

Jürgen Oelkers

*Bildungspolitische Konsequenzen nach PISA:
Neue Steuerung, Standards und Evaluation^{*)}*

In der Schweiz geschieht bildungspolitisch Erstaunliches. Festzustellen ist eine Bewegung hinter oder oberhalb des Förderalismus, ein „Mega-Trend“, wie man neudeutsch sagt, der radikal veränderte Formen der Steuerung des Schulsystems durchsetzen soll. Das erscheint wie ein kühnes Unterfangen, sind doch Volksschulen schon auf der Ebene der Gemeinden zum Teil sehr unterschiedlich, bezogen auf ihre Kulturen, die Steuerung durch die Schulkommissionen und nicht zuletzt, was ihre finanziellen Ressourcen betrifft. Heterogen ist aber nicht nur die unterste Ebene. Faktisch bestehen 26 kantonale Einzelsysteme nebeneinander, die nunmehr, wie es eher euphemistisch heisst, „harmonisiert“ werden sollen. Erstaunlich ist das, weil eine wie immer massvolle, so doch Vereinheitlichung angestrebt wird, die bis vor kurzem noch ausserhalb des Vorstellungshorizontes lag, wenn dieser sich realistisch nennen wollte. Nunmehr geht es ernsthaft um Annäherung, und das ist ein Resultat der PISA-Diskussion, also eine Reaktion der Politik auf Forschung (Bildungsmonitoring Schweiz 2002).

Das lässt auch die Parteien nicht unberührt, hat doch unlängst eine grosse und alte, nicht mehr ganz so mächtige Schweizer Partei alle Bestrebungen dieser Tagung unterstützt und das so drastisch, dass gleich auch noch ein Vogt für Bildungssteuerung in Bern ins Spiel gebracht wurde. Das ist allerdings schon mehrfach in der jüngeren Bildungsgeschichte gescheitert und hat auch diesmal keine Chancen, wobei dann verwundert, warum die Idee zwischen 1874 und heute periodisch attraktiv erscheinen kann. Auch die andere Idee der nicht genannten Partei, von der ersten Klasse an wieder Noten einzuführen, versucht sich am untauglichen Objekt, weil strikte Benotung nicht etwa das Zaubermittel ist, die Leistungen zu verbessern. Aber sonst ist die Schweizerische Bildungspolitik offenbar bereit, sich auf neue Entwicklungen einzulassen und nationale Massnahmen zur Qualitätssicherung abzustimmen, die den Bildungsvogt nicht ersetzen, sondern ihn überflüssig machen.

Dabei geht es primär um neue Verfahren der Systemsteuerung, vor allem um die Entwicklung nationaler Bildungsstandards, die Ersetzung der herkömmlichen Schulaufsicht durch regelmässige Formen der Evaluation sowie den Aufbau eines testbasierten Bildungsmonitoring. Dafür steht die Formel „HarmoS“, deren Realität aber erst im Jahre 2007 vollumfänglich absehbar ist. Vor diesem Publikum muss ich weder erläutern, was die Theorie der Bildungsstandards ausmacht, noch muss ich die Differenz von „externer“ und „interner“ Evaluation darlegen oder erklären, wie das Bildungsmonitoring arbeiten soll. Das setzt mich in eine Verlegenheit, denn worüber soll ich sprechen, wenn es Resultate noch nicht gibt und zugleich den Experten die Entwicklung klar ist?

Das neupädagogische Wort „PISA“ verweist auf *Qualität*, gute und schlechte. Genauer: Die vergleichende Auswertung des PISA-Leistungstests teilte nationalen

^{*)} Vortrag auf der ARGEV-Tagung vom 2. November 2004 in Baden.

Bildungssystemen einen Rang auf einer Skala zu, der mit einer hohen bildungspolitischen Symbolik verbunden ist. Erst danach ist in deutschsprachigen Ländern von „Qualitätssicherung“ die Rede gewesen. Vorher war die Annahme der Qualität der Bildungssysteme eine Selbstzuschreibung, die für Zufriedenheit sorgte, weil keine Daten vorlagen. Bis dahin herrschte ein Vergleichsverbot, das an ROUSSEAUS *Emile* erinnert: Wer sich mit Anderen vergleicht, gefährdet die eigene Souveränität. Aber gilt das auch umgekehrt, wird die eigene Qualität gefördert, wenn man weiss, dass Andere besser sind? Und was lässt sich tun, besser zu werden, wenn fremde System nicht einfach kopiert werden können?

Wie alle pädagogischen Vokabeln und Slogans ist natürlich auch „Qualitätssicherung“ ein leeres Wort und *kein* gehaltvolles Programm, solange man nicht weiss, welche Praxis sich damit verbindet. Alle gut klingenden leeren Wörter der Pädagogik, von denen es viele gibt und die modischen Schwankungen unterworfen sind, beziehen sich auf grosse Versprechungen, bei denen es nicht auf die Praxis, sondern auf die Verheissung ankommt. Das verwandte Wort „Qualitätsmanagement“ ist im pädagogischen Kontext zunächst eine befremdliche Grösse gewesen, das Wort weckte Widerstand, aber dann sehr schnell auch Neugier und schliesslich Interesse, weil es eine neue Lösung anzuzeigen schien, die gleichsam über Nacht zustimmungsfähig wurde. Niemand wusste so Recht, was es war, gleichwohl überzeugte die Lösung, bevor es eine Praxis gab.

Vergleicht man die Situation heute mit der vor fünf oder zehn Jahren, dann fällt auf, dass aus der leeren Hülse „Qualitätsmanagement“ zunehmend eine beschreibbare Praxis geworden ist. Erstaunlich ist, wie rasch das geschehen ist. Alle Kantone führen Massnahmen zur Qualitätssicherung durch, viele adaptieren sogar Zürcher Modelle oder entwickeln sie weiter, die Stichworte sind überall sehr ähnlich, der Konsens scheint zu wachsen, selbst das Geld steht zur Verfügung, bisher wenigstens, tragisch ist vielleicht nur, dass dem Zürcher Stimmvolk vor zwei Jahren unzureichend vermittelt werden konnte, welche Avantgarde da am Werke war. Umso mehr freuen sich die Anderen, sie sparen Entwicklungskosten und gewinnen einen Vorsprung.

Heute enthält die Qualitätssicherung in aller Regel mindestens sieben ausschlaggebende Elemente:

- Aufbau von Schulleitungen mit Kompetenzen und Weisungsbefugnissen,
- höchst mögliche Transparenz des schulischen Angebots,
- Verbesserung der internen Kommunikation,
- Abstimmung von Standards des Unterrichts,
- Mitarbeiterbeurteilung und interne Evaluationen,
- regelmässige externe Evaluationen,
- Offenlegung der dabei erzeugten Daten
- Zielvereinbarungen im Blick auf die nächste Etappe der Schulentwicklung.

Das sind keine Folterwerkzeuge, sondern Instrumente der Professionalisierung, die heute international üblich sind. Die Entwicklung in der Schweiz vollzieht nur nach, was in skandinavischen oder angelsächsischen Systemen vor fünfzehn oder zwanzig Jahren begonnen wurde. Im Blick darauf trägt der Eindruck nicht, dass Bildungssysteme auf dieser Linie modernisiert werden, mit Qualitätsgewinnen, wie der Vergleich zwischen guten und weniger guten Ländern der PISA-Studie zeigt.

Eine in der Diskussion wenig beachtete Funktion dieser Verfahren und Instrumente ist, dass sie die Intransparenz des schulischen Geschehens überwinden sollen. Von der

Leistungsbeurteilung bis zu den Kriterien des guten Unterrichts ist im Schulalltag zuviel undurchsichtig und zu wenig wirklich explizit, also von Eltern, Schülern und Kollegen nachvollziehbar. Die tatsächlichen Kriterien zum Beispiel der Notengebung sind oft nur der Lehrkraft bekannt, während Schüler hier nicht selten dramatische Schicksale erleben, die auf ihre Schulkarriere nachhaltigen Einfluss ausüben. Soll sich das ändern, muss Transparenz zur Grundregel werden, und zwar nach Innen und Aussen gleichermaßen. Und das ist wirklich leichter gesagt als getan.

Kein einziges Element von der Qualitätssicherung lässt sich verordnen, vielmehr müssen alle Elemente überzeugend entwickelt werden. Schulleitungen sind sensible Einrichtungen, die sich vor Ort und also je anders bewähren müssen, Lernprogramme, die das schulische Angebot darstellen, werden keine einheitliche Form annehmen, und selbst die viel zitierten Bildungsstandards müssen lokal angepasst werden, wenn sie praktikabel sein sollen. Die Schulorganisation muss sich selbst gestalten können, was sich verändert, ist das Verhältnis von Autonomie und Kontrolle. Evaluationen sind Versuche, die Stärken und Schwächen einer Schule zu bestimmen, die als mehr sein muss, als die Summe des Kollegiums. Man kann nicht zugleich das Einzelkämpfertum beklagen und es an dieser Stelle nutzen. Das Credo ist *Entwicklung* der Schule, die ein transparentes Klassenzimmer voraussetzt.

Aber soll das gehen? Was sind die Hindernisse? Und was hat das mit dem Tagungsthema zu tun? Ich werde zunächst etwas über den Gegenstand aussagen, den ich nicht einfach „Schule“ oder „Praxis“ nenne, sondern den ich über schulische Wissensformen und Problemlösungen definiere. Wer Schule verändern will, muss ihr *eigenes* Know How in Rechnung stellen, nicht einfach eine bestimmte Variante der Organisationssoziologie (1). In einem zweiten Schritt beleuchte ich einen zentralen Aspekt des Bildungsmonitoring, nämlich das Testen. Dieser Teil ist bislang erstaunlich wenig thematisiert worden, vermutlich weil hier die Praxis noch nicht erfahrbar ist. In anderen Ländern ist das inzwischen erheblich anders (2). Abschliessend gehe ich auf Probleme der Implementation ein, an denen sich die Reform letztlich entscheiden wird. Der Nutzen auch und gerade der Forschung für die Praxis muss nachgewiesen werden, und das vor einem Publikum, das nur begrenzt leidensfähig ist (3).

