Angebot Nutzen ziehen können. Dieser Typ von Chancengleichheit kann nach unserer Einschätzung nicht direkt durch ein „evidenzbasiertes Steuerungssystem“ (z. B. durch die Vorgabe von Leistungs- oder Fortschrittszielen; vgl. Mintrop & Sunderman, 2012, S. 11) erreicht werden. Wohl aber können – wie internationale Beispiele zeigen (vgl. Abschnitt 3.3 und Web-Dokument 9.8) – Informationen aus externen Prüfungen dazu beitragen, benachteiligte Gruppen in spezifischen Schulen und Regionen zu entdecken. Solche Effekte sind aber keine automatische Begleiterscheinung der Implementierung eines evidenzbasierten Steuerungsmodells, sondern treten offenbar nur ein, wenn zentraler oder regionaler politischer oder administrativer Wille besteht, die Ergebnisse gezielt zu analysieren, Bewusstseinsbildung im Schulsystem und in der Öffentlichkeit über die entdeckten Diskrepanzen zu leisten, gegenläufige strukturelle Signale des Bildungssystems zu verändern (z. B. Verhinderung des Abschiebens von Schülerinnen und Schülern; vgl. Schleicher, 2012) und die Ergebnisse als Ansatzpunkte für *spezielle und zusätzliche Förderungsprogramme* zu verwenden.

Weitere Maßnahmen werden benötigt, um Benachteiligungen aufgrund von Herkunft zu verringern

## 4.2 Handlungsoptionen

Angesichts der bislang enttäuschenden Wirkungen und der „recht indifferenten“ Haltung eines erheblichen Teils der Lehrpersonen und Schulleitungen gegenüber den Ergebnissen externer Prüfungen hat Maier kürzlich zwei denkbare „Lösungsvarianten“ diskutiert: „[Die erste ist:] Vergleichsarbeiten reduzieren und die testbasierten Schulreformen einfach ins Leere laufen lassen. Die zweite Variante wäre, die Bedeutung externer Leistungsmessung auf Schulebene deutlich zu erhöhen, z. B. durch Veröffentlichung schulbezogener Daten oder eine konsequente Einbindung der Vergleichsarbeitsergebnisse in neue Formen der externen Schulevaluation bzw. Schulinspektion“ (Maier, 2010, S. 128).

Wenn wir auch meinen, dass es mehr als zwei Ansätze für die Weiterentwicklung gibt, so teilen wir doch Maiers Einschätzung, dass es Zeit für grundsätzliche Entscheidungen ist. Wenn man aus Gründen gesellschaftlicher Transparenz, bislang erfolgter Investitionen und einer engeren Orientierung der Schulen an Systemzielen die evidenzbasierte Steuerungsstrategie weiterführen will, dann braucht es die Klärung einiger grundlegender Probleme und sodann entschiedene Schritte, um die positiven Potenziale der Steuerungsstrategie zu realisieren (vgl. Österreichische Forschungsgemeinschaft, 2009). Der Zeitpunkt ist nach unserer Einschätzung günstig, da durch die Novellierung von § 18 des Bundes-Schulaufsichtsgesetzes im Frühjahr 2011 eine neue rechtliche Basis für Qualitätsbemühungen im österreichischen Schulwesen geschaffen wurde. Die Novelle tangiert an vielen Stellen Bildungsstandards und externe Prüfungen; ihre Umsetzung würde die Bearbeitung einiger der hier angesprochenen Probleme erlauben. Zuvor wird es aber notwendig sein, die Erwartungen an die neuen Steuerungsinstrumente zu dämpfen und die angestrebten Funktionen realistisch zu reformulieren.

Günstiger Zeitpunkt für grundsätzliche Entscheidungen

(1) Die Aufwertung der Frage nach den tatsächlichen Ergebnissen schulischer Bildungsprozesse durch Konzepte evidenzbasierter Steuerung war unserer Einschätzung nach historisch richtig. Ihre technische Umsetzung durch externe Leistungsüberprüfung muss sich notwendigerweise auf einige Bereiche schulischen Lernens konzentrieren und steht daher in *Gefahr, zu einer Verengung der Zieldimensionen von Schule beizutragen*: Erstens im Sinn der Bevorzugung von Leistungen in wenigen inhaltlichen Domänen zuungunsten von anderen Lernbereichen; zweitens im Sinn der Bevorzugung von qualifikatorischen Aufgaben der Schule zuungunsten von Aufgaben im Bereich der Enkulturation und gesellschaftlichen Integration (vgl. Abschnitt 2.3.2).

Zieldimensionen der Schule

E

Chancengleichheit und Bildungsgerechtigkeit sind keine automatische Folge der neuen Steuerungsinstrumente

Gleichsam stellvertretend für diese weiteren Zielbereiche wurde in diesem Kapitel das schulische Qualitätsziel *Förderung von Bildungsgerechtigkeit und Chancengleichheit* geprüft. Dabei zeigte sich, dass externe Leistungsüberprüfungen zwar neue Chancen auf die Entdeckung von sozialen Disparitäten und Ungerechtigkeiten bieten, dass Verbesserungen aber nicht als automatische Folge der Durchsetzung einer evidenzbasierten Steuerung zu erwarten sind, sondern gezielte Anstrengungen bei Analyse, Bewusstseinsbildung und Entwicklung erfordern (vgl. Abschnitt 3.3).

Aufmerksamkeit für Prozesse

(2) Durch die Fokussierung auf Output-Fragen wird in den letzten Jahren „zu wenig nach den *Prozessen* [Hervorhebung v. Verf.] gefragt [...], die zu bestimmten Wirkungen führen“ (van Ackeren, 2003, S. 28). Dies gilt sowohl für die vielfältigen Vermittlungsprozesse, die in einem Mehrebenensystem für das Wirksamwerden von Steuerungsimpulsen notwendig sind, als auch für die Idee der Unterrichtsentwicklung durch Rückmeldung von Leistungsergebnissen (vgl. Resnick et al., 2007, S. 155). Ergebnismaße zeigen Probleme beim Lerngewinn nach dem Ende von Lerneinheiten auf, geben aber keine Hinweise, wie diese im Zuge von Lernprozessen entstanden sind. Die Entwicklung des neuen *Nationalen Qualitätsrahmens* (vgl. § 18 Bundes-Schulaufsichtsgesetz) bietet die Gelegenheit, *prozessbezogene Rückmeldungen* verfügbar zu machen und ihnen einen Platz im Rahmen der Qualitätsbemühungen im Gesamtsystem zuzuweisen. Im Rahmen des Unterstützungssystems *Schulqualität Allgemeinbildung* (SQA) zeichnen sich Umsetzungsschritte in diese Richtung ab (z. B. Materialien zur Pädagogischen Diagnostik; prozessbezogenes Rückmeldeinstrument „SQA-online“; vgl. SQA, 2012).

