

Jürgen Oelkers

## *Schulentwicklung, Forschung und Bildungspolitik<sup>\*)</sup>*

### *1. Ein pragmatischer Blick auf das Bildungssystem*

Mein Vortrag ist ausgeschrieben unter dem Titel

„Anforderungen an Bildungsforschung und Bildungspolitik im Kontext einer wissensbasierten Weiterentwicklung des Schulsystems“.

Der Titel hat mich bei Abfassung des Vortrages doch ziemlich abgeschreckt. Bekanntlich wird heute jedes Problem „im Kontext“ gesehen, was vermutlich die Lösungswahrscheinlichkeit dramatisch sinken lässt, weil „Kontexte“ bis ins Unendliche dehnbare Grössen sind, denen nie genug hinzugefügt werden kann. Es gibt für keinen „Kontext“ eine Sättigungsgrenze.

Das gilt zumal für die „wissensbasierte Weiterentwicklung des Schulsystems“. Dieser Kontext steigt mit jedem neuen Zeitschriftenjahrgang an, ohne dass sich allein deswegen die „Wissensbasierung“ verbessert. Die Verbreiterung des Wissens macht es nicht sicherer, aber genau das ist die Erwartung. Das Schulsystem muss auf der Basis zunehmend besseren Wissens weiterentwickelt werden, aber das ist bei einer rasch steigenden Forschungsmenge und separaten Befunden selbst ein Problem. Ob die Forschung tatsächlich zu einer grösseren Sicherheit in der Systemsteuerung führt, ist die Frage, nicht die Prämisse.

Folgen wir Niklas Luhmann, dann sind Forschung, Praxis und Politik verschiedenartige Funktionssysteme, die einer je eigenen Logik folgen, die nicht auf andere zurückgeführt werden kann. Natürlich kann das eine System versuchen, „Anforderungen“ für die anderen zu formulieren, die Wahrscheinlichkeit, dass sich dadurch die Funktionslogik ändert, ist jedoch gering. Die Praxis bezieht sich ebenso auf sich selbst wie Forschung und die Politik. Natürlich kann sich die Politik mit positiven Forschungsdaten legitimieren, wird dadurch aber nicht anders.

Luhmanns Geist ist seit zwanzig Jahren aus der Flasche, aber sind soziale Systeme wirklich so „selbstreferent“, wie er zu behaupten nicht müde wurde? Die Grundeinheit seiner Theorie ist das Verhältnis von System und Umwelt, vereinfacht gesagt von innen und aussen. Systeme kommunizieren mit ihren Umwelten zur eigenen Bestandsicherung, das heisst, sie nehmen nur das auf, was verträglich ist und nützlich erscheint. Luhmann hat wenig zur Passung und Utilität von Aussenbeobachtung, aber genau an diesen beiden Stellen wird der Transfer definiert.

---

<sup>\*)</sup> Vortrag auf der Tagung „Wissen für Handeln - Ansätze zur Neugestaltung von Bildungsforschung und Bildungspolitik“ am 18. Juni 2010 in der Universität Hamburg.

Allerdings machen das nicht einfach „Systeme“, sondern Akteure in fragilen Interaktionen unter der Voraussetzung einer sozialen Organisation. Auch „Fragilität“ kommt bei Luhmann nicht vor, Kommunikationen dauern an oder werden abgebrochen, aber das sagt über ihre Stabilität nichts aus. In Schulen müssen Akteure nicht nur ständig Missverständnisse ausräumen, sie sind auch auf besondere Weise aufeinander angewiesen. Unterricht ist aus Sicht der Lehrkräfte immer ein Versuch mit unsicherem Ausgang, weil die andere Seite, die Schülerinnen und Schüler, über den Erfolg bestimmt. Ihre Einstellung zum Lernen ist nur begrenzt beeinflussbar, aber Lehrkräfte müssen genau das können, wenn sie die Erfolgswahrscheinlichkeit ihres Handelns verbessern wollen.

In Schulen findet daher nicht einfach eine ständige „Autopoiesis“ statt, wie Luhmanns Theorie suggeriert. Vielmehr versuchen die Akteure für Probleme Lösungen zu finden, die in ihrer Reichweite liegen. Genauer müsste ich sagen: In den meisten Fällen ist das Problem, wie die bereits vorhandenen Problemlösungen angereichert und verbessert werden. Keine soziale Organisation und erst recht keine Schule beginnt bei Null und jede Innovation trifft auf ein gestaltetes Gegenüber, das nicht unbedingt auf Neuerungen wartet. Und selbst bei grosser Gutwilligkeit ist die Frage immer, wie Innovationen im fortlaufenden Betrieb untergebracht werden können.

- Man tut gut daran, sich Schulen als Organisationen vorzustellen, die wohl lernen, aber nicht einfach auf den nächsten Anlass hin.
- Schulen sind historisch gewachsene Erfahrungsräume mit einem langen Gedächtnis.
- Sie gehen vom Bewährten aus und lernen im Blick darauf.
- „Bewährt“ sind Problemlösungen, die institutionell verankert sind und nicht von jeder Referendargeneration neu erfunden werden.

Bewährt heisst nicht „routiniert“, also dauerhaft unflexibel; die Problemlösungen können nur bestehen, wenn sie in einem kommunikativen Feld immer neu getestet und angepasst werden. Das gilt graduell unterschiedlich. Die Stundentafel löst das Problem der Verteilung von Zeit je nach Rang des Schulfaches, und das ist für die Organisation Schule weit gewichtiger als der Morgenkreis oder die Pausenordnung. Aber auch das sind Problemlösungen, wenngleich solche, die leichter zu verändern sind.

Diese historisch-pragmatische Perspektive fehlt der Systemtheorie, aber sie passt viel besser zur Frage, in welchem Verhältnis Bildungspolitik und Bildungsforschung stehen sollen oder können. Bei Luhmann ist das Verhältnis exklusiv und höchstens klandestin: Die Forschung forscht und die Politik entscheidet; wenn sie sich begegnen, dann zum Zweck der Steigerung des wechselseitigen Unverständnisses. Das Verhältnis wird regiert durch Gesten des ständigen Kopfschüttelns.

In pragmatischer Perspektive ist die Hauptfrage, was Politik und Forschung zur Problembearbeitung beitragen können, ohne unentwegt nur neue Probleme zu erzeugen. Die Geschichte von Reformen ist voll von solchen unglücklichen Serien, die nur die Probleme steigern und nicht die Kapazitäten der Problemlösung. Das hat auch mit der Theoriwahl und der darauf aufbauenden Rhetorik zu tun, der es oft egal ist, ob Reformen nur dazu führen, die Probleme zu verschärfen. Genau das verlängert die Lebensdauer der Rhetorik. Die Rede der „Ganzheitlichkeit“ darf daher um ihrer selbst willen kein Problem lösen, aber das kann für Forschung und Politik nicht gelten.

Grundlegend für meine Überlegungen ist eine historische Variante der massgeblich von George Herbert Mead und John Dewey<sup>1</sup> begründeten Theorie des Problemlösens. Die Variante geht davon aus, dass „Problemlösen“ nicht nur auf die Lernenden, sondern auch auf ihre Institutionen angewendet werden kann.

- Die Optik der Theorie ist zweifach:
- sie richtet sich auf die Serie von Problemlösungen, die die Schule als Institution ausmacht,
- und sie kennzeichnet neue Probleme oder neue Lösungsmöglichkeiten.
- Die Zone des Problematischen ist immer eng begrenzt.
- *Gelöste* Probleme sind die Voraussetzung, dass Unterricht stattfinden kann und Schule nicht ständig neu erfunden werden muss.