Was wir „neue Steuerung“ nennen, hat nichts mit einem Tanker zu tun, der eine neue Richtung einschlagen soll und aber schwer beweglich ist. Es gibt kein Äquivalent zum Tanker, auch wenn immer die Systemtheorie bemüht wird, unser Objekt ist nicht einheitlich, sondern hoch differenziert und deswegen schwer zu steuern. Wir haben es mit verschiedenartigen professionellen Feldern zu tun, mit Akteuren in unterschiedlichen Positionen, mit einer kritischen Öffentlichkeit und nicht zuletzt mit Bürgerinnen und Bürger, die wissen wollen, wie die Qualität der Schule tatsächlich beschaffen ist, aber dabei nie *eine* Meinung vertreten. Es gibt komplizierte Feedbackschlaufen und Grenzen der Belastbarkeit, auch mit Reformthemen. Andererseits besteht das Zeitfenster nur jetzt, und es wird darauf ankommen, eine politisch überzeugende Entwicklung einzuleiten.

1. Das Arbeitswissen in Schulen

Eine zentrale Frage der Schulentwicklung ist, wann und unter welchen Umständen Schulen *neue* Probleme akzeptieren. Diese Frage wird sich nur dann beantworten lassen, wenn die Lernprozesse *in* den Schulen beschrieben werden. Das geschieht nicht dadurch, dass Schulen mit Postulaten konfrontiert werden, etwa solchen, die sie „fit“ machen sollen für die

Wissensgesellschaft; vielmehr muss genauer untersucht werden, wie sich Schulen fortlaufend konstituieren und was sie veranlasst, sich in neue Richtungen zu entwickeln, wenn sie mehr davon haben, das *nicht* zu tun. Schulen sind aus guten Gründen eher konservative Institutionen, die nicht jeder pädagogischen Mode nachjagen, sondern die vom Bewährten ausgehen. Das kann nicht einfach „träge“ genannt werden, sondern ist die Folge von bislang nicht überbotenen Problemlösungen.

Eine Möglichkeit der Analyse ist die Untersuchung der Wissensdynamiken in Schulen.¹ Schulen sind Institutionen des Wissens, aber nicht einfach in dem Sinne, dass Schülern auf portionierte Weise Wissen „vermittelt“ wird, sondern in einem weit konstitutiveren Sinne. Schulen organisieren Lernprozesse im Blick auf Wissensangebote, diese Angebote sind historisch sehr stabil und sie werden gestützt durch eine Struktur, die nicht ständigen Wandel, sondern Beharrung belohnt. Wer das ändern will, muss genau untersuchen, die die Transformation des Wissens in den Schulen zustande kommt und was ihre Dynamik bestimmt.

Lehrkräfte verwenden bestimmte Varianten eines tradierten Berufswissens, das kaum Bezug hat zur Forschung und gleichwohl hochgradig wirksam ist. Die Medien des Lernens enthalten didaktische und methodische Entscheidungen, die auf Plausibilität im Wissenshorizont der Lehrkräfte abzielen. Das Vorwissen der Schüler muss auf die Aufgaben und Leistungen im Unterricht abgestimmt werden. Die Theorien der Lehrkräfte werden dem ständigen Test der Erfahrung ausgesetzt, mit dem sie eher bestätigt als verändert werden. Das einmal aufgebaute Repertoire der Lehrkräfte, ihr Know How im Alltag, lässt sich nur schwer beeinflussen, nicht nur weil es sich mit starken Überzeugungen verbindet, sondern weil es auf den zeitlichen Rahmen des professionellen Handelns abgestimmt ist. Lehrkräfte müssen in den ersten Berufsjahren lernen, ihre Kompetenz unter Bewältigung von Stresssituationen aufzubauen, wobei die ersten guten Problemlösungen nachhaltig wirksam sind.

Auf diese Weise wird sichtbar

- die historische Entwicklung und die Präsenz der Lehrmittel,
- das fortlaufend konstituierte Kompetenz der Lehrkräfte,
- das *tacit knowledge* der Schule
- der Wandel und die Stabilität der Kognitionen der Praxis
- der reale Aufbau des Wissens bei den Schülern
- Tradierung und Innovation des Wissens auf allen Ebenen.

„Wissen“ allgemein lässt sich nach einem Vorschlag von DAVENPORT und PRUSAK mit zwei Kriterien so fassen:

- Knowledge derived from minds at work.
- Knowledge is more than data or information.

Bezogen auf Organisationen wie Schule kann von *working knowledge* gesprochen werden, von Arbeitswissen, das in der Organisation ständig verwendet wird und nicht deswegen unintelligent ist, weil es über Jahre in Gebrauch ist. Dafür wird folgende Definition vorgeschlagen:

¹ Begrifflich ist mit GILBERT RYLE (1949) zwischen *knowing that* und *knowing how* zu unterscheiden ist, was im Deutschen etwa der Unterscheidung zwischen „Wissen“ und „Können“ entspricht und sich didaktisch mit „Lehr-“ und „Lernwissen“ in Verbindung bringen lässt.

„Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices and norms” (DAVENPORT/PRUSAK 1998, S. 5).

Die Kernfrage ist, wie das *working knowledge* der Bildungsinstitutionen verbessert werden kann, ohne einfach nur, wie heute etwa bei der Rede von „Bildungsstandards,“ auf den Wechsel der Perspektive zu vertrauen, der die Praxis ja nicht schon anders macht. Schulen operieren mit einem eisernen Bestand an intelligentem Arbeitswissen, das nicht einfach zur Disposition steht, nur weil die Politik Bildungsstandards abverlangt und die Verwaltung tätig wird. *Working knowledge* sind die Lösungen der Praxis, wer sie verbessern will, muss sie erreichen, und dazu genügt es nicht, Diskurse zu führen oder die Bildungspolitik zu beschäftigen.

Notwendig ist dafür eine Veränderung der Theorie der Schule, die weder in einer funktional noch in einer normativen Fassung zutreffend ist. Ich gehe aus von einer historischen Variante der massgeblich von JOHN DEWEY begründeten Theorie des Problemlösens.² Die Variante geht davon aus, dass „Problemlösen“ nicht nur auf die Lernenden, sondern auch auf ihre Institution angewendet werden kann. In dieser Perspektive bestehen Schulen nicht aus einer unbegrenzten Kette von *Problemen*, sondern aus einer begrenzten Serie von *Problemlösungen*. Der fragend-entwickelnde Unterricht, heute schamvoll genannt „Unterricht überwiegend erteilt von der Lehrkraft,“ ist ebenso eine Problemlösung wie die Unterscheidung von Lehrmitteln „für die Hand des Schülers“ und „Lehrmitteln für die Hand des Lehrers“³ oder die Erfindung der Studentafel, des Notenschemas⁴ oder der organisatorischen Grundeinheit der Lektion. Ohne solche erfolgreichen und langfristig stabilen historischen Problemlösungen könnte Schule gar nicht stattfinden.

Eine der heutigen Lösungsvorschläge ist das „lernende System“ oder die „intelligente Organisation.“ Dabei handelt es sich zunächst nur um Metaphern oder Sprachregelungen, die mit politischen Veränderungswünschen verbunden sind und Macht mit Rhetorik zu verbinden suchen. Die politische Rhetorik betont die „Notwendigkeit“ oder „Unausweichlichkeit“ des Wandels und übersieht gerne, dass der Gegenstand Schule *vorhanden* ist und über genügend Erfahrungen verfügt, sich selbst helfen zu können. Es ist auch kein Aufruf nötig, zur „lernenden Organisation“ erst werden zu müssen - das System lernt und hat immer gelernt,

² *How We Think* (1910) (Dewey 2002). Das Konzept des Problemlösens ist freilich älter und geht vor allem auf Theorien des Spiels am Ende des 19. Jahrhunderts zurück (OELKERS 2004).

³ Schon Mitte des 19. Jahrhunderts wird in den Kompendien der Volksschule festgehalten, dass zwischen Lehrmitteln für die *Lehrkräfte* und Lehrmitteln für die *Schüler* zu unterscheiden sei (HERGANG 1852, S. 170ff.).

⁴ Der heutige Schema der Schulnoten geht wesentlich auf den von ANDRES REYHER (1601-1673) verfassten Gothaer *Schulmethodus*. REYHER war von 1639 an Rektor des Gymnasiums in Gotha. Er studierte in Leipzig, machte dort seinen Magister und lehrte an der Philosophischen Fakultät, bevor er eine Karriere als Gymnasialrektor machte. Der Herzog von Gotha, ERNST der Fromme (1601-1675), berief REYHER von Lüneburg nach Gotha, um die Reform der Volksschule voran zu bringen. „Schulmethodus“ wurde der „spezial- und sonderbare Bericht“ zur neuen Volksschulordnung genannt, der 1642 erlassen wurde. Diese Ordnung führte in Gotha die Schulpflicht vom fünften bis zum zwölften Lebensjahr ein, regelte die Rangfolge der Elementarfächer sowie die Einteilung der Schüler nach Klassen, schrieb die Lehrbücher vor, verfügte Regeln der Unterrichtsmethode und der Schuldisziplin, sah Realien vor und fasste schliesslich das Schema der Notengebung. In diesem Sinne handelte es sich um die erste moderne Standardisierung des Volksschulunterrichts in Deutschland.

nur eigeninnig. Neue Verfahren der Systementwicklung müssen sich als wirksam erweisen oder werden bis zur Unkenntlichkeit angepasst. Ich könnte auch sagen, die Reform rechnet nicht mit der Listigkeit des Systems, das schon ganz andere Reformattacken gut überstanden hat.