Systemqualität als Aufgabe aller Ebenen im Schulsystem

(3) Systemqualität wird nicht nur in der Schule, sondern an verschiedenen Stellen in einem Mehrebenensystem gemacht. Die gesetzlichen Grundlagen in Österreich fokussieren die Verantwortung stark auf Lehrpersonen und Einzelschulen. Dadurch entsteht die Gefahr, dass die *Beiträge anderer Akteurinnen und Akteure und die Bearbeitung der Schnittstellen* aus dem Blick geraten. Schulinspektorinnen und -inspektoren setzen beispielsweise große Hoffnungen auf die neuen Steuerungsinstrumente, haben dabei aber unterschiedliche Vorstellungen von ihrer Rolle (vgl. Amtmann et al., 2011). Vielen Lehrpersonen scheint unklar zu sein, was Schüler/innen über den Umgang mit Bildungsstandards wissen müssen. Die Evaluation der Beiträge von Administration und Unterstützungssystemen zur Qualität im Gesamtsystem hat bisher weit weniger geistige und materielle Ressourcen verschlungen als jene der Schülerleistungen. Sowohl die internen Ziele eines evidenzbasierten Steuerungssystems (Systemkenntnis für rationale Entwicklungsentscheidungen) als auch das Prinzip der Fairness verlangen, die Verantwortung unterschiedlicher Akteurinnen und Akteure klarzumachen und in angemessener Weise zu evaluieren. Die im § 18 Bundes-Schulaufsichtsgesetz festgeschriebenen Zielvereinbarungen bieten hier einen neuen Ansatzpunkt. Für die Bilanz- und Zielvereinbarungsgespräche zwischen BMUKK, Schulaufsicht und Schulleitung finden gegenwärtig Qualifizierungsprozesse statt (vgl. SQA, 2012).

Akzeptanz der Akteurinnen und Akteure gewinnen

(4) Wie kann im Schulsystem eine stärkere Verbindlichkeit – eine Bindung an die Systemziele und den Entwicklungsauftrag, der mit den neuen Steuerungsinstrumenten verbunden ist – hergestellt werden? Maier (2010, S. 128), für den Lehrpersonen und Schulleitungen zu „indifferent“ reagiert haben, sieht eine Lösungsvariante in einer stärkeren Bindung an Reformziele durch Konsequenzen und Sanktionen. Dies ist in einem gewissen Sinn konsequent, als die Entwicklung evidenzbasierter Steuerungsstrategien in den deutschsprachigen Ländern in einer Situation verstärkten öffentlichen Drucks aufgrund des PISA-Schocks ausgeformt wurde. Sie trägt einen Zweifel an der professionellen Selbstentwicklungsfähigkeit der Lehrpersonen und Schulen in sich und stärkt den Akteur Staat/Administration gegenüber der Profession (vgl. auch Altrichter, 2010, S. 242; Mintrop & Sunderman, 2012; Schimank, in Druck).

Auf der anderen Seite mehrten sich die Appelle an Lehrpersonen, sich „professionell“ mit den Bildungsstandards und den Ergebnissrückmeldungen aus den externen Leistungsüberprüfungen auseinanderzusetzen: Die neuen Steuerungsinstrumente können eben nicht durch die

Akteurinnen und Akteure „hindurch“-steuern, sondern benötigen deren Akzeptanz, Verständnis und Engagement, um ihre Steuerungswirkung zu entfalten. Dieser Beitrag könnte bei einer weiteren Schwächung der Lehrprofession durch Verstärkung zentralstaatlicher Sanktionen gefährdet sein (vgl. dazu den eher „professionalistisch“ auftretenden Dachverband Schweizer Lehrerinnen und Lehrer [LCH], 2012, S. 1).

Nach unserer Einschätzung sollten vermehrte Anstrengungen unternommen werden, die *Profession dafür zu gewinnen, die Kernelemente der neuen Steuerungspolitik in ihr Verständnis professioneller Kompetenzen und Ziele aufzunehmen*: Kompetenzorientiert zu unterrichten, die Schüler/innen bestmöglich in ihrer Lern- und Leistungsentwicklung zu fördern und dabei auf Kriterien der Chancengleichheit Rücksicht zu nehmen, kann nicht so weit von jenen pädagogischen Idealen entfernt sein, die viele Lehrkräfte leiten (vgl. O'Day, 2004), die aber offenbar gegenüber den vielen „technischen“ und „instrumentellen“ Aspekten des neuen Steuerungsmodus in den Hintergrund getreten sind. Dass Bildungsstandards und Ergebnisrückmeldung nicht der pädagogischen Arbeit äußerlich bleiben (als etwas, das „die da oben“ verordnet haben, den eigenen Unterricht vielleicht stört, aber nicht berührt), sondern dass sie in „bedeutungsvolle Arbeitspraktiken eingebettet“ (vgl. Resnick et al., 2007, S. 179) werden, ist eine Voraussetzung für eine echte Qualitätsentwicklung. Die Konzeption von SQA ist mit diesem Gedanken vereinbar; es wird zu beobachten sein, ob diese Initiative langfristig tatsächlich zu einer Harmonisierung der verschiedenen Aspekte von Qualitätsentwicklung beiträgt. Auch die stärkere Berücksichtigung von Weiterentwicklung und Karriere im aktuell verhandelten Dienstrecht könnte zielorientierte Weiterqualifikation von Lehrpersonen unterstützen.

Neue Steuerungspolitik und  
Lehrerprofessionalität

(5) Die strukturelle Entscheidung, Standards am Ende der Primar- und Sekundar-I-Bildungsstufen im Frühjahr zu messen und im Herbst des darauffolgenden Schuljahrs nach Auflösung der getesteten Klassen aggregiert rückzumelden, kann – angesichts der bisher bekannten Umgangsformen von Lehrpersonen mit Datenrückmeldungen (vgl. z. B. Maier, 2009) – zur *Konsequenzenlosigkeit der Monitoringergebnisse für Schul- und Unterrichtsentwicklung* beitragen (dies gilt nicht für Entwicklungen auf der Makrobene). Alternativen sind, den Anspruch auf Unterrichts- und Schulentwicklung durch Ergebnisrückmeldung fallen zu lassen oder Zeitpunkt und Modus der standardbezogenen Messung und Rückmeldung zu revidieren. Vonseiten des Gesetzgebers wird gegenwärtig eine weitere Alternative erprobt: Die im Nationalen Qualitätsrahmen vorgesehenen Instrumente „Bilanz- und Zielvereinbarungsgespräche“ und „standortbezogene Entwicklungspläne“ sollen einer eventuellen Konsequenzenlosigkeit von Monitoringergebnissen für Schul- und Unterrichtsentwicklung entgegenwirken.

Strukturelle Probleme

(6) Die Erweiterung der Möglichkeiten schulinterner Partizipation und Mitentscheidung für lokale Akteurinnen und Akteure (vgl. Altrichter & Heinrich, 2005) ist im österreichischen Schulsystem gut begründbar; *lokale Rechenschaftslegung* könnte eine Alternative und/oder Ergänzung zur Verstärkung zentralstaatlicher Kontrolle sein. Kogan (1996; ähnlich Mintrop & Sunderman, 2012) unterscheidet drei soziale Akteurinnen und Akteure, in deren Kräftebalance Bildungssysteme gesteuert werden. Deutschsprachige Schulsysteme wurden traditionell durch eine „Doppelsteuerung“ von Staat und Profession geregelt (vgl. Brüsemeister, 2004). Ein potenzieller dritter Akteur, die Eltern und die lokale Gemeinde, haben dagegen wenig Einfluss. Schulautonomie und Profilierung haben ihre Einflussmöglichkeiten etwas gestärkt, allerdings eher über leichteren Zugang zu verschiedenen Schulen und differenzielle Schulwahlentscheidungen, mithin über (schulexterne) Marktmechanismen (vgl. Altrichter, Heinrich & Soukup-Altrichter, 2011).