Der fragend-entwickelnde Unterricht, heute etwas 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“ oder die Erfindung der Studentafel, des Notenschemas und der organisatorischen Zeiteinheit der Schulstunde.

Die Erfindung der Jahrgangsklassen gehört dazu, der frühe Beginn des Schultages am Morgen statt am Mittag oder auch das Schulbuch als exklusives Lernformat, mit dem vermutlich auch noch die Wikipedia-Generation in Schach gehalten werden kann. Die elektronische Enzyklopädie ersetzt nicht die Steuerung durch ein geordnetes Nacheinander oder einen kontrollierten Lernfortschritt, der auf eine Lerngruppe zugeschnitten ist. Eher wird Wikipedia Teil des Lehrmittels.

Ohne solche erfolgreichen und langfristig stabilen historischen Problemlösungen könnte Schule gar nicht stattfinden. Viele dieser Lösungen betreffen *Basisprobleme* und sind institutionell schwergewichtig. Das gilt für

- die Aufteilung der Zeit,
- die Anordnung im Raum,
- die Zuordnung der Personen,
- die Steuerung des Verhaltens durch Regeln,
- den Gebrauch von Medien,
- oder den Einsatz der Ressourcen.

Die Zeit wird in aller Regel starr verteilt, nicht flexibel, weil das Organisationsprobleme löst und Sinnfragen in Schach hält. Jede Schule hat einen genau kalkulierten Zeitplan, der auch für die Aufteilung im Raum und den Einsatz der Ressourcen massgeblich ist. Das Verhalten bezieht sich auf erwartete Leistungen und so auf schematisierte Bewertungen, die Niveauunterschiede darstellen müssen. Unterricht verlangt eine geordnete Abfolge und so eine Lösung des Steuerungsproblems. Zudem wird in aller Regel Anwesenheit verlangt und so eine soziale Organisation mit fest verteilten Rollen und Plätzen. Auch eine situative Begegnung ist immer eine Problemlösung, weil Abstand und Nähe geregelt sein müssen. Jede Veränderung verlangt eine neue Lösung, wobei nicht allein die Lehrkräfte für Lösungsentscheide sorgen. Aber braucht man dafür Forschung? Und wenn ja, in welchem Sinne?

---

<sup>1</sup> *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.

## 2. Der historische Schub der Bildungsforschung

Der heute renommierteste amerikanische Bildungsökonom Henry Levin hatte vor mehr als 35 Jahren eine offenbar gute Idee, die bis heute einleuchtet und die für die Bildungsforschung einen Quantensprung bedeutete. Auf dem Höhepunkt der Diskussion über Kybernetik als allgemeine Steuerungswissenschaft schlug Levin 1974 in der Zeitschrift *School Review* vor, auch Bildungssysteme durch Rückmeldungen und so durch regelmässige Datenerhebungen zu steuern. Der Aufsatz wurde zunächst gar nicht beachtet und entwickelte seine Dynamik erst, nachdem im April 1983 der Krisenreport *A Nation at Risk* veröffentlicht wurde. Mit ihm übernahm die Administration des damaligen Präsidenten Reagan die Rhetorik des Bildungsmonitoring und machte die Steuerung durch Datenerhebung zur Bundesphilosophie.

Levin unterscheidet „goals“ und „objectives“.

- *Goals* sind allgemeine Bildungsziele, die nach Prioritäten geordnet und auf mehreren Ebenen abgestimmt sind. Sie reagieren auf politische Prozesse und stellen ein System von gesellschaftlichen Anforderungen dar, auf die das Bildungssystem reagieren muss, ohne dass zwischen diesen Anforderungen und dem, was Schulen leisten, ein direkter Zusammenhang besteht.
- *Objectives* sind explizite und überprüfbare Teilziele, die im Laufe eines Prozesses erreicht oder verfehlt werden können. Diese Ziele beschreiben den gewünschten Erfolg in verschiedenen Lerndomänen, aber sie dürfen nicht wie eine feste, letzte Grösse betrachtet werden.

Nicht nur die Ziele müssen transparent kommuniziert werden, sondern auch die Erträge. Das fällt dem Erziehungssystem schwer, wusste schon Levin. Die Rhetorik der Ziele steht in keinem Verhältnis zur Diskussion der Resultate, die in offener Form gar nicht stattfindet (ebd., S. 380f.).

Soll sich das ändern, sind verschiedene Massnahmen nötig. Was Levin in diesem Zusammenhang vorschlug, kommt dem ziemlich nahe, was anschliessend die Systementwicklung tatsächlich bestimmt hat. Im Kern werden Schulen wie Unternehmen betrachtet, die Ziele erreichen und Bilanzen ziehen müssen. Zu diesem Zweck sollen besondere Agenturen oder Fachstellen (educational services) eingerichtet werden, die Anreize schaffen, wie die gewünschten Resultate erreicht werden können und die den Weg der Zielerreichung kontrollieren. Bislang - 1974 gesagt - ist das nicht annähernd der Fall.

„The result is that schools now are held accountable not for explicit educational outcomes, but for explicit educational processes and inputs“ (ebd., S. 383).

So entstand das, was heute etwas umständlich „Outcome-Orientierung“ genannt wird. Die Begründung ist einfach: Auch wenn Schulleitungen und Behörden (managers of school) nicht bewusst versuchen, Ziele zu erreichen und Leistungen zu steigern, so produzieren Schulen doch Resultate, und dies unablässig. „The implementation of specific processes and inputs *will* create outcomes“ (ebd.; Hervorhebung J.O.).

„Outcomes“ können unmittelbar solche der Schule sein oder mittelbar solche, die der Gesellschaft zugutekommen oder in ihr Kosten verursachen.<sup>2</sup> In jedem Fall, gewollt oder nicht gewollt, erzielt die Schule Effekte, die in den einen oder anderen Bereich fallen, wobei auch zwischen kurzfristigen und langfristigen Effekten unterschieden werden kann (ebd., S. 387). Wenn nun evaluiert werden soll, ob diese Effekte den Zielen des Unterrichts entsprechen oder nicht, müssen sie gemessen werden (ebd., S. 383). Und die Daten der Evaluation müssen an die Verantwortlichen zurückgemeldet werden, damit bessere Entscheidungssituationen zustande kommen.

„Before the various constituencies and managers of education can decide whether modifications are in order, they need good information on outcomes“ (ebd., S. 383f.).

Das können Daten aus Einzelevaluationen, aber auch aus Longitudinalstudien sein, die ausdrücklich erwähnt werden (ebd., S. 385). Gesteuert werden damit Ziele, Ressourcen sowie Belohnungen und Sanktionen (ebd., S. 386). Das System lernt von dem, was es an Resultaten produziert, also durch verschiedene Formen von Feedback.<sup>3</sup> Zu diesen Formen gehören vor allem Inspektionen, Evaluationen und Leistungsmessungen. Getestet wurde an amerikanischen Schulen schon vor der Einführung des Intelligenztests im Jahre 1920.<sup>4</sup>

Instrumente wie diese waren jahrzehntelang weder mit der Frage der Bildungsfinanzierung verbunden noch auf das Problem bezogen, wofür genau Schulen Verantwortung übernehmen sollen. Das Prinzip der „Accountability“ und der Aufbau von effektiven Rückmeldesystemen sind erst nach *A Nation at Risk* auf breiter Basis realisiert worden. Sie benötigen offenbar eine Krisensemantik wie in Deutschland nach dem PISA-Schock; von dieser Semantik hat niemand stärker profitiert als die empirische Bildungsforschung, die sich dann aber auch die Frage nach ihrem Nutzen gefallen lassen muss. Und diese Frage ist kein Mengenproblem: Die Steigerung des Forschungsoutputs muss nicht die Lösungskapazität verbessern.