Das erklärt sich aus der Geschichte. Die moderne Schule hat keinen namentlichen Begründer, ihre bis heutige gültige Struktur hat niemand persönlich gestiftet, vielmehr ist sie das Resultat eines lang gezogenen Entwicklungsprozesses, der vom Austausch zwischen vielen, zunächst ganz unterschiedlichen kleinen Systemen und ihren Repräsentanten gelebt hat. Was die Schulentwicklung beförderte, war nicht nur staatliches Engagement, sondern *eigenes* Lernvermögen, die stete Verbesserung von Form, Struktur und Gehalt. „Schule“ ist so zunächst und grundlegend *Schulgeschichte*. Das heutige Schulsystem ist das Ergebnis eines eigentümlichen und in vieler Hinsicht bindenden historischen Prozesses.

Genauer müsste ich im Plural sprechen: Das Schulsystem ist das Resultat vielfältiger und vielfach folgenreicher Entwicklungen, die untereinander komplex verknüpft sind. Es ist nicht *eine*, in sich konsistente Geschichte, sondern es sind *verschiedene* Geschichten, die nicht linear und eindimensional vorgestellt werden dürfen. Diese eng verknüpften Geschichten sind vorauszusetzen, wenn es heute erneut um die Modernisierung der Schule gehen soll. Dabei muss eingegrenzt und genauer bestimmt werden, was eigentlich die Innovation ausmachen soll. „Bildungsstandards“ für sich genommen sind dies eben so wenig wie allgemeine Strategien der Qualitätssicherung, die unterstellen, zuvor sei nie versucht worden, die „Schulqualität“ zu verbessern.

Wer ein heutiges Schulproblem verstehen und handhaben will, bezieht sich unwillkürlich auf historischen Markierungen. Diesen Ausdruck verwende ich, um nicht einfach von „Vorgeschichte“ zu sprechen, die mit der heutigen Realität nur so viel zu tun hat, dass sie nominell im gleichen System stattfand, ohne mit markanten Folgen verbunden gewesen zu sein. Das würde „Schulgeschichte“ lediglich in einem historistischen Verständnis interessant machen, als Tatbestand einer irrelevant gewordenen Vergangenheit, während ich davon ausgehe, dass die vielfältigen Geschichten des Systems durchgehende Folgen haben. Sie sind nicht lediglich nachträgliche Konstrukte, sie sind *materiell* im System verankert.

Die Struktur der Schule:

- Lehrpläne und Lehrmittel,
- Methoden des Lehrens und Lernens,
- Grundsituationen des Unterrichtens,
- die Toleranzzonen des Verhaltens,
- Fächer und deren Hierarchien,
- Rollen und Funktionen,
- die Organisation der Zeit
- und die Architektur des Raumes,

verdankt sich historischen Prozessen, die sehr langfristig angelegt sind, die zudem langsam sind, immer wieder bestätigt wurden und die eine *Erfolgsgeschichte* hinter sich haben. Die Wellen der Schulkritik dürfen nicht darüber hinwegtäuschen, dass es sich um ein extrem erfolgreiches System handelt, denkt man an Indikatoren wie

- Grösse und Differenzierung,
- Privilegien des Personals,

- kulturelle Zuständigkeit,
- materielle Ausstattung,
- soziale Akzeptanz,
- Exklusivität,
- Budgetsicherheit und Ähnliches mehr.

Die Kritik ist oft nur deswegen plausibel, weil sie diese Indikatoren übersieht und sich auf moralische Kategorien festlegt. Tatsächlich aber sind historisch schwergewichtige Faktoren wirksam, die zunächst einmal ausschliessen, das System „Schule“ ein zweites Mal grundlegend neu zu erfinden. Es ist anpassungsfähig *in* und *mit* der gegebenen Struktur, daher sind nur solche Veränderungen möglich, die zur Struktur passen und sie nicht gefährden. Alles andere wird abgestossen oder unpassend gemacht. Das wird auch mit „Bildungsstandards“ der Fall sein, wenn sie diesem Kriterium der Passung nicht entsprechen oder ungenügend implementiert werden. Die Formulierung von Standards ist vermutlich das geringste Problem, ihre Umsetzung in das Arbeitswissen der Schule und des Unterrichts dürfte die grösste Herausforderung darstellen.

Diese Analyse erlaubt folgende Schlussfolgerung: Schulen *sind* „lernende Systeme,“ aber solche, die sich auf sich selbst beziehen und ihre relevanten Umwelten wesentlich unter *diesem* Gesichtspunkt wahrnehmen (OELKERS 2000). Das Lernen des Systems erfolgt mit den Vorgaben des Systems und ist keineswegs beliebig veränderbar. Schulen, wenn meine historische Theorie zutrifft, bestehen aus *gelösten* Problemen, die Lösungen werden für den Alltag genutzt, ohne sie ständig in Frage zu stellen, und sie sind nicht einfach stupid. Wer die Absicht hat, etwas zu verbessern, muss die Stelle im System bestimmen, wo das geschehen soll. Appelle sind oft auch deswegen folgenlos, weil der Adressat unbestimmt bleibt. Wenn etwas wirken soll, dann hängt das massgeblich von der Stelle ab, wo es eingesetzt und entwickelt wird. Daher ist die Implementation ein vorrangiges und kein nachgeordnetes Problem.

Das lässt sich an Methoden des Unterrichts gut zeigen. Unterricht wird nicht einfach dadurch besser, dass mehr Geld investiert wird, die formalen Qualifikationen der Lehrkräfte ansteigen oder die Schulen eine bessere Ausstattung erhalten. Vielmehr muss das Know How der Lehrkräfte als *Ressource* verstanden werden, ebenso die Lernerfahrungen der Schüler,⁵ die nicht einfach einen Wert an sich darstellen, wie zum Beispiel die unverbindliche Rede von den „Potentialen“ der Schüler oder Lehrer anzeigt; der Werte der je vorhandenen Ressourcen zeigt sich im Gebrauch. Entscheidend ist nicht, dass Unterricht stattfindet, sondern welche Evidenz er erzielt. Kein Unterricht regt direkt Lernen an, wie viele Lehrkräfte glauben. Lehren besteht aus verschiedenen Aktivitäten, die Schülerinnen und Schüler instand setzen sollen, Lernmaterialien, Aufgaben und andere Ressourcen mehr oder weniger gut zu gebrauchen. Der Wert der Ressourcen ist davon abhängig, nicht ob, sondern wie sie gebraucht werden (COHEN/ RAUDENBUSH/LOEWENBERG BALL 2002, S. 86).

Wer die Unterrichtsqualität verbessern will, kann sich nicht einfach auf einen „Klippert-Effekt“ verlassen, die Nachbesserung der Kompetenz mit Verfahren und Instrumenten, die in der Ausbildung nicht gelernt wurden und zugleich unmittelbar einleuchten. Aber verbessern Methoden des Lehrerhandelns und damit einhergehend der

⁵ “Teachers’ knowledge, skills, and strategic actions can be seen as resources, as can students’ experiences, knowledge, norms, and approaches to learning. These resources attach to the agents of instruction and appear to mediate their use of such conventional resources as time and material” (COHEN/ RAUDENBUSH/LOEWENBERG BALL 2002, S. 85).

Aktivierung oder Motivierung der Schüler die Unterrichtsqualität? Mehr und bessere Bücher sagen über den Nutzen wenig aus, der Besuch einer Fortbildung wird nie auf seinen Effekt hin registriert, von den Eigenschaften der Lehrkräfte kann nicht auf ihren Unterrichtserfolg geschlossen werden. Neuere Studien legen eher einen ganz anderen Zusammenhang nahe:

„Instructional quality instead appears to depend on the mobilization of a complex collection of knowledge and practices, collective actions, and the conventional resources on which those actions and practices draw ... Schools cannot be improved simply by intervening either on individuals' stocks of knowledge or on schools' stocks of conventional resources. Improvement would instead depend on the way in which students, teachers, and school leaders use resources; the knowledge and skill applied in using resources for instruction; the usability of resources; and the conditions that enable resource use" (ebd., S. 86/87).

Die heutige Bildungspolitik ist nicht primär an der Verbesserung der Unterrichtsqualität ausgerichtet, auch nicht an dem Zusammenhang von Ressourcen und Nutzung im Lehr-Lernprozess. Das überragende Interesse gilt der Entwicklung von Bildungsstandards und einer darauf bezogenen Leistungsmessung. Die Unterrichtsqualität wird der Ausbildung überlassen, und die Frage, wie gelehrt und gelernt wird, also wie die Ressourcen genutzt werden, ist Sache der einzelnen Lehrkraft. Aber Bildungsstandards sind nur dann zu gebrauchen, wenn sie Eingang in das Wissenssystem Schule finden, also genutzt werden. Was entwickelt wird, muss in den Köpfen der Lehrkräfte ankommen und die Lehrmittel erreichen.