Lokale  
Rechenschaftslegung

(7) Notwendig ist weiters die *Unterstützung von Schulleitungen und Lehrpersonen bei der Nutzung von Bildungsstandards für Schulentwicklung und kompetenzorientierte Unterrichtsplanung und -gestaltung* (vgl. Grillitsch, 2010, S. 99–100). Aus der Untersuchung von Asbrand et al. (2012, S. 40) lässt sich auf die Notwendigkeit einer Differenzierung der Fortbildungsszenarien

Differenzierte  
Unterstützungs- und  
Fortbildungskonzepte

schließen, die die Unterschiedlichkeit der Lehrererfahrungen stärker berücksichtigen. Für wesentliche Teile der Zielgruppe genügt punktuelle Fortbildung nicht; Unterstützung bei Unterrichtsentwicklung braucht längerfristige Konzepte, die Fortbildung und Entwicklungsarbeit kombinieren (vgl. Altrichter & Geisler, 2012; Lipowsky, 2004). Lehrpersonen und Schulen, die schon Entwicklungsarbeit geleistet haben, können solche Prozesse als „lebende Beispiele der Implementation“ (van Ackeren, 2003, S. 287) wesentlich unterstützen. Diese Aufgabe fällt in der gegenwärtigen Systemlogik den Pädagogischen Hochschulen zu, die dabei – aufgrund der Breite der Entwicklung – vor großen organisatorischen und kompetenzbezogenen Herausforderungen stehen. Mit der Initiative „Entwicklungsberatung in Schulen“ (EBIS) und dem Ausbau der Kapazitäten für fachbezogene Unterrichtsentwicklungsberatung soll diese Entwicklung unterstützt werden (vgl. SQA, 2012).

#### Weitere Begleitmaßnahmen

(8) Diese Implementierungsunterstützung erfordert eine Reihe von *Begleitmaßnahmen*, wie z. B. die Überarbeitung der Lehrpläne mit dem Fokus auf verstärkte Kompetenzorientierung, die Erstellung von kompetenzorientierten Unterrichtsmaterialien und Lernaufgaben (B. Asbrand, elektronische Mitteilung, 13. 06. 2012<sup>3</sup>). Eine besondere Bedeutung haben diagnostische Tests (informelle Kompetenzmessungen; wie auch von Lehrkräften gefordert, vgl. LCH, 2012, S. 1), die Lehrpersonen bei der individuelleren Betreuung von Lernwegen unterstützen sollen. Bei der Implementation muss aber dafür Sorge getragen werden, dass dieses neue Instrument nicht in der Logik eines traditionell selektionsorientierten Systems missverstanden wird (vgl. Maier, 2009) und sich für Schüler/innen nicht in einer Vervielfachung von Prüfungssituationen niederschlägt. Solche Begleitmaßnahmen sind im österreichischen Implementierungskonzept vorgesehen (vgl. BIFIE, 2012b; SQA, 2012); die begleitende Evaluation ihrer tatsächlichen Auswirkungen sollte durchgeführt und veröffentlicht werden.

#### Unterstützung von Unterrichtsentwicklung

(9) Der Einsatz von Rückmeldemoderatorinnen und -moderatoren als Hilfe zur Interpretation von Ergebnissrückmeldungen hat in der Evaluation der österreichischen Pilotphase eine positive Einschätzung erfahren, doch erscheint es langfristig nicht sinnvoll, unterschiedlich spezialisierte Berater/innen im Feld nebeneinander agieren zu lassen. Daher wäre eine *umfassendere Unterstützung von Unterrichtsentwicklung* angebracht: „Für einen erfolgreichen Schritt ‚von der Evaluation zur Innovation‘ [könnte es wichtig sein], dass diese Personen über ihre Übersetzerrolle hinaus auch stärker die Funktion der Entwicklungsberatung bzw. Begleitung übernehmen“ (Grabensberger et al., 2008, S. 75–76). Mit der Initiative „Entwicklungsberatung in Schulen“ werden Schritte in diese Richtung gesetzt (vgl. SQA, 2012).

#### Bildungsstandards in der Lehrerbildung

(10) *Qualifikation für kompetenzorientierten Unterricht, Unterrichtsentwicklung und Umgang mit Datenfeedback* müssen in der *Lehrerbildung* eine größere Rolle als bisher spielen. Zweifellos muss man sich für die Realisierung der Potenziale von Bildungsstandards und Ergebnissrückmeldung auf einen längerfristigen und schrittweisen Prozess der Veränderung von Einstellungen und Praktiken sowie von Kultur- und Strukturmerkmalen in Einzelschulen und im Schulsystem einrichten. Dieser kann fünf bis acht Jahre dauern (Lucyshyn, 2006, S. 9) oder gar „bis zu zehn Jahre“ (Oelkers & Reusser, 2008; vgl. für Wirkungszeiträume bei Schulreformen Borman, Hewes, Overman & Brown, 2003). Dies ist aber keine Zeit des Abwartens: Zwischenziele müssen formuliert und evaluiert werden, wo notwendig, Umsteuerungsinitiativen gesetzt werden.

### 4.3 Forschungsoptionen

Die bisherigen Erfahrungen mit der Implementation von Bildungsstandards von Ergebnissrückmeldungen sind vom BIFIE bzw. von seinen Vorgängerinstitutionen kontinuierlich begleitend evaluiert wurden, sodass in Österreich für die vergangenen Jahre eine vergleichs-

3 Inhalt der Mitteilung ist die Kurzexpertise „Educational (performance) standards and external testing of student competences: Potential contributions of external measurement for achieving the quality goals of schooling“ für Deutschland.

weise gute Datenbasis über die Rezeption und Bewertung der neuen Instrumente durch Lehrpersonen und Schulleitungen vorliegt.

(1) Diese Studien sollten *kontinuierlich weitergeführt* werden. Ihr Design sollte erlauben, Entwicklungen der Wahrnehmungen und Handlungen von wesentlichen Zielgruppen zu beobachten und Trendaussagen abzuleiten. Eine Öffnung des Projekts oder der Datenbasis für externe Wissenschaftler/innen würde die Chance auf unterschiedliche Zugänge und frühere Entdeckung von Implementations- und Entwicklungsproblemen erhöhen.

Kontinuierliche  
Begleituntersuchungen

(2) Dennoch kann sich die Erforschung der neuen evidenzbasierten Steuerungsinstrumente nicht nur auf diesen Ansatz beschränken. Insbesondere die *Umsetzung des Nationalen Qualitätsrahmens und seine Beziehung zu den bisherigen Steuerungsinstrumenten* erfordern neue Forschungsanstrengungen.

Nationaler  
Qualitätsrahmen

(3) Bisher erfolgte die Erforschung der neuen Steuerungsinstrumente häufig mittels der Erhebung von Selbstberichten durch Fragebögen. Dieser Ansatz unterbietet längerfristig die selbstformulierten Ansprüche einer outputorientierten Schulreform und sollte mit *Informationen über Ergebnismerkmale* verknüpft werden.

Verknüpfung mit  
Ergebnismerkmalen

(4) Zum besseren Verständnis der Verarbeitung und Nutzung von Bildungsstandards und Ergebnisrückmeldungen sowie zur Erarbeitung von Möglichkeiten zur Unterstützung dieser Prozesse sind (oft mit qualitativen Methoden arbeitende) Studien sinnvoll, die solche *Verarbeitungsprozesse in ihren besonderen Schul- und Gruppensettings beobachten, sowie Interventionsstudien*, die spezifische Prozesse und Rahmenbedingungen erproben (z. B. Zeitler et al., 2012).

Studien zur Nutzung und  
Umsetzung

(5) Die bisherige Forschung konzentriert sich oft stark auf die Perspektive von Lehrpersonen und Schulleitungen und *vernachlässigt andere Akteurinnen und Akteure*. Auch der *Mehrebenencharakter von Innovationen im Schulwesen* wird selten in den Blick genommen. Dies ist eine empirisch anspruchsvolle Aufgabe, umso mehr müssen hier Weiterentwicklungen in Angriff genommen werden.