Seit Levins Begründung ist das Thema „Bildungsmonitoring“ und Schulevaluation politisch stark aufgewertet worden, ohne an dem grundlegenden Modell von Levin viel zu verändern. Es beschreibt, was Qualitätssicherung im Kern ausmachen und wozu sie in der Praxis dienlich sein soll.

- Die Verantwortung der Schule kann nicht abstrakt definiert werden, sondern verlangt die Beobachtung des Prozesses und so eine Datenbasis.
- Anders kann jeder Akteur Verantwortung reklamieren, ohne zur Rechenschaft verpflichtet zu sein.

---

<sup>2</sup> „Educational outcomes can be thought of in two ways: proximate and ultimate. The proximate outcomes are those effects of schooling such as school induced increases in cognitive proficiencies and alterations of attitudes that are the immediate outputs of the educational process. The ultimate outcomes refer to the social effects of these changes in terms of their benefits or costs to society“ (Levin 1974, S. 383).

<sup>3</sup> Was auf diesem Weg angestrebt wird, ist ein *Bildungsmonitoring*, ohne das nationale Bildungssysteme kaum wirksam gesteuert werden können. Auch die dabei auftretenden Probleme wurden 1974 bereits klar benannt: „The fact that educational objectives of a specific nature rarely emerge from the political process and the lack of incentives for measuring outcomes lead to a situation where educational outputs *are not commonly monitored*. Moreover, there exist measurement problems in assessing outcomes, and there is a particularly severe evaluation problem in ascertaining the unique contribution of the educational sector to particular proficiencies and behaviors“ (Levin 1974, S. 387; Hervorhebungen J.O.).

<sup>4</sup> Der von Lewis Terman (1877-1956) weiter entwickelte Test wurde 1917 in die amerikanische Armee eingeführt und ist als „National Intelligence Test“ seit 1920 in den Schulen verwendet worden. Zehn Jahre später war das etablierte Praxis.

- Oder es kann jeder Verantwortung bestreiten, obwohl die Zuständigkeit gegeben ist.
- Die Rechenschaft über das Erreichte und der Nutzen des bisherigen Prozesses stehen als Erwartungen hinter dem Modell.

Soweit das kybernetische Modell und seine Begründungen, die bis heute einleuchten und denen abstrakt auch kaum widersprochen werden kann. Schulen können nicht nur durch den Input gesteuert werden, sondern verlangen auch die Orientierung am Resultat - das ist der Kern der Qualitätssicherung und sicher auch ein demokratietheoretisch gut begründetes Postulat. Die Frage ist nur, was „Resultate“ sind, wie sie zustande kommen und was davon in welcher Form die Öffentlichkeit erreicht.

- Im Modell von Levin fehlt eine Mengenvorstellung, es sagt nicht, wie *vielen* Datenerhebungen nötig sind und wie *oft* sie wiederholt werden müssen,
- damit Resultate zustande kommen, sichtbar werden und gute Entscheide auf den Weg gebracht werden.
- Das Modell sagt auch nicht, was geschehen soll, wenn zu viele Daten, widersprüchliche Aussagen und nicht beherrschbare Datenmengen vorliegen.

„Resultate“ kann es auf verschiedenen Ebenen geben und die müssen keineswegs harmonieren. Aber nicht nur die Menge, auch die Art der Datenerhebung ist nicht ohne Probleme.

Einmalige Erhebungen sind Momentaufnahmen und über den Moment hinaus kaum aussagekräftig, zumal nicht in beliebiger Kombination mit anderen Momentaufnahmen quer zu verschiedenen Bildungskulturen; Mehrfacherhebungen zum gleichen Thema führen gewollt oder ungewollt ein Konkurrenzprinzip ein, wie etwa an den vorliegenden Studien zur Freien Schulwahl gezeigt werden kann, wo verschiedene Forschungsequipen sich schon im Ansatz und dann auch in den Resultaten signifikant unterscheiden (Oelkers 2008); und Langzeitstudien eignen sich naturgemäss nicht für eine kurzfristige Themenbearbeitung, die aber politisch immer angesagt ist.

Weiter fällt auf, dass es kaum noch einen Bereich der öffentlichen Verwaltung gibt, der *nicht* ständig evaluiert wird. Das ist die Folge des Wirksamkeitsprinzips im „New Public Management“, und auch hier wird das Problem der ständig steigenden Datenmenge unterschätzt. Die Erforschung der „Wirksamkeit“ von allem und jedem hat so selbst ein Wirksamkeitsproblem, weil es keine Studie gibt, die über alle anderen entscheiden könnte. Man behilft sich dann mit Metastudien, die ihren Zweck, Übersicht zu schaffen, meistens durchaus erfüllen, oft auch zu praktikablen Vorschlägen kommen, aber angesichts der begrenzten Kapazität und der schnell wachsenden Fülle der Studien immer eine Selektion darstellen.

Schliesslich: Das Wort „Public“ im New Public Management bleibt weitgehend unbestimmt, und auch das schafft Probleme. Es geht ja nicht nur um eine neue Steuerungsphilosophie der öffentlichen Verwaltung, sondern auch um den Wandel der Öffentlichkeit. Die öffentliche Verwaltung führt staatliche Aufgabenbereiche wie den der Bildung unter der besonderen Beobachtung einer Öffentlichkeit, die primär medial verfasst ist und sich seit der Einführung des Internets radikal verändert hat. Internet heisst vor allem unmittelbare Zugänglichkeit ohne Ortsbindung und Zeitlimit.

Der englische Ökonom und erfahrene Evaluator Michael Power von der London School of Economics hat 1997 den Ausdruck „Audit Society“ geprägt, der auf die Ritualisierung der fortlaufend dichtereren Verifikation durch Datenerhebung und Rückmeldungen hinweisen soll. Wenn sich die Gesellschaft an die permanente Evaluation von allem und jedem gewöhnt hat, wird die Reaktion darauf wie die Evaluation selbst zum eingespielten Ritual (Power 1997). Die fünfte PISA-Studie hat nicht denselben politischen Impact wie die erste, auch und gerade dann nicht, wenn sie zu keinen besseren Resultaten führt und aber nicht absehbar ist, wie damit umgegangen werden soll. Dann verliert auch die „bildungsferne Schicht“ ihren Alarmierungscharakter; der Appell, etwas dagegen zu tun, wird zum folgenlosen Ritual.

- Audits, so Power, verlangen ein dreifaches „s“:
- *Samples, Specialists, and Systems* (ebd., S. 69ff.)
- und ein zweifaches „a“:
- *auditees and the autitable performance* (ebd., S. 91ff.).

Die Datenerhebung durch die Spezialisten ist die eine Seite, die öffentliche Kommunikation die andere. Auditoren übernehmen Aufträge aufgrund ihrer fachlichen und methodischen Expertise, die sich oft nicht zur Darstellung der Botschaften eignet. „Auditees“, also diejenigen, die von der Evaluation betroffen sind, müssen mit einer anderen als der Expertensprache angesprochen werden, das gilt auch für die grössere Öffentlichkeit. Für die Kommunikation ausserhalb der Experten ist die Wortwahl bei der Interpretation der Daten entscheidend. Keine Studie vermittelt sich von selbst und jeder Befund wird in der Öffentlichkeit politisch wahrgenommen. Die Experten mögen neutral sein, die Öffentlichkeit ist es nicht.