Die heutige Diskussion bezieht die Rede von den Bildungsstandards fast immer auf Inhalte (contents) bestimmter Fächer (domains) in bestimmten Klassenstufen (grades). Standards im Blick auf Unterrichtsqualität oder Ressourcennutzung sind noch kaum angedacht. Die Inhalte der Schulfächer jedoch stehen nicht einfach neu zur Disposition, sondern liegen weitgehend fest und müssen nur gefasst werden. Die Idee der „Bildungsstandards“ formuliert einfach eine Zielerreichung auf verschiedenen Stufen oder Kompetenzniveaus, die fortlaufend getestet werden soll (Zur Entwicklung nationaler Bildungsstandards 2003). Aber was weiss man über diese Politikstrategie? Die Frage wird mich in einem zweiten Schritt beschäftigen, auch weil oft hier das grosse Problem der Zukunft gesehen wird.

2. *Teaching and Testing*

Dort, wo man Bildungsstandards seit längerem kennt, etwa in den Vereinigten Staaten, sind Warnungen oder Relativierungen nicht zu überhören. Die Warnungen sind zweigeteilt, entweder stellen sie generell in Frage, dass Testen die Bildungsqualität verbessert oder aber sie kritisieren die durchgehend schlechte Qualität des Testens. Zu der Kritikern der zweiten Kategorie gehört EVA BAKER, Co-Direktorin des *National Center for Research on Evaluation, Standards, and Student Testing* in Los Angeles.⁶ Sie verweist zunächst auf zwei

⁶ Das *National Center for Research on Evaluation, Standards, and Student Testing* (CRESST) ist entstanden aus dem Ende der sechziger Jahre gegründeten *Center for the Study of Evaluation* (CSE) der University of California at Los Angeles. CRESST ist ein Zusammenschluss verschiedener amerikanischer Universitäten. Die fünf Programme von CRESST beziehen sich auf die Validität von Qualitätssicherungssystemen im Bildungsbereich, die Entwicklung neuer Instrumente der Beobachtung und Methoden der Forschung. Ein Programm ist speziell

Punkte: Zum einen sind Standards von Unterrichtsfächern sehr schnell für alle Beteiligten eine Überforderung, die wohl mit guten Absichten, zugleich aber auch mit unerreichbaren Zielen zu tun haben (BAKER 2004, S. 152). Zum anderen schwillt mit der Entwicklung die Menge der Standards, also der unverzichtbaren Inhalte, sofort bemerkenswert an, weil keine Autorität vorhanden ist, die Prioritäten setzen könnte. Zudem ist ausgeschlossen, dass Experten für Standards sich je einigen könnten, wenn schon die grundlegenden Sichtweisen unvereinbar sind. Und die Frage der Standards selbst ist strittig: Warum sollten wir inhaltliche Standards in zentralen Wissensbereichen der Schule vorgeben und rigoros abverlangen,

„when in the true spirit of constructivism, students should reinvent algebra if it is warranted by applied problem-solving tasks?“ (ebd., S. 154)

Andererseits reagieren Schulen und Schulleitungen auf die Sprache des Konstruktivismus nicht so, wie die Experten sich das vorstellen. Für die Praxis sendet die Sprache „signals of laxity“ (ebd.), also das Absenken der Anforderungen, auch weil kaum ein Experte in Rechnung stellt, dass alle praktischen Vorschläge wie „Problemlösung“, „Teamwork“ oder das „self management“ des Lernens seit Beginn des 20. Jahrhunderts bekannt sind (DIESBERGEN 2000) und das System nicht grundlegend verändert haben. Die Experten kalkulieren die „hohen Standards“ oft am System vorbei, und das System weist Standards zurück, wenn nicht verstanden wird, was ihren didaktischen Mehrwert ausmachen soll oder weil praktikable Modelle der Implementation fehlen.

Es hilft nichts, bei einer Überfülle von inhaltlichen Standards mit Bündelungen zu arbeiten, die auf ein- und dieselben Lernleistungen oder Kompetenzen abzielen. Daraus resultieren nur Standard-Klumpen (clumps), die dem ähnlich sind, was wir bereits haben. Viele Wege führen zu einem Ziel, das sich nicht genauer bestimmen lässt, einfach weil zuviel offen gelassen wird. Schliesslich führt auch die Rede der „Vernetzung“ (alignment) aller Ebenen und Bereiche nur dazu, sich nicht festlegen zu müssen. Und dann erscheint wiederum alles gleich wichtig,

- die Qualität der Lehrkräfte,
 - die Ressourcen des Unterrichts,
 - die soziale Kohärenz im Klassenzimmer,
 - das Klima der Schule,
 - die Ergebnisse des Lernens
 - oder das Engagement der Eltern
- (BAKER 2004, S. 156).

Letztlich lässt sich kein Faktor gegenüber allen anderen determinieren, und das bremst jede Innovation aus (ebd., S. 157). Ohne Priorität und Festlegung kann nichts wirklich entwickelt werden.

Die hohen Investitionen in die Beschreibung der Ergebnisse führen leicht dazu, die Inhalte zu vernachlässigen und nur noch die *Testergebnisse* zu beachten, einschliesslich der dazu nötigen Motivierungstechniken. Damit werden kurzfristige Erfolge gesucht und der langfristige Aufbau der Kompetenz in schulischen Lernbereichen (domains) vernachlässigt. Je mehr man auf diese Weise testet, desto weniger stabil sind die Resultate. Der Sinn des Testens ist die Kontrolle der Resultate. Damit verbunden ist folgende Erwartungshaltung:

auf die Entwicklung leistungsschwacher Schüler ausgerichtet, und ein weiteres dient der Kommunikation der Forschungsergebnisse.

- (1) You know what you want;
- (2) you measure it well;
- (3) you make reasonable choices based on the interpretations of results;
- (4) everyone knows the consequences of the system;
- (5) management accounts for key system variables; and
- (6) the outcomes improve (ebd., S. 158).

Wer das Erreichen von Standards messen will, muss, was immer sie sind, Standards präzise und akkurat formulieren, sie mit den richtigen Mitteln versehen und auf verschiedene Typen von Lernern ausrichten. Die bisherigen Erfahrungen zeigen, dass Lehrmittel wie Tests entweder auf die *besten* oder die *durchschnittlichen* Schüler ausgerichtet sind. Die schwierigste Gruppe, die *leistungsschwachen* Schüler, werden regelmässig nicht erfasst. Schwache Schüler wechseln in den Vereinigten Staaten oft die Schule, haben begrenzte Kenntnisse der Unterrichtssprache, erhalten vielfach schlechtere Lehrer und werden mit geringeren Ressourcen unterrichtet. Wer das Schulsystem ernsthaft von den Resultaten aus entwickeln will und zugleich Förderabsichten vertritt, muss dieses Problem lösen können, was derzeit nicht der Fall ist (ebd., S. 159).

Nicht nur die amerikanische Bildungspolitik hat drei Probleme nachhaltig unterschätzt: Bildungsstandards sind nicht einfach nur Festlegungen, sondern auch Entwicklungsaufgaben, die am Ende zu unterschiedlichen Qualitäten führen können. Es gibt bessere und schlechtere Lösungen eines schwierigen Problems, nicht etwa einheitlich gute Standards. Zweitens sind auch die Beschreibungen der Resultate und die ist die Bearbeitung der Folgen an verschiedenen Orten unterschiedlich gut. Und drittens richten sich die Lehrkräfte und die Schüler kognitiv und vor allem motivational darauf ein, dass und wie getestet wird (O'NEIL/ABEDI/MIYOSHI/MASTERGEORGE 2004).⁷ Tests sind nicht einfach nur Leistungsmessungen, sondern selbst Lernobjekt, das von den Schülern mehr oder weniger ernst genommen wird. Sie lernen, sich regelmässig auf Tests einzustellen, so wie sie gelernt haben, regelmässig Klassenarbeiten zu schreiben und am Ende eine Durchschnittsnote entgegen zu nehmen.

EVA BAKER argumentiert nicht gegen, sondern *für* Bildungsstandards, allerdings solche, die der modernen Testpsychologie entsprechen und die frei gehalten werden von einer ambitionierten pädagogischen Reformsprache, die sich lediglich in der Bildungspolitik nutzen lässt. Die politische Diskussion bemüht Metaphern und unterstellt Realitäten, ohne eine kritische Differenz zwischen Sprache und Praxis anzunehmen. Metaphern werden nicht sprachlich verstanden, sondern als Projekte, die in einem pädagogischen Programm zielgesteuert Realität annehmen. Das ist nichts Neues, seit der Barockdidaktik des 17. Jahrhundert sind für die Promotion von Erziehungsreformen immer Anpreisungssysteme verwendet worden, die als möglich und machbar erscheinen liessen, was noch gar nicht entwickelt war. Man denke an die Geschichte des „Nürnberger Trichters“, die sich bis in die Metaphorik des „Input“ und „Output“ fortsetzt.

Speziell kritisiert werden drei Metaphern, die tatsächlich bis in die PISA-Philosophie hinein diskursbestimmend sind, nämlich

- All Children Can Learn
- High Standards, Aligned Systems

⁷ Vermutet wird ein „senioritis-effect“, also eine Abnahme der Motivation am Ende der Schulzeit (O'NEIL/ABEDI/MIYOSHI/MASTERGEORGE 2004, S. 199), was zum Beispiel für die PISA-Studie, die Leistungen am Ende der Schulzeit testet, erhebliche Konsequenzen hätte.

- Management by Results.

Die erste Metapher ist von der Bush-Administration im Jahre 2001 mit dem Bundesgesetz „No Child Left Behind“ präzisiert worden. Die politische Idee ist, dass alle Kinder am Ende der zwölfjährigen High-School ein gleiches Bildungsminimum in zentralen Fächern erreichen können. Aber „alle Kinder“ ist tatsächlich nur eine metaphorische Grösse. Es ist ausgeschlossen, dass jedes Kind die Schule auf einer einheitlichen und hohen Stufe der Kompetenz abschliessen wird; höchstens kann es darum gehen, ein gleiches Minimum zu sichern, und auch das wird sich nie einheitlich erreichen lassen.