Schule =  
Mehrebenensystem

(6) Dazu bedarf es *umfassenderer theoretischer Modelle*, die der angenommenen Relevanz von Bildungsstandards und Datenfeedback für mehrere Ebenen sowie ihrer Kommunikation über mehrere Ebenen hinweg gerecht werden. Frühe Studien in diesem Feld wirken oft wie anlassbezogene Rezipientenbefragungen; in der Zwischenzeit können höhere Ansprüche an die „professions-, organisations- und/oder institutionentheoretische Rahmung empirischer Projekte“ (Maier & Kuper, 2012, S. 96) gestellt werden.

Umfassende theoretische  
Modelle

(7) Da evidenzbasierte Steuerungsssettings gute Chancen bieten, *Bildungsungerechtigkeiten zu entdecken* (diese aber nicht ohne weitere Interventionen verhindern), sollten darauf bezogene Analysen und Vorschläge für Interventionsprogramme regelmäßig in Auftrag gegeben und veröffentlicht werden (z. B. als Nationaler Bericht zur Chancengerechtigkeit im Bildungswesen).

Entdeckung von  
Bildungsungerechtigkeiten

(8) *International vergleichende Forschung* könnte wichtige Hinweise auf spezifische Kulturmerkmale des österreichischen Schulsystems, die die Arbeit mit Bildungsstandards und Ergebnisrückmeldungen behindern oder fördern, geben.

International  
vergleichende Forschung

(9) Zur Stimulierung und Koordination der eben genannten Forschungsanstrengungen könnte ein *nationales Schwerpunktprogramm* – vergleichbar dem Forschungsschwerpunkt Steuerung im Bildungssystem (SteBis), der vom deutschen Bundesministerium für Bildung und Forschung gefördert wird – ausgeschrieben werden.

Nationaler  
Forschungsschwerpunkt

## Verzeichnis der Web-Dokumente:

- Web-Dok. 9.1: *Mehrebenencharakter des Schulsystems* (zu Abschnitt 2.3 des Kapitels). Verfügbar unter <https://www.bifie.at/buch/1915/9/2/1>
- Web-Dok. 9.2: *Bedeutungen von Chancengerechtigkeit*. Verfügbar unter <https://www.bifie.at/buch/1915/9/2/2>
- Web-Dok. 9.3: *Konzeption der internationalen Kurzexpertisen*. Verfügbar unter <https://www.bifie.at/buch/1915/9/3/1>
- Web-Dok. 9.4: *Standards und standardbezogene Prüfungen in verschiedenen Steuerungsregimes*. Verfügbar unter <https://www.bifie.at/buch/1915/9/3/2>
- Web-Dok. 9.5: *Globale Vergleiche zwischen high-stake- oder low-stake-Systemen*. Verfügbar unter <https://www.bifie.at/buch/1915/9/3/3>
- Web-Dok. 9.6: *Die Implementation von Bildungsstandards in der Studie von Asbrand et al. (2012)*. Verfügbar unter <https://www.bifie.at/buch/1915/9/3/4>
- Web-Dok. 9.7: *Unerwünschte Wirkungen von Leistungsvorgaben und externen Leistungsüberprüfungen unter high-stake Bedingungen*. Verfügbar unter <https://www.bifie.at/buch/1915/9/3/5>
- Web-Dok. 9.8: *Nutzung von Leistungsdaten für spezielle Förderungsprogramme*. Verfügbar unter <https://www.bifie.at/buch/1915/9/3/6>

## Literatur

Abdelfattah, F. (2010). The relationship between motivation and achievement in low-stakes – examinations. *Social Behavior and Personality*, 38 (2), 159–167.

Ackeren, I. van. (2003). National curriculum tests‘ und ‚final examinations‘ – Zur nationalen Test- und Prüfungspraxis im englischen Schulsystem. *PÄD Forum*, 31/22 (5), 275–278.

Ackeren, I. van. (BMBF Hrsg.). (2007). *Nutzung großflächiger Tests für die Schulentwicklung: Exemplarische Analyse der Erfahrungen aus England, Frankreich und den Niederlanden* (Bildungsforschung, Bd. 3). Bonn: BMBF. Zugriff am 12. 10. 2012 unter [http://www.bmbf.de/pub/nutzung\\_grossflaechiger\\_tests\\_fd\\_schulentwicklung.pdf](http://www.bmbf.de/pub/nutzung_grossflaechiger_tests_fd_schulentwicklung.pdf)

Aiglsdorfer, B. & Aigner, M. (2005). *Implementierung nationaler Bildungsstandards in Österreich*. Unveröffentlichte Diplomarbeit, Johannes Kepler Universität.

Altrichter, H. (2008). Veränderungen der Systemsteuerung im Schulwesen durch die Implementation einer Politik der Bildungsstandards. In T. Brüsemeister & K.-D. Eubel (Hrsg.), *Evaluation, Wissen und Nichtwissen* (S. 75–115). Wiesbaden: VS Verlag für Sozialwissenschaften.

Altrichter, H. (2010). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 219–254). Wiesbaden: VS Verlag für Sozialwissenschaften.

Altrichter, H. & Geisler, B. (2012). Unterrichtsentwicklung in Zeiten der Systemreform. Erscheint in *Schweizerische Zeitschrift für Bildungswissenschaften*.

Altrichter, H. & Heinrich, M. (2005). Schulprofilierung und Transformation schulischer Governance. In X. Büeler, A. Buholzer & M. Roos (Hrsg.), *Schulen mit Profil* (S. 125–140). Innsbruck: StudienVerlag.

Altrichter, H., Heinrich, M. & Soukup-Altrichter, K. (2011). *Schulentwicklung durch Schulprofilierung?* Wiesbaden: Verlag für Sozialwissenschaften.

Altrichter, H. & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 15–39). Wiesbaden: VS Verlag für Sozialwissenschaften.

Altrichter, H. & Posch, P. (Hrsg.). (1999). *Wege zur Schulqualität*. Innsbruck: StudienVerlag.

Amrein, A. L. & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, 60 (5), 32–38.

Amtmann, E., Grillitsch, M. & Petrovic, A. (2011). *Bildungsstandards in Österreich. Die Ergebnisrückmeldung im ersten Praxistest. Das Rückmeldedesign zur Baseline-Testung (8. Schulstufe) aus der Sicht der Adressaten* (BIFIE-Report 7/2011). Leykam: Graz. [Verfügbar am 11. 10. 2012 unter <https://www.bifie.at/node/1397>].

Asbrand, B., Heller, N. & Zeitler, S. (2012). Die Arbeit mit Bildungsstandards in Fachkonferenzen. Ergebnisse aus der Evaluation des KMK-Projekts for.mat. *Die Deutsche Schule*, 104 (1), 31–43.

Attali, Y., Neeman, Z. & Schlosser, A. (2011). *Rise to the challenge or not give a damn: Differential performance in high vs. low stakes tests* (Discussion paper series, Forschungsinstitut zur Zukunft der Arbeit, Nr. 5693). Zugriff am 04. 10. 2012 unter <http://hdl.handle.net/10419/51974>

Bacher, J. (2005). Bildungsungleichheit und Bildungsbenachteiligung in Österreich. *SWS Rundschau* 45 (1), 37–63.

Baeriswyl, F., Wandeler, C., Trautwein, U. & Oswald, K. (2006). Leistungstest, Offenheit von Bildungsgängen und obligatorische Beratung der Eltern. *Zeitschrift für Erziehungswissenschaft*, 9 (3), 373–392.