### 3. *Forschung, Politik und Öffentlichkeit*

Daten und Resultate aus dem Bildungsbereich finden vor allem dann das Interesse der Medien und der grösseren Öffentlichkeit, wenn sie sich mit starken Erfolgen oder Misserfolgen verbinden lassen, wobei das Ungewöhnliche und Hervorstechende der Meldung ausschlagend ist. Der klassische Vorwurf an die Medien: *Only bad news are good news*, ist verkürzt, aber auch „good news“ müssen für Aufmerksamkeit sorgen und so eine bestimmte Dimension erreichen. Das normale Geschäft der Schulen ist nicht sehr aufregend und hat kaum medialen Wert. Zur Abschätzung der Realität in den Schulen sind Normalitätsannahmen grundlegend, die sich mit dem herausragenden Einzelfall nicht erfassen lassen und aber genau damit bestritten werden können.

Was wir „Öffentlichkeit“ nennen, organisiert sich heute bekanntlich anders als im klassischen Print-Zeitalter. Damit werden auch die Meinungen anders gemacht, viel direkter und zugleich unbegrenzt. Ein Blog ist ein neues Medium der öffentlichen Kommunikation, bei dem die Person hinter der frei wählbaren Anschrift zurücktritt. Entsprechend drastisch können dann die Äusserungen sein; dass man „kein Blatt vor den Mund nimmt“, erhält so eine ganz neue Bedeutung. Bei Erziehungsthemen wird - nicht nur hier - ganz schnell geurteilt, und zwar immer vor dem Hintergrund einer pädagogischen Idealfolie, die die Erwartungen bestimmt. Das Versagen des Systems hat dann Methode.

- Man kann nämlich immer bezweifeln, dass es sich um einen „Einzelfall“ handelt;

- in der Wahrnehmung entsteht ein „Trend“ nicht so wie in der Statistik, sondern einfach durch die Koppelung mit anderen Fällen, die man nicht zählen muss, um zu einer Meinung zu gelangen.
- Und die Alarmierung der Öffentlichkeit ist im pädagogischen Bereich immer möglich, zumal dann, wenn auf jeden relativierenden Vergleich, der die Schlagzeile kosten würde, verzichtet wird.
- So erscheint als Fakt, was zu einem erheblichen Teil mediale Inszenierung ist, die auf ihren Erzeugungskontext mit keinem Wort eingeht.

Das macht die Public Relations zwischen Behörde und Publikum schwierig, auch weil es eher vorwurfsvolle als unterstützende Schlagzeilen gibt, die oft mit impliziten oder expliziten Schuldwürfen kommuniziert werden. Im Sinne von Walter Lippmann macht die Schlagzeile die Meinung, nicht umgekehrt, wie die Medien gerne selbst und im Blick auf die Grenzen ihrer Macht behaupten. Auch der Hinweis auf das Gallup-Paradox hilft nicht recht weiter. Zwar ist es seit George Gallup eine alte Erfahrung in der Demoskopie, dass auf die persönliche Situation *nicht* zutrifft, was im Allgemeinen für schlecht gehalten wird, aber genau deswegen entsteht das Problem, weil die Meinung nicht von der Erfahrung kontrolliert wird.

Die Funktionsweise medialer Kommunikation lässt sich an einem Fallbeispiel zeigen. Vor zwei Jahren hat die ETH-Zürich erstmalig ein Ranking veröffentlicht, in dem die Gymnasien danach gruppiert wurden, wie viele ihrer Absolventen ein Studium begonnen und mit welchem Erfolg absolviert haben. Das Ranking sorgte für helle Aufregung, weil damit ein Tabubruch verbunden war. Jeder weiss, dass sich die Gymnasien in ihrer Qualität unterscheiden, aber niemand hat sie je auf einer Folie vor der Presse in eine Rangfolge gebracht. Das Tabu betraf nicht die Unterschiede, sondern die öffentliche Kommunikation darüber. Das Tabu besteht, weil vor allem die Lehrkräfte befürchten, dass sich mit Rankings umso mehr Nachteile verbinden, je tiefer der Platz ihrer Schule auf der Rangskala ist. Was *genau* die Nachteile sind, darf allerdings bei einem Tabu, wenn es wirken soll, nicht kommuniziert werden.

Rankings von Schulen aufgrund von Leistungsdaten sind in England oder in den Vereinigten Staaten seit langem Alltag, mit dem keine Aufregung mehr verbunden ist. Sie sind Teil der *good educational governance*, die in Konkurrenz steht zu den Schul- und Lehrerbeurteilungen im Internet. Hier ist inzwischen ein eigener Wirtschaftszweig entstanden, der sich zur Forschung bei ihren Methoden bedient, natürlich mit einem erheblichen qualitativen Abstand, aber immerhin. Auch das ist eine Art Tabubruch, der die Exklusivität bedroht und zugleich als Demokratisierungsprozesse verkauft werden kann.

*Ratemyteachers.com* lautet die Webanschrift eines amerikanischen Anbieters, der Gelegenheit bietet, die Lehrkräfte zu bewerten und dann eine Rangfolge für jede Schule zu erstellen.<sup>5</sup> Es gibt eine *Hall of Fame* mit den besten und eine *Wall of Shame* mit den schlechtesten Schulen. Weniger weit, aber schon so gefürchtet, dass verschiedentlich Klagen eingereicht wurden, ist das deutsche *spickmich.de*<sup>6</sup> oder für die Hochschulen *meinprof.de*.<sup>7</sup> Hier werden einzelne Lehrkräfte beurteilt und miteinander verglichen, ohne dass irgendeine

<sup>5</sup> <http://www.ratemyteachers.com>.

<sup>6</sup> <http://www.spickmich.de> Die Seite wird als Hilfe für die Lehrer und als Orientierung für die Eltern angeboten.

<sup>7</sup> <http://www.meinprof.de> Die Bewertung setzt eine bestimmte Anzahl von Nennungen voraus. Die Auflistung geschieht alphabetisch und für jede Hochschule einzeln. Die Professoren erhalten von ihren Studierenden Noten, aus denen ein Durchschnitt gebildet wird.



Behörde eingreifen könnte. Eltern können auf der Seite *schulradar.de*<sup>8</sup> Wertungen abgeben und so die Entwicklung in den Schulen ihrer Kinder kommentieren.<sup>9</sup>

Die Frage ist, wie Kunden auf diese Art Governance reagieren. Auf jeden Fall sind Rankings im angelsächsischen Sprachraum inzwischen Gewohnheitssache und so entdramatisiert. Selbst das in konservativen Kreisen immer noch als die Skandalschule schlechthin wahrgenommene Internat von Summerhill besteht inzwischen das regelmässige Audit der Behörde zur Überprüfung der Standarderreichung, wobei man sich allerdings fragt, was wohl Alexander Neill dazu gesagt hätte, dass seine Schule nicht nur von aussen beurteilt wird, sondern auch noch Noten erhält. Das Verfahren selbst ist inzwischen soweit standardisiert, dass nur die wirklich schlechten Schulen in kurzen Intervallen und eng evaluiert werden. Alles Andere wäre unbezahlbar, nachdem ohnehin eine Bürokratie aufgebaut werden musste, die es vorher nicht gab.