Die zweite Metapher bezieht sich darauf, dass hohe Leistungsanforderungen gestellt werden können, wenn und soweit das Schulsystem auf einer Linie darauf ausgerichtet ist. Aber auch das ist mehr Wunsch als Realität (BAKER 2004, S. 153), weil hinter der Sprache der Reform keine einheitliche oder gleichsinnige Praxis angenommen werden kann oder darf. Die dritte Metapher ist mit globalen Interpretationen verbunden, die etwa die PISA-Diskussion nach sich gezogen hat, also Interpretationen wie „wir sind gut“ oder „wir sind schlecht,“ ohne geeignete Mittel zu haben, wie aus Daten die geeigneten Konsequenzen gezogen werden können. Man kann Lehrer auszeichnen, die besten Schüler belohnen, Schulen als „gut“ oder „schlecht“ klassifizieren und je nachdem „Entwicklungsanreize“ schaffen, ohne damit schon über effektive Strategien des Wandels und der Förderung zu verfügen (ebd., S. 158).

Die EDK hat eine Art Dauer-PISA vor Augen, wenn sie nationale Bildungsstandards entwickeln lässt. Die Kernfrage wird auch in der Schweiz sein, ob die starke Ausrichtung auf Tests die Qualität von Schule und Unterrichts tatsächlich verbessert oder ob die Schulgeschichte nicht besser beraten war, auf sehr strikte Kontrollen zu verzichten und die Effizienzfrage *nicht* in den Mittelpunkt zu stellen. Ein reines *Teaching for Testing* kann niemand wollen, andererseits sind faire Vergleiche ohne Tests in irgendeiner Form nicht zu haben. Wenn wirklich Leistung und Qualität die Beurteilung der Schule bestimmen soll, und nicht Ideologien und Partikularinteressen, dann sind neue Instrumente unerlässlich, die Risiken mit sich bringen, auf die Schweizer Schule angesichts ihrer Geschichte bislang kaum vorbereitet ist.

Generell wird die Frage des Testens zu einer Schlüsselfrage des Systems, wobei die amerikanischen *Test-Wars* nicht unbedingt den Massstab für die europäische Entwicklung abgeben sollten. Amerikanische Erfahrungen zeigen, dass der Zusammenhang zwischen Standardisierung und Testing problematisch ist (ORFIELD/KORNHABER 2001). Rigide Testprogramme können die soziale Ungleichheit verstärken und die ohnehin gegebene Benachteiligung bestimmter Gruppen erhöhen, gute Testprogramme müssen aufwendig entwickelt werden, die interne Kommunikation der Resultate ist ein andauerndes Problem (ebd., S. 149f.) und der Aufwand macht nur Sinn, wenn er bei den politischen Entscheidungsprozessen auch genutzt wird. Datenbezogene Bildungspolitik ist aber immer noch die Ausnahme (ebd., S. 155). Letztlich ist also diese Art von Qualitätssicherung selbst ein Qualitätsproblem.⁸

Aber Bildungsstandards im Sinne einer gestuften Präzisierung der Erwartungen sind in einer demokratischen Leistungsschule unverzichtbar, wobei nicht einfach die Ziele des Lehrplans erneuert werden dürfen, sondern tatsächlich kompetenzbezogene und gestufte Standards für jedes einzelne Schulfach entwickelt werden müssen, einhergehend mit genauen Bestimmungen, was unter überfachlichen Kompetenzen bestimmt werden soll (GROB/MAAG-

⁸ Sie auch SACKS (1999) und MCNEIL (2000).

MERKI 2001). Ohne Standards ist es sinnlos, von einem Bildungsmonitoring zu reden, man würde Kompetenzen zu testen, die das Feld gar nicht als Zielsetzung verfolgt hat (OELKERS 2003, S. 85ff.). Viele Tests testen nicht, was sie vorgeben, wenn ernsthaft Tests die Entwicklung steuern sollen, dann müssen sie mit klaren inhaltlichen Anforderungen verknüpft sein, kognitive Ansprüche stellen und sich auf Leistungen beziehen.

Aber das verlangt:

- Sensibilität für den tatsächlich geleisteten Unterricht: Der Test muss erfassen, dass der gute Unterricht den Unterschied macht;
- hohe Qualität der technischen Seite des Testens;
- langfristige Anlage der Tests;
- klare und exklusive Ziele/Standards, keine vagen Additionen;
- Evidenz, dass der Test für verschiedene Zwecke und jede Subgruppe von Schülern geeignet ist
(BAKER 2004, S. 162).

Das ist internationalen Vergleich offenbar nicht die Testrealität. Oft werden Tests verwendet, die gleich hohe Massstäbe für alle anlegen, statt hohe, aber differenzierte Standards zu testen (LINN 2000, 2003). Die Prämisse in der Anlage des Tests ist vielfach nicht, dass es individuelle Unterschiede im Entwicklungsstand der Schüler gibt, sondern dass sie alle das gleiche Ziel hätten erreichen können, was nicht der Fall ist und sein kann (LAM 2004, S. 141). Ohne Berücksichtigung des sozialen und kulturellen Hintergrundes der Schüler geht die Testkonstruktion an den Realitäten vorbei (so schon PORTER 1993).

Die technische Seite des Problems lässt sich entwickeln, auch weil genügend Erfahrungen vorliegen, was besser *nicht* getan werden sollte. Die Akzeptanz des Bildungsmonitoring in der Lehrerschaft wird erheblich mehr Mühe bereiten. Die Schulleitungen müssen lernen, mit Testergebnissen umzugehen, sie nicht in der Schublade verschwinden zu lassen und mit ihnen die Entwicklung nachhaltig zu gestalten. Das ist leichter formuliert als getan, vor allem wenn regelmässig Daten auf die Lehrkräfte zukommen, und sie lernen müssen, sie auf ihr Arbeitswissen zu beziehen, auch weil sie sie nicht ignorieren können.

In meinem letzten Teil stelle ich Fragen der Implementation, die heute noch weitgehend übersehen werden, aber an denen sich zentrale Fragen wie die Förderns angesichts erhöhter und an Standards orientierter Leistungsanforderungen entscheiden werden. Was geschieht mit den Schülerinnen und Schülern, die den Mittelschicht-Normen der Tests nicht entsprechen und bereits mit Eintritt in die Schule benachteiligt sind? Amerikanische Erfahrungen zeigen deutlich, dass rigoroses Testen der gegebenen Qualität die soziale Differenz verstärkt und die leistungsschwachen Schüler nochmals benachteiligt (HARGREAVES 2003, S. 88ff.). Wie also geht man mit dem Matthäus-Effekt um?

Es gibt auf diese Frage drei Antworten abseits der Testpsychologie, die mich abschliessend beschäftigen werden. Die Antworten gehen ihrerseits aus von drei Fragen: Haben wir die Bildungsforschung, die notwendig ist, das Bildungssystem zu entwickeln? Gibt es gute Modelle der Evaluation? Und: Wie lässt sich Leistungsorientierung mit Förderung verbinden? Ich verlasse also das Testen und komme auf die Entwicklung der Schule zurück. Was wirkt, sind nicht nur Tests, wie man gelegentlich den Eindruck erhält.

3. Fragen der Forschung und Implementation

Es genügt nicht, einfach die Bildungsforschung mit Qualitätssicherung zu beauftragen. Auch die Forschung übersieht oft die eigensinnige Struktur der Schule und offeriert Lösungen, die schon im Ansatz an der Praxis vorbeigehen. Das Problem ist nicht nur, dass Lehrkräfte gut unterrichten können, ohne sich ständig mit Bildungsforschung zu befassen; der vorgängige Typus von Forschung legt praktischen Nutzen gar nicht nahe, weil das Ziel die wissenschaftliche Publikation ist und nicht die Verbesserung der Praxis. Selbst gute Vorschläge verschwinden in der Flut der Veröffentlichungen, sie sind interessant, vielleicht sogar für die Praxis, aber sicher nicht folgenreich für Schule und Unterricht. Nicht zufällig interessieren die Daten und Aussagen von OECD-Berichten die Bildungspolitik, nicht oder nur am Rande die Lehrkräfte, und dies nicht nur aus dem Grunde, weil seit GEORG PICHT mit solchen Berichten ständig „Bildungskatastrophen“ nahe gelegt werden.

Das amerikanische Stichwort für die Neuausrichtung der Forschung lautet „use-inspired basic research.“ Der Ausdruck geht auf den amerikanischen Politologen DONALD STOKES (1997) zurück und hat in angelsächsischen Bereich zu einer intensiven Diskussion geführt, die inzwischen auch die offizielle Forschungspolitik zu bestimmen beginnt. Angestrebt wird eine Forschung, die nicht in den Journalen verschwindet, sondern für präzise definierte Anwendungsbereiche Nutzen hat und diesen auch nachweist. Das Thema ist nunmehr, *wie* - mit welchen Instrumenten oder Tools - die Praxis der Schule und die Kompetenz der Lehrkräfte gezielt verbessert werden kann, was Objekt einer eigenen Forschung sein muss. Wir müssen beobachten, was wirkt, während eine ganze Didaktikindustrie straflos Wirksamkeit behaupten kann, ohne sie je nachweisen zu müssen. Soll sich das ändern, müssen die Regeln der Forschung geändert werden.