Balog, A. & Cyba, E. (2004). Erklärung sozialer Sachverhalte durch Mechanismen. In M. Gabriel (Hrsg.), *Paradigmen der akteurszentrierten Soziologie* (S. 21–42). Wiesbaden: VS Verlag für Sozialwissenschaften.

Beer, R. (2007). *Bildungsstandards: Einstellungen von Lehrerinnen und Lehrern*. Wien: LIT.

Belfield, C. R. & Levin, H. N. (2009). Market reforms in education. In G. Sykes, B. Schneider, D. N. Plank & T. G. Ford (Hrsg.), *Handbook of Education Policy Research* (S. 513–527). New York: Routledge.

Bishop, J. H. (1999). Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review*, 6, 349–398.

Bishop, J. H. & Wößmann, L. (2004). Institutional Effects in a Simple Model of Educational Production. *Education Economics*, 12 (1), 17–38.

Blömeke, S., Herzig, B. & Tulodziecki, G. (Hrsg.). (2005). *Gestaltung von Schule. Eine Einführung in die Schulpädagogik*. Bad Heilbrunn: Klinkhardt.

Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.). (2006). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen*. Berlin: Cornelsen Scriptor.

Bonsen, M., Büchter, A. & Peek, R. (2006). Datengestützte Schul- und Unterrichtsentwicklung. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung* (Bd. 14, S. 125–148). Weinheim: Juventa.

Bonsen, M. & Gathen, J.v.d. (2004). Schulentwicklung und Testdaten. Die innerschulische Verarbeitung von Leistungsrückmeldungen. In H. G. Holtappels, K. Klemm, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung* (Bd. 13, S. 225–252). Weinheim: Juventa.

Borman, G. D., Hewes, G. M., Overman, L. T. & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73 (2), 125–230.

Böttcher, W. & Dicke, J. N. (2008). Implementation von Standards. In W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive* (S. 143–156). Münster: Waxmann,

Böttger-Beer, M. & Koch, E. (2008). Externe Schulevaluation in Sachsen – ein Dialog zwischen Wissenschaft und Praxis. In W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive* (S. 253–264). Münster: Waxmann.

Breit, S., Friedl-Lucyshyn, G., Furlan, N., Kuhn, J.-T., Laimer, G., Längauer-Hohengaßner, H. et al. (2012). *Überprüfung und Rückmeldung* (3. aktualisierte Aufl.). Zugriff am 04. 07. 2012 unter [https://www.bifie.at/system/files/dl/bist\\_ueberpruefung-und-rueckmeldung\\_201205\\_0.pdf](https://www.bifie.at/system/files/dl/bist_ueberpruefung-und-rueckmeldung_201205_0.pdf)

Bruneforth, M., Weber, C. & Bacher, J. (2012). Chancengleichheit und garantiertes Bildungsminimum in Österreich. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen*. Graz: Leykam.

Brüsemeister, Th. (2004). *Schulische Inklusion und neue Governance*. Münster: Monsenstein & Vannerdat.

Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE). (2012a). *Bildungsstandards*. Zugriff am 10. 05. 2012 unter <https://www.bifie.at/bildungsstandards>

Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE). (2012b). *Text „Impelmentierung“*. Zugriff am 23. 10. 2012 unter <https://www.bifie.at/node/50>

Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19 (3), 476–509.

Coburn, C. E. & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, 9 (4), 173–206.

Cole, J. S., Bergin, D. A. & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology*, 33 (4), 609–624.

Dachverband Schweizer Lehrerinnen und Lehrer (LCH). (2012). *Positionspapier „Qualität durch Stärkung der Profession“*. Verabschiedet an der Delegiertenversammlung LCH vom 16. Juni 2012. Zugriff am 11. 07. 2012 unter [http://lch.ch/cms/upload/pdf/stellungnahmen/2012/120616\\_Position\\_Steuerung\\_Qualitaet.pdf](http://lch.ch/cms/upload/pdf/stellungnahmen/2012/120616_Position_Steuerung_Qualitaet.pdf)



Darling-Hammond, L. (1994). National standards and assessments: Will they improve education? *American Journal of Education*, 102 (4), 478–510.

Demmer, M. (2003). Bildungsstandards: Selektion perfektionieren oder überwinden? In Gewerkschaft Erziehung und Wissenschaft (GEW; Hrsg.), *Nationale Bildungsstandards – Wundermittel oder Teufelszeug?* (S. 8–12). Zugriff am 15. 10. 2012 unter <http://www.gew.de/Binaries/Binary35637/bildungsstandards.pdf>

Eberle, F., Gehrler, K., Jaggi, B., Kottonau, J., Oepke, M. & Pflüger, M. (2008). *Evaluation der Maturitätsreform 1995 (EVAMAR) Phase II*. Zürich: Universität. Zugriff am 03. 07. 2012 unter [http://www.ife.uzh.ch/igb/forschungsprojekte/evamar/2008\\_11\\_13\\_Kurzbericht\\_EVAMARII\\_d.pdf](http://www.ife.uzh.ch/igb/forschungsprojekte/evamar/2008_11_13_Kurzbericht_EVAMARII_d.pdf)

Eder, F. (2009). Die Schule der 10- bis 14-Jährigen als Angelpunkt der Diskussion um Struktur und Qualität des Schulsystems. In W. Specht (Hrsg.), *Nationaler Bildungsbericht Österreich 2009. Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 33–53). Graz: Leykam.

Eder, F. & Altrichter, H. (2009). Qualitätsentwicklung und Qualitätssicherung im österreichischen Schulwesen: Bilanz aus 15 Jahren Diskussion und Entwicklungsperspektiven für die Zukunft. In W. Specht (Hrsg.), *Nationaler Bildungsbericht. Österreich 2009. Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 305–322). Graz: Leykam.

Eder, F., Posch, P., Schratz, M., Specht, W. & Thonhauser, J. (Hrsg.). (2002). *Qualitätsentwicklung und Qualitätssicherung im Österreichischen Schulwesen* (1. Aufl., Bildungsforschung des Bundesministeriums für Bildung, Wissenschaft und Kultur, 17). Innsbruck: StudienVerlag.

Eder, F., Neuweg, G. H. & Thonhauser, J. (2009). Leistungsfeststellung und Leistungsbeurteilung. In W. Specht (Hrsg.), *Nationaler Bildungsbericht Österreich 2009, Band. 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 247–269). Graz: Leykam.

Education, Audiovisual & Culture Executive Agency (EACEA; Hrsg.). (2009). *National testing of pupils in Europe: Objectives, organisation and use of results*. Zugriff am 03. 07. 2012 unter [http://eacea.ec.europa.eu/education/eurydice/documents/thematic\\_reports/109EN.pdf](http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf)

Ehren, M. C. M., Altrichter, H., McNamara, G. & O'Hara, J. (in Druck). *Impact of school inspections on teaching and learning. Describing assumptions on causal mechanisms in seven European countries*. Eingereicht zur Publikation.

Ehren, M. C. M., Leeuw, F. L. & Scheerens, J. (2005). On the impact of the Dutch Educational Supervision Act; analyzing assumptions concerning the inspection of primary education. *American Journal of Evaluation*, 26 (1), 60–76.

Entwicklungsberatung in Schulen (EBIS). (2012). *EBIS-Kompetenzprofil – Erster Bewerbungsauftrag zur Eintragung in die EBIS-Liste*. Zugriff am 25. 09. 2012 unter [http://work.sqa.at/pluginfile.php/368/mod\\_page/content/25/20120625\\_EBIS\\_Kompetenzprofil\\_1st\\_CALL.pdf](http://work.sqa.at/pluginfile.php/368/mod_page/content/25/20120625_EBIS_Kompetenzprofil_1st_CALL.pdf)

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Polity Press.