In der Schweiz werden die öffentlichen Schulen erst seit wenigen Jahren evaluiert, kantonale durchaus unterschiedlich und mit immer noch spürbarer Abwehr an der Basis. Wenn die Resultate schlecht sind, wird die Methode angezweifelt, was umgekehrt nicht gilt; und wem die ganze Richtung nicht passt, kann im Kanton Zürich unmittelbar an die untergegangene Bezirksschulpflege erinnern. Das sind vermutlich nur Anpassungsprobleme, was für das Thema „Ranking“ nicht in gleicher Weise gesagt werden kann, weil damit ein sehr grundsätzlicher Konflikt verbunden ist, den die Aufregung über das ETH-Ranking auch deutlich spiegelt.

- Leistungsmessungen implizieren Rankings,
- die Frage ist nur, wie sie dargestellt werden,
- wen sie erreichen sollen
- und ob sie öffentlich zugänglich sind.

Die Aufregung über den Vorstoss der ETH-Leitung mag zum Teil damit erklärt werden, dass keine Zürcher Schule auf dem ersten Platz war und ein Gymnasium aus Winterthur vor den Schulen der Stadt Zürich rangierte. In der Aufregung ist völlig übersehen worden, dass die Abstände zwischen den Schulen gering waren und nur ein Parameter - Studium an der ETH - überhaupt erhoben wurde. Weil das eine Eliteanstalt ist, die manche Eltern für ihre Kinder fest eingeplant haben, erhielten Schulleitungen besorgte Anrufe und füllten sich die Leserbriefspalten und Blogs zum Thema.

Die ästhetische Form der Rankings ist die einer gestuften Rangfolge mit festen Plätzen, die nie genau das wiedergeben kann, was empirisch erfasst wird. Rankings sind Vereinfachungen zur schnellen Lesart, die unmittelbar mit Assoziationen von „oben“ und „unten“ verbunden werden. In einer Leistungsgesellschaft will niemand „unten“ sein, während ein Ranking nur dann überzeugend ist, wenn es genau das leistet und eine Hierarchie abbildet, die zwischen exzellent, gut und schlecht unterscheiden kann. Rankings sind umso besser, je mehr sie die Datenlage auch tatsächlich abbilden. Das ist nie ganz der Fall, wohl aber näherungsweise. Insofern basierte der Tabubruch auf keinem so guten Beispiel.

<sup>8</sup> <http://www.schulradar.de> Es gibt Top-Listen mit den Schulen, die von den Eltern am besten bewertet wurden.

<sup>9</sup> Bewertet werden insgesamt zehn Kategorien: Qualität der Lehrer, technische Ausstattung, Fächer/AG-Angebot, Sportmöglichkeiten, Essensmöglichkeiten, Schulgebäude, Stimmung unter den Mitschülern, Schulleitung, Unterrichtsausfälle, Mitbestimmungsmöglichkeiten. Die Anzahl der Bewertungen ist nur für registrierte Mitglieder zugänglich.

Solange keine wirklich politische Entscheidung gefallen ist, bleibt das Reizthema erhalten, und damit zugleich alle Erzählungen von Fortschritt und Untergang, die sich um das Thema ranken, Klarheit kann nur die Praxis schaffen, was in diesem Falle gleichbedeutend wäre mit der Entwicklung und Etablierung eines geeigneten Instruments. Problematisch ist dann höchstens die Akzeptanz, die dann zunimmt, wenn das Instrument als valide und fair gilt. Viel schwieriger liegen Fälle, in denen sich aus den vorliegenden Daten wohl Schlüsse ziehen lassen, bei denen aber die Umsetzung nicht einfach instrumentell möglich ist, sondern die Akteure verändern muss. Die Bedingung dafür heisst in der amerikanischen Bildungsforschung „sense-making“ (Oelkers/Russer 2008).

#### 4. *Unangenehme Resultate*

Eine zentrale Frage ist, wie die Bildungspolitik mit für sie unangenehmen Daten umgeht, ein Fall, der bei Levin gar nicht vorkommt. Auch hier kann ich mich auf ein Fallbeispiel beziehen, nämlich die 2004 in Auftrag gegebene Langzeitstudie über die Entwicklung des Lernstandes im Kanton Zürich. Die Schülerinnen und Schüler sind bei Schuleintritt untersucht worden und nachfolgend am Ende der dritten Klasse, der sechsten und der neunten Klasse. Die Ergebnisse der ersten beiden Studien (Moser/Stamm/Hollenweger 2005; Moser/Keller/Tresch 2008; Moser/Hollenweger 2008) liegen vor und sind sogar vergleichsweise breit kommuniziert worden, auch weil die Studie tatsächlich sehr aufschlussreich ist.

Einige Ergebnisse lauten so:

- Bei Schuleintritt haben etwa vier Fünftel der Kinder den Lernstoff der ersten Klasse in Mathematik teilweise bewältigt und ein knappes Fünftel befindet sich leistungsmässig bereits in der zweiten Klasse.
- Die Lesekompetenzen liegen weiter auseinander. 5 Prozent der Kinder kennen keine Buchstaben, etwa 35 Prozent erfassen die Verbindung zwischen Buchstaben und Lauten nur im Ansatz, 15 Prozent lesen bereits schwierige Wörter und längere Sätze.
- Noch disparater ist der Wortschatz, also die Fähigkeit, Bilder mit eigenen Wörtern richtig zu bezeichnen.

Nur knapp ein Drittel der Kinder beherrscht die richtigen Bezeichnungen, ein weiteres Drittel überbrückt geschickt die Lücken, das letzte Drittel beherrscht nur ganz elementare Bezeichnungen, also kann sagen, dass auf einem Bild ein Hammer ein „Hammer“ ist, mehr jedoch nicht. Korrekte Bezeichnungen für komplexere Bilder konnten von einem Drittel der Kinder nicht genannt werden.

Die Schülerinnen und Schüler erreichen die erste Klasse der Volksschule mit grossen Unterschieden im Lernstand. Unterschiede können ausgeglichen werden, wie die zweite Studie nach drei Jahren Schulzeit gezeigt hat. Der Unterricht hat starke, aber auch ungleiche Effekte, je nachdem, wie er stattfindet, und das ist für Steuerungsphilosophen ein ganz gemeines Datum, denn wie will man die Akteure steuern? Und vor allem, wie will man die Unterschiede ausgleichen, die die Lehrkräfte ja in guter Absicht schaffen? Wenn es aber stimmt, dass der Unterricht den Unterschied macht und nicht einfach nur die soziale Herkunft,

dann will die Öffentlichkeit wissen, wie sich die Qualität des Unterrichts beeinflussen lässt, damit die Ziele der Volksschule erreicht werden.

Man sieht, Rankings sind trotz aller Aufregung eigentlich das leichtere Problem, denn was muss man bei der Unterrichtsqualität alles „beeinflussen“?

- Die Kompetenz amtierender Lehrkräfte,
- die Zusammensetzung der Klasse,
- die eingesetzten Lehrmittel
- den je erreichten Lernfortschritt
- die Risikogruppen, also diejenigen Schülerinnen und Schüler, die die Lernziele nicht erreicht haben
- und mit hoher Wahrscheinlichkeit im weiteren Verlauf ihrer Schulkarriere weiter zurückbleiben.

Das sind die wirklichen Probleme, die von der Öffentlichkeit oft übersehen werden, die Resultate sehen will und nach stichhaltigen Erklärungen verlangt. Aber Probleme wie diese sind schwer zu bearbeiten, weil sie *in* den Klassen entstehen, trotz erheblichem Einsatz der Lehrkräfte auftreten und vielfach gar nicht bemerkt werden. Die vorliegenden Studien verweisen darauf, dass die Entwicklung des Lernstandes zwischen verschiedenen Klassen auch in ein- und derselben Schule höchst unterschiedlich sein kann. Das gilt ebenso für die Effekte der Förderung. In den einen Klassen gelingt es, die Leistungsschere zu verringern, in anderen jedoch nicht, was vor allem damit zu tun hat, dass sehr verschieden unterrichtet wird.