Dabei geht es nicht um die „Anwendung“ von Ergebnissen, sondern um die Erforschung neuer oder anderer Wege des Unterrichtens. Lehrkräfte werden nicht einfach Standards für guten Unterricht erreichen, wenn wir nicht wissen, welche Qualität ihr Coaching haben muss. Und die leistungsschwachen Schüler werden nicht gehaltvoll Lesen lernen, wenn es dafür keine neuen, forschungsbezogenen Verfahren gibt, die speziell für diese Gruppe entwickelt wurden. Solche Entwicklungen geschehen heute ohne Forschung, meistens über staatliche Massnahmen zur Fortbildung, die eine Art Erfahrungsaustausch darstellen und nicht etwa auf *best practice* bezogen sind. Welche Lehrkraft für welches Problem über das beste Know How verfügt, wissen in der Regel weder die Schulleitungen noch die Schulkommissionen, weil Daten fehlen. Dass die gewählten Massnahmen zur Fortbildung *nicht* forschungsgestützt sind, macht geradezu ihren Wert aus.

Das muss nicht so sein. Ich gebe abschliessend einige Beispiele, wie sich praxisnahe Forschung mit Schulentwicklung verbinden lässt und wie Probleme der Implementation etwa besserer Fördersysteme bearbeitet werden können. Der medizinische Ausdruck „Implementation“ hat einen wahren Kern, er verweist darauf, dass auch Schulen über ein Immunsystem verfügen. Schulen sind aus guten Gründen eher konservative Institutionen, die nicht jeder pädagogischen Mode nachjagen, sondern die vom Bewährten ausgehen. Das kann nicht einfach „träge“ genannt werden, sondern ist die Folge von bislang nicht überbotenen Problemlösungen. In diesem Sinne ist es ein kardinales Problem der Schulreform, wann und unter welchen Umständen die Schule ein *neues* Problem akzeptiert und eine neue Lösung annimmt.

Eine der neuen Forderungen ist die regelmässige Evaluation der Qualität einer Schule. Das Konzept der *accountability*, also der Verantwortung der Schule für das Zustandekommen von Qualität, geht auf einen Aufsatz zurück, den HARRY LEVIN⁹ im Jahre 1974 in der amerikanischen Zeitschrift *School Review* veröffentlichte. Seitdem ist das Thema politisch stark aufgewertet worden, ohne an dem einfachen Modell viel zu verändern. Es umfasst¹⁰

1. *Ziele*: Erreichbare Endzustände eines Prozesses.
2. *Mittel*: Massnahmen, die sicherstellen, dass der begonnene Prozess in Richtung Ziele verläuft.
3. *Feedbackschleifen*: Regelmässige Überprüfung der erreichten Qualität.
4. *Folgen*: Anpassung des Prozesses in Richtung Feedback-Daten.

Das ist die Basisidee hinter dem, was heute „Evaluation“ genannt wird. Evaluieren werden können die Leistungen der Schüler, die Kompetenz der Lehrkräfte oder die Qualität der Schulen. Zugrunde gelegt wird jeweils das Modell von LEVIN. „Qualität“ wird nicht mit abschliessenden Noten, sondern mit Hilfe von Feedbackschleifen *im* Prozess beschrieben. Damit wird es theoretisch möglich, Lernen nicht einfach auf Ziele hin auszurichten, sondern mit Ergebnissen den Prozess zu beeinflussen. Es ist im Sinne dieses Prozessmodells falsch oder irreführend, aus allgemeinen Lernzielen operative abzuleiten und dabei stehen zu bleiben; entscheidend ist, die tatsächlichen Erfahrungen einschätzen und korrigieren zu können.

Die Entwicklung einer Schule muss vor dem Hintergrund einer fortlaufenden Evaluationen ihrer Stärken und Schwächen erfolgen, Lehrkräfte und Schulleitung müssen bereit sein, sich auf eine Analyse ihrer Schwächen einzulassen, was als einer der heiklen Punkte der gesamten Qualitätsdiskussion verstanden werden muss. Hier setzt ein Verfahren an, das im Kanton Zürich auf der Basis der Ideen von LEVIN entwickelt worden ist. Die Idee hat noch nicht zwischen „internen“ und „externen“ Evaluationen unterschieden, was inzwischen aber die Norm ist. Das Zürcher Projekt heisst „Neue Schulaufsicht“ und ist Teil der Qualitätssicherung des öffentlichen Dienstes mit Hilfe externer Evaluationen. Das Verfahren basiert auf dem Modell der „Peer-Reviews“, also der Besichtigung von Schulen durch externe Experten.¹¹

Wenn der Zürcher Soverän das neue Volksschulgesetz im zweiten Anlauf annimmt, was nicht unmöglich ist, wird die „Neue Schulaufsicht“ vom versuch zur Dauereinrichtung.

⁹ Harry M. Levin ist derzeit William Heard Kilpatrick Professor of Economics and Education am Teachers College der Columbia University. Er leitet das *National Center for the Study of Privatization in Education* (NCSPE).

¹⁰ „An accountability system is a closed loop reflecting a chain of responses to perceived needs or demands; an activity or set of activities that emerges to fill those demands; outcomes that result from those activities; and feedback on outcomes to the source of the demands. The feedback may generate new demands or a regeneration of the old ones; in either case, the previous set of activities may be modified or remain intact; a new or altered set of activities may be produced; and the loop is completed again with feedback to the source of the demands“ (LEVIN 1974, S. 375).

¹¹ Die Evaluationsteams setzen sich aus Fachpersonen zusammen, die sämtlich über Lehrpatente und Schulerfahrung verfügen müssen, anders würden sie im Feld nicht akzeptiert. Man kann Evaluationsaufträge nicht einfach Evaluationsbüros übertragen, sondern muss sehr sorgfältig auf die Nähe zum Berufsfeld achten, weil sonst sofort eine Glaubwürdigkeitslücke entsteht. Schulen sollten von Lehrkräften evaluiert werden, wengleich solchen, die auch über weitergehende Berufserfahrungen verfügen und die kompetent sind für das Evaluationsgeschäft. Wenn Schulen lernen sollen, ihre Karten offen zu legen und sich so transparent wie möglich darzustellen, müssen sie feldspezifische Kompetenz voraussetzen.

Alle Schulen des Kantons werden dann in regelmässigen Abständen extern evaluiert. Die Berichte der Experten beschreiben Stärken und Schwächen, sie dienen der weiteren Entwicklung der Schulen und werden in Zielvereinbarungen umgesetzt. Das Konzept verlangt drei hauptsächliche Akteure, nämlich Evaluationsteams, die Schulpflege in den Gemeinden und eine professionelle Schulleitung. Nur mit Schulleitungen neuer Art lässt sich eine längerfristige Entwicklung auf der Basis von Zielsteuerungen erreichen, die verantwortlich umgesetzt werden müssen, also nicht lediglich einen stillschweigenden kollegialen Konsens voraussetzen. Die Evaluation selbst besteht aus zwei Teilen, einer Selbstevaluation der Schule und einer externen Evaluation durch die Teams der „Neuen Schulaufsicht.“

Das Modell der „Neuen Schulaufsicht“¹² im Kanton Zürich ist bislang mit über sechzig Schulen ausprobiert worden, die sich sämtlich freiwillig beteiligt haben. Die „Neue Schulaufsicht“ ist ihrerseits evaluiert worden. Die Daten zeigen, dass ein solches Verfahren vor allem auf Zustimmung bei den Schulen stösst, die sich ernst genommen fühlen und die erfahren, dass und wie sie von den ursprünglich gefürchteten externen Evaluationen profitieren. Die Evaluation dieses 1999 begonnenen Versuchs erbrachte folgende Resultate:

- Die Schulen befürworten das Projekt und stufen es als Qualitätsgewinn ein. Die Verfahren der Datenerhebung werden als zweckmässig eingeschätzt und die Ergebnisse erhalten hohe Glaubwürdigkeit.
- Schwierigkeiten bereitet noch die genaue Festlegung der Folgen, die Evaluationsberichte sind nicht in jedem Falle Grundlage für wirksame Zielvereinbarungen und so für eine gezielte Entwicklungspolitik der einzelnen Schule.
- Ein Manko ist auch, dass der Unterricht vor allem aus Zeitgründen zu wenig berücksichtigt wird. Die Qualitätseinschätzung betrifft also den Kernbereich eher nur marginal (BINDER/TRACHSLER 2002).

Daraus folgte die Verbesserung des Verfahrens bis hin zur Entwicklung eines Handbuchs für die präzise Beschreibung der Kriterien der Evaluation, die letzten Endes den professionellen Wert der Beschreibungen bestimmen. Auch diese Kriterien sind forschungsgestützt, wobei sich wiederum zeigt, welches Problem der Komplexitätssteigerung gegeben ist, wenn ernsthaft der Forschungsstand eine Rolle spielen soll.

Aber das Beispiel zeigt auch, dass neue Probleme von den Schulen akzeptiert werden, wenn sie mit aufschlussreichen und nützlichen Lösungen verbunden sind. Sie dürfen nicht einfach erfahren, was sie ohnehin wissen, und was sie durch Evaluation erfahren, muss sich umsetzen lassen, und dies möglichst unmittelbar nach Kommunikation der Ergebnisse. Das gilt generell; wenn sich Datenaufnahmen gleich welcher Art *nicht* mit Nutzen verbinden, werden sie von der Praxis ignoriert. Wer also Schulen mit neuen Problemen und Aufgaben konfrontiert, muss imstande sein, praktikable Lösungen anzubieten. Und nicht nur das,

¹² Das Verfahren und seine Rahmenbedingungen lassen sich allgemein so beschreiben: Die Teams der „neuen Schulaufsicht“ sind unabhängig, sie arbeiten im Auftrag der Behörde, aber sind nicht der Hierarchie der Behörde unterstellt. Die Evaluationen werden mit höchstmöglicher Transparenz der Kriterien und Verfahren durchgeführt, die vorher offen gelegt werden. Die Schulen erstellen Selbstevaluationsberichte, die Evaluationsteams besuchen die Schulen, erheben Daten und legen ihrerseits Berichte vor, die die Grundlage sind für das weitere Vorgehen. Die Evaluationsberichte werden im Kollegium sowie mit den Eltern und Schülern ziel- und entwicklungsbezogen diskutiert. Schulleitung und Schulpflege vereinbaren auf dieser Basis die Entwicklungsziele und besondere Massnahmen für den nächsten Berichtszeitraum.