Faulkner, S. A. & Cook, C. M. (2006). Testing versus teaching: The perceived impact of assessment demands on middle grade instructional practices. *Research in Middle Level Education*, 29 (7), 1–13.

Fend, H. (2006). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Fend, H. (2008). *Schule gestalten*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Figlio, D. & Loeb, S. (2011). School accountability. In E. Hanushek, S. Machin & L. Woessman (Hrsg.), *Handbook of the economics in education, Vol. 3* (S. 383–421), Amsterdam: Elsevier.

Finnish National Board of Education. (2009). *Education*. Zugriff am 05. 07. 2012 unter [www.oph.fi/english/education](http://www.oph.fi/english/education)

Fitz-Gibbon, C. T. (1997). *The value added national project final report: Feasibility studies for a national system of value-added indicators*. London: School Curriculum and Assessment Authority.

Franklin, C. A. & Snow-Gerono, J. L. (2007). Perceptions of teaching in an environment of standardized testing: Voices from the field. *The Researcher*, 21 (1), 2–21.

Freidson, E. (2001). *Professionalism. The third logic*. Chicago: University Press.

Freudenthaler, H. H. & Specht, W. (2005). *Bildungsstandards aus Sicht der Anwender* (Zentrum für Schulentwicklung, ZSE Report Nr. 69). Graz: ZSE.

Freudenthaler, H. H. & Specht, W. (2006). *Bildungsstandards: Der Implementationsprozess aus der Sicht der Praxis*. (Zentrum für Schulentwicklung, ZSE Report Nr. 71). Graz: ZSE.

Goldhaber, D. & Hannaway, J. (2004). Accountability with a Kicker: Preliminary observations on the Florida A+ accountability Plan. *Phi Delta Kappan*, 85 (8), 598–605.

Grabensberger, E., Freudenthaler, H. H. & Specht, W. (2008). *Bildungsstandards: Testungen und Ergebnisrückmeldungen auf der achten Schulstufe aus der Sicht der Praxis. Ergebnisse einer Befragung von Leiterinnen, Leitern und Lehrkräften der Pilotschulen*. BIFIE. Zugriff am 15. 10. 2012 unter <https://www.bifie.at/node/555>

Grillitsch, M. (2010). *Bildungsstandards auf dem Weg in die Praxis* (BIFIE-Report 6/2010). Graz: Leykam. [Verfügbar am 15. 10. 2012 unter <https://www.bifie.at/buch/1235> ].

Groß Ophoff, J., Koch, U., Helmke, A. & Hosenfeld, I. (2006). Vergleichsarbeiten für die Grundschule – und was diese daraus machen (können). *Journal für Schulentwicklung*, 10 (4), 7–12.

Haider, G. & Schreiner, C. (2006). PISA-Leistung und Schulnoten. In G. Haider & C. Schreiner (Hrsg.), *Die PISA-Studie. Österreichs Schulsystem im internationalen Wettbewerb* (S. 229–236). Wien: Böhlau.

Hannover, B. & Kessels, U. (2011). Sind Jungen die neuen Bildungsverlierer? Empirische Evidenz für Geschlechterdisparitäten zuungunsten von Jungen und Erklärungsansätze. *Zeitschrift für Pädagogische Psychologie*, 25 (2), 89–103.

Hanushek, E. & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24 (2), 297–327.

Helmke, A. (2004). Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule. *Seminar*, 2, 90–112.

- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsentwicklung. In G. Brägger, B. Bucher & N. Landwehr (Hrsg.), *Schlüsselfragen zur externen Schulevaluation* (S. 127–152). Bern: h.e.p.
- Herman, J., Dreyfus, J. & Golan, S. (1990). *The effects of testing on teaching and learning. Center for Research and Evaluation, Standards, and Student Testing, University of California at Los Angeles* (CSE Technical Report 327). Zugriff am 06. 07. 2012 unter <http://www.cse.ucla.edu/products/reports/TR327.pdf>
- Herman, J. L. & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement. Issues and Practice*, 12 (4), 20–25.
- Hölzl, L. & Rixinger, G. (2007). *Implementierung von Bildungsstandards in Österreich – das zweite Jahr*. Unveröffentlichte Diplomarbeit. Linz: Johannes Kepler Universität.
- Ingram, D., Louis, K. S. & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106 (6), 1258–1287.
- Jones, B. D. & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12 (39), 1–34.
- Kelly, A. & Downey, C. (2010). *Using effectiveness data for school improvement: Developing and utilizing metrics*. London: Routledge.
- King, K. & Zucker, S. (2005). *Curriculum narrowing* (Pearson Policy Report). Zugriff am 06. 07. 2012 unter <http://www.pearsonassessments.com/NR/rdonlyres/D3362EDE-7F34-447E-ADE4-D4CB2518C2B2/0/CurriculumNarrowing.pdf>
- Klieme, E. (2003). Bildungsgerechtigkeit zu fördern, ist ein wesentliches Ziel. Interview. In Gewerkschaft Erziehung und Wissenschaft (GEW; Hrsg.), *Nationale Bildungsstandards – Wundermittel oder Teufelszeug?* (S. 14–19). Zugriff am 15. 10. 2012 unter <http://www.gew.de/Binaries/Binary35637/bildungsstandards.pdf>
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (BMBF Hrsg.). (2007). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* (Bildungsforschung Bd. 1). BMBF. Zugriff am 15. 10. 2012 unter [http://www.bmbf.de/pub/zur\\_entwicklung\\_nationaler\\_bildungsstandards.pdf](http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf)
- Kogan M. (1996). Monitoring, control and governance of school systems. In OECD (Hrsg.), *Evaluating and Reforming Education Systems* (S. 25–45). Paris: Herausgeber.
- Koretz, D. (2008). Test-based educational accountability. Research evidence and implications. *Zeitschrift für Pädagogik*, 54, 777–790.
- Kornhaber, M. (2004a). Appropriate and inappropriate forms of testing, Assessment, and accountability. *Educational Policy*, 18 (1), 45–70.
- Kornhaber, M. (2004b). Assessment, standards, and equity. In J.A. Banks & C. A. M. Banks (Hrsg.), *Handbook of multicultural education*. San Francisco: Jossey-Bass.
- Kühle, B. (2010). *Zentrale Lernstandserhebungen – Ergebnisorientierte Unterrichtsentwicklung?* Berlin: Köpfer.

Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. *Empirische Pädagogik*, 21 (4), 428–447.

Kuper, H. (2005). *Evaluation im Bildungssystem*. Stuttgart: Kohlhammer.

Kussau, J. & Brüsemeister, T. (2007). Educational Governance: Zur Analyse der Handlungskoordination im Mehrebenensystem der Schule. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance – Handlungskoordination und Steuerung im Bildungssystem* (S. 15–54). Wiesbaden: VS Verlag für Sozialwissenschaften.