Damit will ich nicht auf die Unmöglichkeit der Bearbeitung dieses Problems hinweisen, sondern nur auf den Komplexitätsgrad, die lange Dauer und den durchaus ungewissen Ausgang. Eine gute politische Botschaft ist das nicht. Ungewissheit belastet die Deutungshoheit, Politik ist kein Universitätsseminar und *wie* mit der Öffentlichkeit kommuniziert wird, ist von ausschlaggebender Bedeutung, weil jedes ungelöste Problem nachteilig ausgelegt wird. Unlösbare Probleme mag es geben; was vermieden werden muss, ist der Anschein der Untätigkeit.

Ein Zusatz sei mir noch erlaubt: Auch der Bezug auf andere Studien, zumal ausländische, sollte genau geregelt sein. Ergebnisse aus finnischen Gesamtschulen mögen politisch attraktiv sein, aber lassen sich kaum auf Schweizer Verhältnisse übertragen. Zum Glück liegen inzwischen für die Schweiz und besonders für den Kanton Zürich genügend eigene Studien vor, auf die man sich beziehen kann. In der Physik oder in der Experimentalpsychologie ist es egal, wo man forscht, wenn nur die Ausstattung stimmt; in der Bildungsforschung muss die historisch-kulturelle Besonderheit des Gegenstandes in Rechnung gestellt werden. Aussagen über deutsche Gymnasien lassen sich nur schwer mit Schweizer Realitäten in Einklang bringen, obwohl sich beide Seiten auf Humboldt berufen.

Abgesehen davon, allgemein gilt: Wer die Absicht hat, mit Reformen auch etwas zu verbessern, muss die Stelle im System bestimmen, wo das geschehen soll. Appelle aus Reformprojekten 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 und - ob es Sinn macht. Projekte nutzen sich ab und können das Gegenteil von dem bewirken, was ursprünglich intendiert war, wenn nicht genau darauf geachtet wird, wo ein Projekt zum Einsatz kommt und wie es sich entwickelt.

Wie leicht oder wie schwer das ist angesichts knapper Ressourcen, lässt sich anhand verschiedener Beispiele aus meinem Arbeitsbereich zeigen. Diese Beispiele haben einen

unterschiedlichen Bezug zur Forschung, sie zeigen aber, dass zwischen dem Bildungsmonitoring für die Politik und Reformen, an denen die Akteure beteiligt sind, unterschieden werden muss. Der Nutzen der Forschung muss sich auch im Feld finden lassen, was dann der Fall ist, wenn konkrete Projekte begleitet werden.

### 5. Reformen, Effekte und Forschungsbasierung

Das erste Beispiel bezieht sich auf das Ende der Schulzeit und den Übergang zur Berufsausbildung. Das Beispiel stammt aus St. Gallen und ist sogar vom Kanton Zürich vorbehaltlos übernommen worden. Man kann an dem Beispiel auch sehen, wie „HarmoS“ tatsächlich wirkt, als unterschwellige und passende Harmonisierung, nicht als Versuch mit der Brechstange. Das Beispiel ist nicht Teil des Konkordats, aber sicher eine Folge der gestiegenen Bereitschaft, nicht in jedem Kanton das Rad selbst und immer wieder neu zu erfinden, sondern auf gute Lösungen anderer Kantone zurückzugreifen.

Die Rede ist von *Stellwerk*, also dem Leistungstest Mitte der achten Klasse, der heute in vielen Kantonen durchgeführt wird. Er ist entwickelt worden, um der schleichenden Entwertung des Volksschulabschlusses vorzubeugen und die Bewertungen der Lehrkräfte sinnvoll zu ergänzen.<sup>10</sup> In einem Stellwerk der SBB werden die Weichen gestellt. Im Sinne dieser Metapher werden alle Schülerinnen und Schüler Mitte der 8. Klasse getestet, über welche Kompetenzen in zentralen Fächern sie tatsächlich verfügen. Das Ziel ist, die Anforderungen zu erfüllen, die heute an qualifizierte Berufslehren gestellt werden. Die Volksschule ist dafür das „Stellwerk“. Sie ist verantwortlich, dass Lücken geschlossen werden und am Ende grundlegende Kompetenzen vorhanden sind.

- Der Test ist die Grundlage für gezielte Nachbesserungen während der verbleibenden Schulzeit.
- Die Schüler erhalten in einem Standortgespräch mit Lehrern und Eltern eine objektivierte Rückmeldung, wo sie stehen und sie können fehlende Kompetenzen aufholen.
- Fördern in diesem Sinne setzt einen Treffpunkt voraus, an dem die genaue Richtung und der Ressourceneinsatz bestimmt werden.

Ohne solche direkten Rückmeldungen ist die Wahrscheinlichkeit gross, dass sich am Lernen nur wenig verändert. Die Schülerinnen und Schüler können bekanntlich einiges Geschick in Einsichtsverweigerung entwickeln, da sind objektivierte Daten durchaus willkommen.

In einer akteursnahen Evaluation über die Nutzung von „Stellwerk“ im Kanton Zürich zeigte sich, dass neben dem Test vor allem die Standortgespräche bei den Lehrkräften auf grosse Zustimmung stossen, auch wenn damit zusätzliche Belastungen verbunden waren (Kammermann/Siegrist/Lempert 2007). Die Belastungen werden mit der Wirksamkeit verrechnet, die Lehrkräfte nehmen Mehrarbeit in Kauf, wenn sich ein neues Instrument als nützlich erweist und mit Erfolg eingesetzt werden kann. Das gilt auch umgekehrt: Belastungen mit manifestem Sinnlosigkeitsverdacht werden unterlaufen oder eben ausgebremst.

---

<sup>10</sup> <http://www.stellwerk.ch>

Auf der Linie dieser Entwicklung wird im Kanton Zürich auch das neunte Schuljahr verändert.

- Die Lektionentafel unterscheidet neu zwischen einem minimalen und einem maximalen Angebot, das vor Ort in den Schulen festgelegt wird.
- Das Angebot der Fächer wird reduziert, die Schüler verfolgen aufgrund ihrer Stärken und Schwächen auch individuelle Ziele.
- Verbindlich sind zudem drei Lektionen Projektunterricht pro Woche sowie eine grössere, selbständig erstellte Abschlussarbeit, ähnlich wie das in den Gymnasien der Fall ist.
- Die Schülerinnen und Schüler lernen auch, wie man die im Projekt erstellten Produkte dokumentiert und präsentiert.

Damit öffnet sich die Volksschule ein Stück weit in Richtung Lehrstellenmarkt und das ist eine Reformoption, die trotz Mehrbelastung wiederum auf Zustimmung stösst. Die vorliegenden Forschungen waren in dieser Hinsicht aufschlussreich und halfen bei der politischen Entscheidung.

Ein weiteres Beispiel für lohnende Reformen ist die gezielte Entwicklung der Lehrmittel. Unterrichtsqualität hat - abgesehen von den Randbedingungen - zwei zentrale Parameter, die systematisch beeinflusst werden können, nämlich die Lehrmittel und die Kompetenz der Lehrkräfte. Lehrmittel aber entstehen überwiegend immer noch auf althergebrachte Weise, darüber darf die Entwicklung im E-Learning-Sektor nicht hinwegtäuschen.