Schulen kann man begeistern, wenn man originell und kreative Lösungen vorlegt, die das *working knowledge* spürbar verbessern.

Aus der Schulgeschichte ist bekannt, dass neben wirksamen Verfahren die Entwicklung brauchbarer Instrumente immer einen wesentlichen Teil der Implementation ausgemacht hat. FRIEDRICH FRÖBELS Theorie des Kindergarten ist bis heute unverständlich, seine Spielgaben waren praktikabel und erklären seinen Einfluss. Die Lehrkräfte mögen fern der Bildungsforschung gut unterrichten können, aber sie werden neue Instrumente für den Unterricht nicht ignorieren, wenn diese die Praxis sichtbar anreichern. Die ganze Diskussion um E-Learning ist bislang an den Schulen weitgehend vorbei gegangen, weil die neue Technik nicht zur gegebenen Raum-Zeit-Struktur des Klassenzimmers passt (CUBAN 2001). Das wird sofort anders, wenn nicht mehr Schiefertafeln, sondern elektronische Wandtafeln das Klassenzimmer bestimmen. Mit ihnen lässt sich der Unterricht, den die Lehrkräfte können, auf technisch neue Weise realisieren, bis dahin, dass die Tafelbilder Schülern, die fehlen, per Mail nach Hause geschickt werden können, und zwar im Unterricht. Sie werden dadurch gefördert, dass sie nichts versäumen.

Eine anderes Beispiel für gelungene Implementation ist der im Kanton St. Gallen erprobte und vom Kanton Zürich nicht ignorierte *Klassenscockpit*,¹³ bekanntlich ein Internet-Angebot für die Bestimmung des Leistungsstandes einer Klasse in bestimmten Fächern. Es handelt sich im Kern um einen Test oder ein Evaluationsinstrument, das den Lehrkräften erlaubt, den Lernerfolg ihrer Klasse mit dem Lernerfolg anderer Klassen zum gleichen Thema im gleichen Fach zu vergleichen.¹⁴ Die Schüler bearbeiten standardisierte Leistungstests, die von den Lehrkräften korrigiert werden. Die Lösungen der einzelnen Schüler werden mit Hilfe einer Eingabemaske über das Internet in die Datenbank des *Klassenscockpit* eingelesen. Die Datenbank enthält die Daten anderer Klassen, die einen kantonalen Durchschnitt bilden, so dass ein Vergleich der Leistungen im Prozess möglich wird, was bislang nie der Fall war. Im Blick auf diesen Vergleich weiss man dann, wo man steht.

Warum war diese Implementation erfolgreich? Das Instrument lässt sich für den täglichen Unterricht nutzen, die Zustimmung der Lehrkräfte ist gross, die Kritik betrifft eher den noch restriktiven Einsatz des Instruments. Die Vergleichsdaten können klassenspezifisch wie schülerspezifisch abgerufen werden, die Eingabe ist anonym. In Befragungen zeigt sich, dass die Lehrerinnen und Lehrer den Test und seine Auswertung nicht nur zur Standortbestimmung nutzen wollen, sondern auch zur Anpassung der Schülerbeurteilung, zur Planung der Übertritte von der Primar- in die Sekundarschule und nicht zuletzt zur Optimierung des eigenen Unterrichts (MOSER 2003). Der Grund für diese hohe Akzeptanz ist nicht nur die Nützlichkeit des Instruments, sondern auch der Einsatz von Vergleichstests, die *nicht* für die schulische Selektion genutzt werden. Die Tests sollen leistungsfördernd wirken, die Förderung des Leistungsverhaltens setzt voraus, dass der Leistungsstand der Klasse eingeschätzt werden kann, ohne dass die Bezugsnorm allein die Normalverteilung der jeweiligen Klasse wäre.

¹³ <http://www.klassenscockpit.ch>

¹⁴ Derzeit werden ausgewählte Bereiche in den Fächern Mathematik und Deutsch angeboten. Anhand von Aufgabensätzen (Orientierungsmodulen), die dreimal jährlich zur Verfügung gestellt und im Internet ausgewertet werden, können die Lehrkräfte die Leistungen ihrer Klasse im Blick auf diese Module vergleichen. Derzeit bietet *Klassenscockpit* von der dritten bis zur neunten Klasse solche Orientierungsmodule an, auf der Oberstufe (Sekundarstufe I) sind zwei Niveaus erhältlich. *Klassenscockpit* wird bereits in verschiedenen Kantonen angewendet, darunter demnächst auch im Kanton Zürich.

Gezielte Förderung in diesem Sinne steht nicht im Widerspruch zu Tests und der Implementation neuer Verfahren und Instrumente, eher im Gegenteil. Die Förderung setzt Datenerhebungen voraus, die auch zur genauen Problembestimmung eingesetzt werden. Ein zentrales Problem ist die Schere im Leistungsverhalten der Schülerinnen und Schüler, sobald fachbezogene Leistungsanforderungen einsetzen, also etwa mit der vierten Klasse (MOSER/RHYN 2000; MOSER/KELLER/TRESCH 2002). Eines der grossen Probleme wird sein, diesen Schereneffekt zu minimieren,¹⁵ was nur mit wirksamen Förderprogrammen möglich ist. Dazu kommen die bekannten Probleme der geschlechtsdifferierten Leistungen und der offenbar frühen Domänenwahl, also die Festlegung der Interessen für oder gegen bestimmte Fächer, die anschliessend offenbar nur noch schwer zu korrigieren ist. „Fördern“ heisst, nicht auf das Ergebnis zu warten, sondern früh genug einzugreifen, ohne dass der Zeitpunkt sicher wäre. Das setzt die Lehrkräfte unter Druck, sie müssen Massnahmen ergreifen und können sich in der Regel nur auf sich selbst verlassen.

Auch das lässt sich verbessern. Wirksame Programme zur Förderung setzen rechtzeitiges Erkennen der Leistungsdifferenzen voraus, was nur dann zu erreichen ist,

- wenn früh nicht-selektive Tests eingesetzt werden,
- flexible Lektionpools zur Verfügung stehen,
- eine kompetente Schulleitung für Abstimmung und Prioritätenbildung sorgen kann,
- verbindliche Standards vorhanden sind
- und die Qualität der Schulen sich am Ergebnis bemisst.

Das ist schwer genug, bedenkt man, dass die Kompetenzen der Schülerinnen und Schüler in den entscheidenden Leistungsfächern bereits bei Schuleintritt weit auseinander liegen (STAMM/MOSER/HOLLENWEGER 2004) und jede Klasse eine Leistungshierarchie entwickelt, deren Rangverteilung auch beim Wechsel der Lehrkraft mit hoher Wahrscheinlichkeit erhalten bleibt.

Ein weiteres Förderprojekt in der Schweiz heisst „Stellwerk“. Es dient der Bestimmung des individuellen Standortes der Schülerinnen und Schüler Mitte des achten Schuljahres. Mit Hilfe einer Analyse des Leistungsstandes jedes einzelnen Schülers soll ermittelt werden, wie und in welchen Bereichen die Schüler bis zum Ende des 9. Schuljahres gezielt gefördert werden können, um das Niveau zu erreichen, eine Lehrstelle finden zu können. Dieses Förderprogramm ersetzt in Teilen den Lehrplan, der zugunsten von gezielten Nachbesserungen in bestimmten Kompetenzbereichen gelockert oder suspendiert wird. Es ist illusorisch zu erwarten, dass die Schülerinnen und Schüler bis Mitte der 8. Klasse einen angeglichenen Leistungsstand erreichen werden, was immer man tun kann, die Öffnung der Schere zu verringern. Aber die Schule kann mit gezielten Förderprogrammen die Chancen auf dem Arbeitsmarkt verbessern.

Im Kanton Zürich wie überhaupt in der Schweiz sind vor allem grössere Firmen dazu übergegangen, Lehrlinge mit eigenen Tests und Assessments auszuwählen (MOSER 2004). Der Volksschulabschluss ist dadurch systematisch entwertet worden, heute bekommen Schulabgänger eine Lehrstelle nicht mehr, weil sie gute Schulnoten nachweisen können, sondern weil sie firmeneigenen Aufnahmeprüfungen bestanden haben. Dabei werden Tests verwendet, die private Büros entwickelt haben und auf die die Schulen keinerlei Einfluss

¹⁵ Der Effekt hat nicht nur mit den Herkunftsmilieus der Schülerinnen und Schüler zu tun, die auch in den Schweizer PISA-Daten als zentraler, aber nicht einziger Faktor für das Zustandekommen der Leistungen nachgewiesen ist (Bildungsmonitoring 2002).

nehmen können. Das Projekt „Stellwerk“ soll dieser Entwertung der allgemeinbildenden Schule entgegenarbeiten, indem die für den Bewerbungsmarathon und die Lehren selbst erforderlichen Kompetenzen nachgebessert werden. Das Projekt dient vor allem den leistungsschwächeren Schülern.