Kuusela, J. (2006). *Thematic approaches to equity and equity in basic education* (Assessment of learning results, 6/2006). Finnish National Board of Education. Zugriff am 28. 06. 2012 unter [http://www.oph.fi/download/47684\\_thematic\\_approaches\\_to\\_equality\\_and\\_equity\\_in\\_basic\\_education.pdf](http://www.oph.fi/download/47684_thematic_approaches_to_equality_and_equity_in_basic_education.pdf)

Lange, M. de & Dronkers, J. (2007). *Hoe gelijkwaardig blijft het eindexamen tussen scholen in Nederland? Discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen in het voortgezet onderwijs tussen 1998 en 2005* (European University Institute [EUI] Working Paper, Nr. 2007/03). Zugriff am 06. 07. 2012 unter [http://cadmus.eui.eu/bitstream/handle/1814/6665/SPS2007\\_03.pdf?sequence=3](http://cadmus.eui.eu/bitstream/handle/1814/6665/SPS2007_03.pdf?sequence=3)

Langer, R. (2008). Analyse verborgener Mechanismen im Bildungssystem. In: W. Böttcher, W. Bos, H. Döbert & H. G. Holtappels (Hrsg.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive* (S. 129–141). Münster: Waxmann.

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation*, 24 (1), 5–20.

Lersch, R. (2010). *Wie unterrichtet man Kompetenzen? Didaktik und Praxis kompetenzfördernden Unterrichts*. Wiesbaden: Institut für Qualitätsentwicklung.

Lipowsky, F. (2004). Was macht Fortbildungen für Lehrkräfte erfolgreich? *Die deutsche Schule*, 4, 462–479.

Lucyshyn, J. (2004). *Bildungsstandards – Ein weiterer Qualitätssprung für das österreichische Bildungswesen*. Salzburg: Typoskript.

Lucyshyn, J. (2006). *Implementation von Bildungsstandards in Österreich*. Arbeitsbericht. Zugriff am 15. 10. 2012 unter [http://www.gemeinsamlernen.at/siteVerwaltung/mOBibliothek/Bibliothek/Arbeitsbericht\\_Bildungsstandards\\_14\\_02\\_06.pdf](http://www.gemeinsamlernen.at/siteVerwaltung/mOBibliothek/Bibliothek/Arbeitsbericht_Bildungsstandards_14_02_06.pdf)

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 Neue Steuerung im Schulsystem* (S. 145–169). Wiesbaden: VS Verlag für Sozialwissenschaften.

Maier, U. (2006). Können Vergleichsarbeiten einen Beitrag zur Schulentwicklung leisten? *Journal für Schulentwicklung*, 10 (4), 20–28.

Maier, U. (2008). Rezeption und Nutzung von Vergleichsarbeiten aus der Perspektive von Lehrkräften. *Zeitschrift für Pädagogik*, 54, 95–117.

Maier, U. (2009). Testen und dann? Ergebnisse einer qualitativen Lehrerbefragung zur individualdiagnostischen Funktion von Vergleichsarbeiten. *Empirische Pädagogik*, 23, 191–207.

- Maier, U. (2010). Effekte von testbasiertem Rechenschaftsdruck auf Schülerleistungen: Ein Literaturüberblick zu quasi-experimentellen Ländervergleichsstudien. *Journal of Educational Research Online*, 2 (2), 125–152.
- Maier, U. & Kuper, H. (2012). Vergleichsarbeiten als Instrumente der Qualitätsentwicklung. *Die Deutsche Schule*, 104 (1), 88–99.
- Mintrop, H. & Sunderman, G. (2012). Zentrale Steuerung von Schulentwicklung mit Quoten und Sanktionen: Das US-amerikanische Großexperiment in Schulen für benachteiligte Bildungsschichten. *Die Deutsche Schule*, 104 (1), 8–30.
- Mintrop, H. & Trujillo, T. (2007). The practical relevance of accountability systems for school improvement: A descriptive analysis of California schools. *Educational Evaluation and Policy Analysis*, 29 (4), 319–352.
- Morris, A. (2011). *Student standardised testing: Current practices in OECD countries and a literature review* (OECD Education Working Papers, Nr. 65). Zugriff am 30. 08. 2012 unter <http://dx.doi.org/10.1787/5kg3rp9qbnr6-en>
- Muijs, D., Chapman, C., Collins, A. & Armstrong, P. (2010) *Maximum Impact Evaluation: The impact of Teach First teachers in schools: An evaluation funded by the Maximum Impact Programme for Teach First. Final Report*. University of Manchester. Zugriff am 04. 07. 2012 unter [http://teachforall.org/articles/max\\_impact.pdf](http://teachforall.org/articles/max_impact.pdf)
- Muller, C. & Schiller, K. S. (2000). Levelling the playing field? Students' educational attainment and states' performance testing. *Sociology of Education*, 73, 196–218.
- Netzwerk empirische Schulentwicklung (EMSE). (2008). *Nutzung und Nutzen von Schülerrückmeldungen im Rahmen standardisierter Lernstandserhebungen/Vergleichsarbeiten*. 2. Positionspapier verabschiedet auf der 9. EMSE-Fachtagung am 16.–17. Dezember 2008, Nürnberg.
- Newton, P. E. (2007). *Evaluating assessment systems*. Qualifications and Curriculum Authority, London. Zugriff am 08. 10. 2012 unter [http://archive.teachfind.com/qcda/orderline.qcda.gov.uk/gempdf/1445900599/Evaluating\\_Assessment\\_Systems1.pdf](http://archive.teachfind.com/qcda/orderline.qcda.gov.uk/gempdf/1445900599/Evaluating_Assessment_Systems1.pdf)
- Nichols, S. L. & Berliner, D. C. (2007). *Collateral damage. How high-stakes testing corrupts America's schools*. Cambridge: Harvard Education Press.
- O'Day, J. A. (2004). Complexity, accountability, and school improvement. In S. H. Fuhrman & R. F. Elmore (Hrsg.), *Redesigning accountability systems for education* (S. 12–43). New York: Teachers College Press.
- Oelkers, J. & Reusser, K. (2008). *Qualität entwickeln – Standards sichern – mit Differenz umgehen*. (Bildungsforschung Bd. 27). BMBF. Zugriff am 15. 10. 2012 unter [http://www.bmbf.de/pub/bildungsforschung\\_band\\_siebenundzwanzig.pdf](http://www.bmbf.de/pub/bildungsforschung_band_siebenundzwanzig.pdf)
- Österreichische Forschungsgemeinschaft. (2009). *Bildungsstandards in Österreich – Empfehlungen für eine erfolgreiche und nachhaltige Implementierung*. Zugriff am 06. 07. 2012 unter [http://www.oefg.at/text/stellungnahmen/Positionspapier\\_Bildungsstandards.pdf](http://www.oefg.at/text/stellungnahmen/Positionspapier_Bildungsstandards.pdf)
- Ozga, J. & Jones, R. (2006). Travelling and embedded policy: the case of knowledge transfer. *Journal of Education Policy*, 21 (1), 1–17.

Pant, H.A., Emmrich, R., Harych, P. & Kuhl, P. (2011). Leistungsüberprüfung durch Schulleistungsstudien und Vergleichsarbeiten. In H. U. Grunder, K. Kansteiner-Schänzlin & H. Moser (Hrsg.), *Professionswissen für Lehrerinnen und Lehrer* (Bd. 4, S. 123–141). Baltmannsweiler: Schneider.

Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008). Offenheit für Innovationen. *Zeitschrift für Pädagogik*, 54 (6), 827–845.

Pedulla, J. J. (2003) State-mandated testing: What do teachers think? *Educational Leadership*, 61 (3), 42–46.

Peek, R. (2004). Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QUASUM) – Klassenbezogene Ergebnisrückmeldung und ihre Rezeption in Brandenburger Schulen. *Empirische Pädagogik*, 18, 82–114.

Prenzel, M. & Seidel, T. (2010). Bildungsmonitoring und Evaluation. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.), *Bildungspsychologie* (S. 329–345). Göttingen: Hogrefe.