- Lehrmittel sind Autoren-Produkte, die vor ihrer Implementation wohl beurteilt, aber keinen nennenswerten empirischen Kontrollen unterliegen; sie werden eingeführt, ohne Testserien im Feld vorauszusetzen.
- Auch ihr Gebrauch wird nicht erhoben, so dass wir nicht wissen, ob die Entscheide der Autoren, ihr Produkt zu verändern, sinnvoll gewesen sind oder nicht, wobei es dafür eigentlich nur *ein* Kriterium gibt, nämlich den Lernerfolg der Schülerinnen und Schüler.

Hier liegt erheblicher Forschungsbedarf vor. Die Autoren von Lernmitteln rufen üblicherweise keine Daten ab, mit denen sie ihr Produkt verbessern könnten oder aufgrund derer das Produkt vom Markt genommen werden müsste. Für die Datenerhebung müssten im Internet elektronische Fragebögen zugänglich sein, die fortlaufend benutzt werden können. Damit könnte die Weiterentwicklung der Lehrmittel von den Rückmeldungen der Benutzer abhängig machen werden. Weil das bislang kaum geschieht, sind viele neue Lehrmittel oft Neuauflagen von alten, auch wenn sie anders aussehen. Was aber den Unterricht steuert, sind Lehrmittel und nicht Lehrpläne, wie umfangreich und wohlmeinend diese auch immer formuliert sein mögen.

Für die Kunst des Unterrichtens gibt es bislang kaum verbindliche Standards, sondern oft nur das Ergebnis von Versuch und Irrtum, wobei es für die Praxis spricht, dass gleichwohl in vielen Fällen eine hohe Qualität entwickelt wird. Das berufliche Können der Lehrkräfte ist jedoch fragil, die tatsächliche Leistungsfähigkeit sollte daher fortlaufend untersucht und verbessert werden. Sie ist für den Erfolg der Schule zu wichtig, um stillschweigend zermürbt zu werden. Dabei hilft kein Zureden, sondern nur professionelle Entwicklung und - gute Lehrmittel. Sie sind ein Kern der Qualitätssicherung, der nicht durch fliegende Kopien im

Klassenzimmer ersetzt werden kann. Die „Standards“ der Bildung sind in hohem Masse Standards der Lehrmittel, die besser als bisher in ihrer Wirksamkeit überprüft werden müssen.

Anpassungen zwischen den Lehrplänen und den zugelassenen Lehrmittel sind aufwändig und nicht billig, aber möglich. Ehrgeizige Projekte gehen dahin, kompetenzbasierte Lehrpläne und darauf bezogene Lehrmittel zu entwickeln. Ein Beispiel ist das Lehrmittel Mathematik für die Sekundarstufe I des Kantons Zürich, das der kantonale Lehrmittelverlag in Auftrag gegeben hat und mit dem die Unterrichtspraxis auf eine neue Grundlage gestellt werden soll. Die Entwicklung dieses Lehrmittels wird erstmalig im Feld erprobt, sie umfasst mehrere Kantone, die Entwicklungszeit beträgt bis zu 4 Jahren und der Kostenaufwand ist erheblich.

Der Versuch ist als so vielversprechend wahrgenommen worden, dass sich weit mehr Gemeinden beworben haben, als aufgenommen werden konnten. Mathematik ist ein zentrales Leistungsfach, das von vielen Schülerinnen und Schülern als problematisch erlebt wird; Verbesserungen in der Breite sind nur möglich, wenn ein Durchbruch bei den Lehrmitteln erzielt wird. Was ein „Durchbruch“ ist, entscheidet sich im Versuch und nicht am Schreibtisch. Heute werden Lehrmittel ausgezeichnet und mit Preisen bedacht, weil sie Expertenkriterien genügen. Aber worauf es ankommt, ist die Verwendbarkeit im Unterricht und die Gretchenfrage lautet, ob es gute Lehrmittel für schwache Schülerinnen und Schüler gibt. Hier lohnt fast jede Reform.

Aber das ist bislang keineswegs der Regelfall. Denn was ist mit Reformen, die wohl neue Ideen entwickelt, aber die Umsetzung offen gelassen und die Effekte zunächst gar nicht durch Forschung kontrolliert haben? Ich gebe dafür ein weiteres Beispiel: Seit Mitte der achtziger Jahre des letzten Jahrhunderts ist mit hohem Aufwand versucht worden, das historisch gewachsene System der Schweizer Gymnasien grundlegend zu erneuern und anzupassen. Ich gehe davon aus, dass sich die Erfahrungen mit diesem Versuch auch auf andere Schulreformen übertragen lassen.

Ziele der MAR-Revision waren

- die Loslösung der Ausbildung von den fünf bestehenden Maturitätstypen,
- die stärkere Individualisierung des Lernens durch Eröffnung von Wahlmöglichkeiten,
- interdisziplinärer Unterricht,
- Vermittlung überfachlicher Kompetenzen
- und die Vermeidung von zu früher Spezialisierung im Blick auf bestimmte Studienrichtungen.

Das einzige Mittel, den Prozess in Richtung dieser Ziele zu steuern, war seinerzeit der neue Rahmenlehrplan, der erste gesamtschweizerische Lehrplan für die Gymnasien, der je vorgelegen hat. Er wurde in aufwändiger Arbeit von den Fachgruppen der Gymnasiallehrkräfte entwickelt und konnte im Juni 1994 verabschiedet werden. Der Plan sah vor für alle Fächer

- Allgemeine Bildungsziele,
- Begründungen und Erläuterungen des Faches
- sowie Richtziele, mit denen die fachlichen Kompetenzen erfasst werden sollten.

- Zusätzlich und quer zu den Fächern wurde auch die Ausbildung überfachlicher Kompetenzen abverlangt.

Acht Jahre nach Beginn der Reform der Gymnasien im Jahre 1995 ist das Ergebnis flächendeckend und mit einem hohen Forschungsaufwand evaluiert worden (Evaluation 2004; Oelkers 2008). Die grundlegenden Ergebnisse dieser Untersuchung lauten so: Die Maturitätstypen sind unter neuen Bezeichnungen durchaus lebendig; damit hält die Spezialisierung für bestimmte Studiengänge auf der Sekundarstufe II an; weder ein nennenswerter Wandel hin zu interdisziplinärem Unterricht noch die Ausbildung überfachlicher Kompetenzen konnte festgestellt werden. Weiterhin dominiert gymnasialer Fachunterricht die Erfahrung der Schülerinnen und Schüler.

Das Beispiel ist in mehrfacher Hinsicht lehrreich:

- Mit einem Rahmenlehrplan verändert man nicht die Schulpraxis.
- Zusätzliche Aufgaben verlangen zusätzliche Ressourcen, ohne neue Gefässe gibt es keine Erweiterung des Unterrichts;
- wenn überfachliche Kompetenzen gestärkt werden sollen, müssen sie den Unterricht erreichen,
- und das gelingt nur mit einer Repertoireanpassung der Lehrkräfte, die ohne Ressourceneinsatz nicht zu haben ist.
- Das kann man auch als *circulus vitiosus* der Reform bezeichnen; wer ihn *nicht* durchbricht, scheitert.