Zur Qualitätssicherung und Förderung auf Seiten der Lehrkräfte sind verschiedene Ideen entwickelt worden, darunter eine, die mit *best practice* zu tun hat. Lehrerinnen und Lehrer unterscheiden sich hinsichtlich ihrer Qualität zum Teil erheblich, jeder weiss, dass es bessere und schlechtere Lehrkräfte gibt, nur dass dies bislang ein Tabuthema war, an das nicht gerüttelt werden sollte. Der Schweizer Think-Tank *Avenir Suisse* hat eine Studie in Auftrag gegeben, die ausgehend von Leistungsmessungen die Unterschiede zwischen den Lehrkräften darstellt, also bessere und schlechtere Leistungen der Schüler auf die Kompetenz der Unterrichtenden zurückführt (MOSER/TRESCH 2003). Aus dieser Studie ist ein weiteres Projekt¹⁶ entstanden, mit dem die Vergleichsdaten ins Feld rückübersetzt werden sollen. Die Lehrkräfte werden mit ihren persönlichen Daten konfrontiert und erhalten gezielt Unterstützung, ihre Qualität zu verbessern. Ohne solche Strategien der Implementation wird die Forschung die Lehrkräfte nie erreichen, als das grösste Manko nie überwunden werden.

Ich fasse zusammen: Was wir „neue Steuerung“ nennen, ist nicht auf Bildungsstandards zu reduzieren, sondern verlangt eine umfassende Perspektive der Schulentwicklung. Die Veränderung der Organisation ist weitgehend abgeschlossen, fast alle Kantone haben geleitete Schulen eingeführt oder entwickelt, Leitbilder gibt es vermutlich überall, Massnahmen zur Qualitätssicherung wie die Entwicklung von Teams in den Schulhäusern oder von Lernprogrammen im Unterricht sind an vielen Stellen realisiert worden. Wenn jetzt die Bildungsstandards im Mittelpunkt stehen, dann weil dies der nächste Schritt sein muss. Die Bestimmung der Inhalte hat in der Organisationsentwicklung kaum eine Rolle gespielt, aber hier liegt der Kern des Unterrichts. Das New Public Management hat die Schule bereichert, aber nur weil die Prinzipien erfolgreich angepasst worden sind.

Das wird beim Bildungsmonitoring auch der Fall sein, wenn Praxis erreicht werden soll. Tests sind notwendig, aber ein Teaching-for-Testing muss vermieden werden. Die Testpsychologie hat dienenden Charakter und verlangt Implementationen eigener Art. Das wäre meine hauptsächliche Botschaft: Wenn die Praxis verändert werden soll, ist der Aufbau tauglicher Verfahren und Instrumente notwendig. Dazu gehören curriculare Standards anstelle untauglicher Lehrpläne, aber auch neue Formen der Diagnostik und Evaluation sowie nicht zuletzt spezielle Verfahren der Förderung. Ohne Standards weiss man nicht, wohin gefördert werden soll, ohne besondere Massnahmen der Förderung bleibt ein Drittel der Schüler zurück. So einfach ist das Problem, leider nicht die Lösung.

Literatur

BAKER, E.L.: Reforming Education Through Policy Metaphors. In: Th. FITZNER (Hrsg.): Bildungsstandards. Internationale Erfahrungen - Schulentwicklung – Bildungsreform. Bad Boll: Evangelische Akademie 2004, S. 150-163.

Bildungsmonitoring Schweiz: Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Nationaler Bericht der Erhebung PISA 2000. Neuchâtel 2002.

¹⁶ Dissertationsprojekt von SARAH TRESCH am Pädagogischen Institut der Universität Zürich (Fachbereich Allgemeine Pädagogik).

- BINDER, H.-M./TRACHSLER, E.: *wif!* –Projekt “Neue Schulaufsicht an der Volksschule. Externe Evaluation. Luzern 2002.
- COHEN, D.K./RAUDENBUSH, St. W./LOWENBERG BALL, D.: Resources, Instruction, and Research. In: F. MOSTELLER/R. BORUCH (Eds.): *Evidence Matters. Randomized Trials in Education Research*. Washington, D.C.: Brookings Institution Press 2002, S. 80-119.
- CUBAN, L.: *Oversold and Underused. Computers in the Classroom*. Cambridge/Mass., London: Harvard University Press 2001-
- DAVENPORT, T./PRUSAK, L.: *Working Knowledge: How Organisations Manage What They Know*. Boston: Harvard Business School Press 1998.
- DEWEY, J.: *Wie wir denken*. Mit einem Nachwort neu hrsg v. R. HORLACHER/J. OELKERS. Zürich: Verlag Pestalozzianum 2002. (amerik. Orig. 1910)
- DIESBERGEN, C.: *Radikal-konstruktivistische Pädagogik als problematische Konstruktion. Eine Studie zum Radikalen Konstruktivismus und seiner Anwendung in der Pädagogik*. 2. Aufl. Bern et.al.: Peter Lang 2000. (= *Explorationen. Studien zur Erziehungswissenschaft*, hrsg. v. J. OELKERS, Band 22)
- FITZNER, Th. (Hrsg.): *Bildungsstandards. Internationale Erfahrungen - Schulentwicklung - Bildungsreform*. Bad Boll: Evangelische Akademie 2004.
- GROB, A./MAAG-MERKI, K.: *Überfachliche Kompetenzen. Theoretische Grundlegung und empirische Erprobung eines Indikatorensystems*. Bern et. al.2001. (= *Explorationen. Studien zur Erziehungswissenschaft*, hrsg. v. J. OELKERS, Bd. 31)
- HARGREAVES, A.: *Teaching in the Knowledge Society. Education in the Age of Insecurity*. New York/London: Teachers College Press 2003.
- LAM, T.C.M.: *Issues and Strategies in Standards-Based School Reform: The Canadian Experience*. In: Th. FITZNER (Hrsg.): *Bildungsstandards. Internationale Erfahrungen - Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie 2004, S. 103-149.
- LEVIN, H.M.: *A Conceptual Framework for Accountability in Education*. In: *School Review* Vol. 82, No. 3 (May 1974), S. 363-391.
- LINN, R.L.: *Assessment and Accountability*. In: *Educational Researcher* Vol. 29, No. 2 (2000), S. 6-16.
- LINN, R.L.: *Accountability: Responsibility and Reasonable Expectations*. In: *Educational Researcher* Vol. 32, No. 7 (October 2003), S. 3-13.
- MCNEIL, L.: *Contradictions of School Reform. Educational Costs of Standardized Testing*. New York/London: Routledge 2000. (= *Critical Social Thought*, ed. M. APPLE)
- MOSER, U.: *Klass Cockpit im Kanton Zürich. Ergebnisse einer Befragung von Lehrerinnen und Lehrern der 6.Klassen über ihre Erfahrungen im Rahmen der Erprobung von Klass Cockpit im Schuljahr 2002/2003*. Bericht zuhanden der Bildungsdirektion des Kantons Zürich. Ms. Zürich 2003.
- MOSER, U.: *Jugendliche zwischen Schule und Berufsbildung. Eine Evaluation des Übergangs von der obligatorischen Schulbildung in die berufliche Grundbildung bei Schweizer Grossunternehmen unter Berücksichtigung des internationalen Schulleistungsvergleichs PISA*. Ms. Zürich 2004.
- MOSER, U./RHYN, H.: *Lernerfolg in der Primarschule. Eine Evaluation der Leistungen am Ende der Primarschule*. Aarau 2000.
- MOSER, U./KELLER, F./TRESCH, S.: *Evaluation der 3. Primarschulklassen. Schlussbericht zuhanden der Bildungsdirektion des Kantons Zürich*. Ms. Zürich 2002.
- MOSER, U./TRESCH, S.: *Best Practice in der Schule. Von erfolgreichen Lehrerinnen und Lehren lernen*. Buchs: Lehrmittelverlag des Kantons Aarau 2003.
- OELKERS, J.: *Schulreform und Schulkritik*. 2., vollst. überarb. Aufl. Würzburg 2000. (= *Erziehung, Schule, Gesellschaft*, hrsg. v. W. BÖHM u.a., Bd. 1)
- OELKERS, J.: *Wie man Schule entwickelt. Eine bildungspolitische Analyse nach PISA*. Weinheim/Basel/Berlin 2003.

- ORFIELD, G./KORNHABER, M.L. (Eds.): Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education. New York: The Century Foundation Press 2001.
- O'NEIL, H.F./ABEDI, J./MIYOSHI, J./MASTERGEORGE, A.: Monetary Incentives for Low-Stakes Tests. In: Th. FITZNER (Hrsg.): Bildungsstandards. Internationale Erfahrungen - Schulentwicklung - Bildungsreform. Bad Boll: Evangelische Akademie 2004, S. 164-201.
- PORTER, A. C.: School Delivery Standards. In: Educational Researcher Vol. 22, No. 5 (1993), S. 24-30.
- RYLE, G.: The Concept of Mind. London: Hutchinson 1949.
- SACKS, P.: Standardized Mind. The High Prize of America's Testing Culture and What Can We Do to Change It. Cambridge/Mass.: Perseus Publishing 1999.
- STAMM, M./MOSER, U./HOLLENWEGER, J.: Lernstandserhebung in den 1. Klassen des Kantons Zürich. Schlussbericht zuhanden der Bildungsdirektion des Kantons Zürich. Ms. Zürich 2004.
- STOLL, L./EARL, L./FINK, D.: It's about Learning; It's about Time. London: Falmer Press 2002.
- STOKES, D.E.: Pasteur's Quadrant. Basic Science and Technological Innovation. Washington DC: The Brookings Institution Press 1997.
- Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Herausgegeben vom Bundesministerium für Bildung und Forschung. Bonn 2003.