*Qualität in Multikulturellen Schulen (QUIMS)*. (2012). Zugriff am 03. 07. 2012 unter [http://www.vsa.zh.ch/internet/bildungsdirektion/vsa/de/schulbetrieb\\_und\\_unterricht/qualitaet\\_multikulturelle\\_schulen\\_quims.html](http://www.vsa.zh.ch/internet/bildungsdirektion/vsa/de/schulbetrieb_und_unterricht/qualitaet_multikulturelle_schulen_quims.html)

Resnick, L. B., Besterfield-Sacre, M., Mehalik, M. M., Sherer, J. Z. & Halverson, E.R. (2007). A framework for effective management of school system performance. In P. A. Moss (Hrsg.), *Evidence and decision making. Yearbook of the National Society for the Study of Education*, 106 (1), 155–185. Malden, MA: Blackwell Publishers.

Rosenshine, B. (2003). High-stakes testing: Another analysis. *Education Policy Analysis Archives* 11, 24. [Verfügbar am 12. 10. 2012 unter <http://epaa.asu.edu/ojs/article/view/252>].

Sahlberg, P. (2009). Educational change in Finland. In A. Hargreaves, M. Fullan, A. Lieberman & D. Hopkins (Hrsg.), *International Handbook of Educational Change*, (S. 1–28). Dordrecht: Kluwer Academic Publishers.

Schimank, U. (in Druck). Governance und Professionalisierung – Notizen zu einem Desiderat. In K. Maag Merki, R. Langer und H. Altrichter (Hrsg.), *Educational Governance als Forschungsperspektive. Strategien, Methoden und Forschungsansätze in interdisziplinärer Perspektive*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Schleicher, A. (2012, Juni). *Das schweizerische Schulsystem im internationalen Kontext und Vergleich*. Vortrag im Rahmen des Workshops „Zukunft Bildung Schweiz“. Thun, 21. Juni 2012.

Schrader, F.-W. & Helmke, A. (2003). Evaluation - und was danach? *Schweizerische Zeitschrift für Bildungswissenschaft*, 25 (1), 79–110.

Schulze, F. (2012). Folgen zentraler Lernstandserhebungen: Unterscheiden sich Lehrkräfte in der Auseinandersetzung mit Lernstandserhebungen? *Die Deutsche Schule*, 104 (1), 43–56.

Sheldon, K. M. & Biddle, B. J. (1998). Standards, accountability, and school reform: Perils and pitfalls. *Teachers College Record*, 100, 164–180.

Simola, H. (2005). The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41 (4), 455–470.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C. & Cherland, M. (1987). *The role of testing in elementary schools*. Los Angeles: Center for Research on Educational Standards and Student Tests and the Office of Educational Research and Improvement, Department of Education.

Specht, W. (2006). Von den Mühen der Ebene. Entwicklung und Implementation von Bildungsstandards in Österreich. In F. Eder & A. Gastager & F. Hofmann (Hrsg.), *Qualität durch Standards? Beiträge zum Schwerpunktthema der 67. Tagung der AEPF* (S. 13–37). Münster: Waxmann.

Spiel, C., Schober, B. & Litzenberger, M. (2008). *Evaluation der Eignungstests für das Medizinstudium in Österreich* (Projektbericht für das Bundesministerium für Wissenschaft und Forschung). Zugriff am 23. 10. 2012 unter [http://www.bmwf.gv.at/uploads/tx\\_contentbox/Spiel\\_Studie.pdf](http://www.bmwf.gv.at/uploads/tx_contentbox/Spiel_Studie.pdf)

Spillane, J. P. (2005). Standards deviation: How Schools Misunderstand Education Policy. *CPRE Policy briefs, RB-43*, 1–12.

SQA – Schulqualität Allgemeinbildung [Webpage]. BMUKK. [Zugriff am 14. 11. 2012 unter <http://www.sqa.at/>].

Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In: L. S. Hamilton, B. M. Stecher & S. P. Klein (Hrsg.), *Making sense of test-based accountability in education* (S. 79–100). Santa Monica, CA: Rand.

Steffens, U. (2009). *Lernstandserhebungen in den deutschen Ländern – Probleme und Perspektiven*. Unveröffentlichtes Manuskript, Institut für Qualitätsentwicklung, Wiesbaden.

Stojanov, K. (2007). Bildungsgerechtigkeit im Spannungsfeld zwischen Verteilungs-, Teilhabe- und Anerkennungsgerechtigkeit. In M. Wimmer, R. Reichenbach & L. Pongratz (Hrsg.), *Gerechtigkeit und Bildung* (S. 29–48). Paderborn: Schöningh.

Tymms, P. & Merrell, C. (2007). *Standards and quality in English primary schools over time: The national evidence* (Primary Review Research Survey 4/1). University of Cambridge, Faculty of Education. Zugriff am 06. 07. 2012 unter <http://image.guardian.co.uk/sys-files/Education/documents/2007/11/01/overtime.pdf>

Urdan, T. C. & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy*, 8, 137–156.

Visscher, A. J. & Coe, R. (2003). School performance feedback systems: Conceptualisation, Analysis and Reflection. *School Effectiveness and School Improvement*, 14 (3), 321–349.

*Vorblatt und Erläuterungen zum BGBl. I Nr. 117/2008*. (2008). Zugriff am 23. 10. 2012 unter [http://www.bmukk.gv.at/medienpool/16327/ris\\_schugmaterialien\\_2.pdf](http://www.bmukk.gv.at/medienpool/16327/ris_schugmaterialien_2.pdf)

Weir, S., Archer, P., Flaherty, A. & Gilleece, L. (2011). *A report on the first phase of the evaluation of DEIS*. Dublin: Educational Research Centre. Zugriff am 06. 07. 2012 unter [http://www.erc.ie/documents/deis\\_p1\\_main.pdf](http://www.erc.ie/documents/deis_p1_main.pdf)

Wikeley, F., Stoll, L. & Lodge, C. (2002). Effective school improvement: English case studies. Educational Research and Evaluation. *School Effectiveness and School Improvement in a European Context* 8 (4), 363–385.

---

Windzio, M., Sackmann, R. & Martens, K. (2005). *Types of governance in education – A quantitative analysis* (TranState Working Papers, Nr. 25). Bremen: Universität. Zugriff am 03. 07. 2012 unter <http://www.econstor.eu/dspace/bitstream/10419/28275/1/501321926.PDF>

Wortman, P. M. (1983). Evaluation research: A methodological perspective. *Annual Review of Psychology*, 34 (1), 223–260.

Zeitler, S., Heller, N. & Asbrand, B. (2012). *Bildungsstandards in der Schule. Eine rekonstruktive Studie zur Implementation der Bildungsstandards*. Münster: Waxmann.

Zukunftskommission (Haider, G., Eder, F., Specht, W. & Spiel, C.; BMBWK Hrsg.). (2003). *Das Reformkonzept der österreichischen Zukunftskommission. Zukunft: Schule. Strategien und Maßnahmen zur Qualitätsentwicklung*. Zugriff am 04. 10. 2012 unter [http://schule.salzburg.at/faecher/didaktik/reformkonzept\\_zukunft\\_schule.pdf](http://schule.salzburg.at/faecher/didaktik/reformkonzept_zukunft_schule.pdf)

Zukunftskommission (Haider, G., Eder, F., Specht, W., Spiel, C. & Wimmer, M.; BMBWK Hrsg.). (2005). *Abschlussbericht der Zukunftskommission an Frau Bundesministerin Elisabeth Gehrler*. Zugriff am 08. 07. 2012 unter <http://www.plattform-educare.org/Datenbank/Abschlussbericht%20Zukunftskommission%20April%202005.pdf>