Ein letztes Beispiel zum Verhältnis von Reform, Zeitdruck und Effekt: Im Kanton Zürich werden inzwischen alle Schulen alle vier Jahre extern evaluiert. Die Evaluationen werden von der unabhängigen „Fachstelle für Schulbeurteilung“ durchgeführt, die Expertinnen und Experten einsetzt. Im vergangenen Jahr bezogen sich die Evaluationen auf rund ein Viertel der Zürcher Schulen, ausgenommen die Gymnasien und Berufsschulen. Das Verfahren basiert auf neun Qualitätsanforderungen (QA), die sich auf drei zentrale Bereiche der Schulentwicklung beziehen, nämlich Lebenswelt Schule, Lehren und Lernen sowie Führung und Management. Für jeden dieser Bereiche stehen drei bis vier Indikatoren zur Verfügung stehen, die die Beurteilung leiten (ausführlich dargestellt in: Handbuch 2005).

Der Bericht über das Schuljahr 2007/2008 zeigt das Kernproblem von Schulreformen in aller Deutlichkeit (Fachstelle 2009). Die Ergebnisse sind dort gut bis sehr gut, wo das normale Geschäft der Schulen betroffen ist, also

- die Pflege der Schulgemeinschaft,
- die Entwicklung und Sicherstellung von Verhaltensregeln,
- die Strukturierung des Unterrichts
- oder die Herstellung eines lern- und leistungsfördernden Klassenklimas.

Aber auch in diesen Bereichen sind die Unterschiede zwischen den einzelnen Lehrkräften und den Klassen gross. Das gilt nicht zuletzt für die Individualisierung und Differenzierung im Unterricht. Das geschieht selbst individuell und sehr differenziert, also mit uneinheitlichen Regeln und Massstäben, die persönlich interpretiert werden müssen.

Qualitätsansprüche, die sich auf Entwicklung, Management und Führung der Schulen beziehen, werden von den Evaluatoren wesentlich schwächer beurteilt. Gesetzlich müssen alle Schulen über eine Leitung verfügen. Deren Leistung wird wie folgt beurteilt:

Ein weiteres Ergebnis betrifft die Praxis der internen Evaluation von Schule und Unterricht, ein bildungspolitisches Thema seit mindestens zehn Jahren und eigentlich auch gesetzlich vorgeschrieben. Hier liegen folgende Befunde vor:

- Eine systematische Evaluation von Jahreszielen oder Projekten ist in der Schule noch nicht üblich.
- Die Schule und die Lehrpersonen überprüfen den Unterricht im Blick auf seine Qualität noch selten.
- Die Eltern werden in vielen Schulen kaum nach ihrer Meinung gefragt.

Die Unterschiede im Entwicklungsstand erklären sich mit den Verhältnissen vor Ort. Mit Annahme des neuen Volksschulgesetzes von 2005 mussten alle Schulen Leitungen einführen. Vorher fand ein mehrjähriger Schulversuch statt, an dem sich etwa ein Viertel der Schulen des Kantons beteiligten. Die Schulen, die sich an dem Versuch nicht beteiligt hatten, waren auf die Konsequenzen des Gesetzes nicht vorbereitet. Sie reagierten teilweise wie auf einen Kulturschock, weil plötzlich Leitung vorhanden war und die Daten offengelegt werden mussten. Auch Zielvereinbarungen waren für viele der unvorbereiteten Schulen Neuland, das manche lieber nicht betreten hätten. Und das Verfahren der externen Evaluationen stiess zum Teil auf heftigen Widerstand.

Das Beispiel zeigt die Probleme, aber lässt sich nicht als Scheitern verstehen, was erst dann der Fall wäre, wenn die *Umsetzung* misslingt, also die Berichte schubladisiert werden, die Leistungsvereinbarungen ohne Folgen bleiben und sich in vier Jahren in den einzelnen Schulen nichts geändert hat. Bis das definitiv beurteilt werden kann, vergehen mindestens noch sieben Jahre, also überschreitet der Zeitrahmen meine Pensionsgrenze. Allein das zeigt, wie viel Zeit erforderlich ist, einmal begonnene Reformen halbwegs bilanzieren zu können.

- Der Versuch mit externen Evaluationen ist in Zürich 1999 begonnen worden,
- der Meilenstein war das Volksschulgesetz von 2005
- und eine sichere Beurteilung der Wirksamkeit dürfte kaum vor 2015 möglich sein.

Aber so schliesse ich nicht. Schulen sind keine Aktiengesellschaften, die vierteljährlich Bilanz ziehen müssen. Die grösste Gefahr ist nicht die Gewinnwarnung der Rating-Agenturen, sondern die Abnutzung von Motivation und Engagement, also das Verhalten der Akteure. Sie müssen sich in einem Rahmen bewegen, der ihre Arbeit herausfordert und unterstützt, und die Frage ist dann, wie die Organisation der Schule sinnvoll und schonend verbessert werden kann. Wenn beide Kriterien *nicht* erfüllt werden, bekommt jedes Reformprojekt ein Problem.



- Dewey, J.: *Wie wir denken*. Mit einem Nachwort neu hrsg v. R. Horlacher/J. Oelkers. Zürich: Verlag Pestalozzianum 2002. (amerik. Orig. 1910)
- Evaluation der Maturitätsreform 1995 (EVAMAR). Im Auftrag der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) und des Bundesamtes für Bildung und Wissenschaft (BBW). *Neue Fächerstruktur - Pädagogische Ziele - Schulentwicklung*. Schlussbericht der Phase 1. Bern: EDK/BBW 2004.
- Fachstelle für Schulbeurteilung: *Jahresbericht 2007/2008*. Zürich: Bildungsdirektion des Kantons Zürich 2009.
- Handbuch Schulqualität. *Qualitätsansprüche an die Volksschule des Kantons Zürich*. Zürich: Kantonale Drucksachen und Materialzentrale kdmz 2005.
- Kammermann, M./Siegrist, M./Sempert, W.: *Begleitende und abschliessende Auswertung der Erfahrungen mit dem neu gestalteten Schuljahr an der Sekundarschule des Kantons Zürich*. Schlussbericht zur zweiten Erhebung (April-Juni 2007). Vervielf. Ms. Zürich: Interkantonale Hochschule für Heilpädagogik 2007.
- Levin, H.M. (1974). *A Conceptual Framework for Accountability in Education*. In: *School Review*, Vol. 82, no. 3 (May), S. 363-391.
- Moser, U./Stamm, M./Hollenweger, J. (Hrsg.): *Für die Schule bereit? Lesen, Wortschatz, Mathematik und soziale Kompetenzen bei Schuleintritt*. Oberentfelden: Sauerländer 2005.
- Moser, U./Keller, F./Tresch, S.: *Schullaufbahn und Leistung. Bildungserfolg und Lernverlauf von Zürcher Schülerinnen und Schülern am Ende der dritten Volksschulklasse*. Bern: h.e.p. Verlag 2008.
- Moser, U./Hollenweger, J. (Hrsg.): *Drei Jahre danach. Lesen, Wortschatz und soziale Kompetenzen am Ende der dritten Klasse*. Oberentfelden: Sauerländer 2008.
- Oelkers, J.: *Die Schule nach Wahl? Privatschulen, Regelschulen und Bildungsgutscheine im internationalen Vergleich*. Ms. Zürich 2008.
- Oelkers, J.: *Die Qualität der Gymnasien. Eine Expertise für den Kanton Zürich*. Bern: h.e.p. Verlag 2008a.
- Oelkers, J./Reusser, K.: *Expertise: Qualität entwickeln - Standards sichern - mit Differenz umgehen*. Unter Mitarbeit von E. Berner/U. Halbheer/St. Stolz. Berlin: Bundesministerium für Bildung und Forschung 2008.
- Power, M.: *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press 1997.
- Schmid, Chr.: *Lernen und Transfer. Kritik der didaktischen Steuerung*. Bern: h.e.p. Verlag 2006.