

Jürgen Oelkers

Good Governance im Umgang mit empirischen Bildungsdaten ^{*)}

1. Der Auftrag: Fünf Fragen

Mein Vortrag soll gemäss Auftrag und Mail vom 28. Juli 2009 folgende fünf Fragen beantworten:

1. Was bedeutet das IDG, also der Grundsatz der Öffentlichkeit aller (nicht personalisierten) Bildungsdaten, für eine gelingende Qualitätspolitik?
2. Was bedeutet dieses Öffentlichkeitsprinzip auch für die Durchführung von Datenerhebungen?
3. Welche Leitideen und Strategien einer „Good Educational Governance“ im Umgang mit Bildungsdaten schlagen Sie für den Kanton Zürich vor?
4. Wo sehen Sie Regelungsbedarf?
5. Was würde dies für den Umgang mit den Bildungsdaten der eingegebenen Fallbeispiele bedeuten?

Hätte ich das vorher gewusst, könnte ich sagen, wäre ich jetzt nicht hier, denn natürlich kann kein einzelner Vortrag diese Fragen schon wegen der zahlreichen Fallbeispiele irgendwie erschöpfend beantworten. In dieser Situation gibt es eine bewährte Strategie zur Eingrenzung, man historisiert das Thema und reduziert die Komplexität mit Blick auf seine Entwicklung. Das nimmt dem Thema seine Einzigartigkeit, wie sich an den Reaktionen der Lehrerschaft auf Qualitätspolitik zeigen lässt. Damit werde ich beginnen.

2. Ein Blick zurück

Der heute renommierteste amerikanische Bildungsökonom Henry Levin hatte vor 35 Jahren eine offenbar gute Idee, die er auch gleich öffentlich machte. 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.

^{*)} Vortrag auf der Retraite der Bildungsdirektion des Kantons Zürich am 20. August 2009.

Terrel Bell,¹ der Secretary of Education im ersten Kabinett von Ronald Reagan, hatte seinen Präsidenten davon überzeugt, eine National Commission on Excellence in Education einzusetzen. Die Kommission konstituierte sich am 26. August 1981 unter der Leitung des Präsidenten der Universität von Utah, David P. Gardner.² Curricular und schulorganisatorisch strebte die Kommission eine Richtung „back to the basics“ an, also weg von der schülerzentrierten Pädagogik, dem offenen Unterricht und der weitgehend freigesetzten Fächerwahl. *A Nation at Risk* war der Auslöser vor allem für curriculare Reformen, die in eine Diskussion über nationale Bildungsstandards, Konzentration der Fächer in der High School und die Frage der Wirksamkeit einmündeten. Erst jetzt kam die Stunde der Bildungsökonomien.

Wäre Harry Levin Historiker, hätte er wissen können, dass seine Idee vielleicht gut, aber nicht neu war. Bereits 1912 hatte der Bildungsstatistiker Leon Porter Ayres³ angeregt, mit Daten über die Wirksamkeit der Schulen Transparenz zu erzeugen und eine neue Steuerungstechnik ins Leben zu rufen. Von Ayres stammt die Formel

„measuring educational processes through results“
(Ayres 1912).

Die erste externe Evaluation im grösseren Umfang ausserhalb der Schulinspektion hat allerdings weder ein Bildungsökonom noch ein Statistiker, sondern ein Kinderarzt durchgeführt. Zufall ist das nicht. Wir brauchen nur auf den Kanton Zürich zu schauen, um zu erkennen, welche Meinungsmacht in der Erziehung Pädiater entfalten können, zumal dann, wenn sie Bücher schreiben.

Der Kinderarzt hiess nicht Lago, sondern Joseph Mayer Rice, der einige Jahre in New York praktiziert hatte,⁴ 1888 seine Praxis aufgab und nach Europa ging. Er studierte zwei Jahre lang Psychologie und Pädagogik an den Universitäten Jena und Leipzig. Der Grund war die Fundierung einer Erziehung, die auf dem Prinzip der Induktion beruhen und das Lernen der Kinder in den Mittelpunkt stellen sollte. Kinder, so die Theorie, bauen ihre kognitiven Fähigkeiten durch Schlüsse aus ihren eigenen Erfahrungen auf, sie können nicht deduziert werden; es ist klar, dass eine solche Position sozusagen von selbst in einen Gegensatz zur Schule gerät, die Lernen für grosse Zahlen in eine organisierbare Form bringen muss.

Nach seinem Studium in Deutschland untersuchte Rice die Qualität der öffentlichen Schulen in den Vereinigten Staaten. Er wird heute zu Recht der „Vater der Evaluation“ genannt, weil durch ihn erstmalig die Schulen mit einem Blick von aussen und vergleichend observiert wurden. Rice war kein Inspektor, sondern ein unabhängiger Beobachter, der für seine erste Studie sechs Monate in Schulen verbrachte, den Unterricht beobachtete und seine

¹ Terrel Howard Bell (1921-1996) war High School-Lehrer und promovierte 1961 in Pädagogik an der University of Utah. Bell war Utah Commissioner of Higher Education in Utah, bevor er 1981 in das erste Kabinett von Präsident Reagan eintrat. Bell trat Ende 1984 aus persönlichen Gründen zurück. Seine Autografie schildert das Zustandekommen von *A Nation at Risk* (Bell 1988).

² David Pierpont Gardner war seit 1973 und so zum Zeitpunkt des Berichts Präsident der University of Utah. Von 1983 bis 1992 war Gardner der 15. Präsident der University of California. Gardner promovierte 1966 an der University of California und wurde danach auf eine Professur for Educational Leadership an die Graduate School of Education an die University of Utah berufen.

³ Leonard Porter Ayres (1879-1946) war nach seinem Studium sechs Jahre lang als Lehrer und Superintendent auf Puerto Rico tätig. Ayres war Statistiker und leitete von 1908 an das Bureau of Education and Statistics der Russell Sage Foundation.

⁴ Joseph Mayer Rice (1857-1934) stammte aus Philadelphia und promovierte 1881 am College of Physicians and Surgeons in New York. Er war zunächst Arzt am Mount Sinai Hospital in New York und eröffnete 1886 eine eigene Praxis.

Notizen auswertete. Er besuchte insgesamt 36 Schuldistrikte und veröffentlichte seine Resultate 1893 in dem Buch *The Public School System in the United States* (Rice 1893). Rice nannte sich den Anwalt (advocate) des Kindes (ebd., S. 4); er wollte wissen, nicht welche guten Absichten die Lehrkräfte vertreten, sondern wie die Schülerinnen und Schüler im Unterricht tatsächlich lernen. Absichten sind Hypothesen, die über den tatsächlichen Effekt nichts aussagen.

Entsprechend waren die Beurteilungskriterien. Der Bericht von Rice basierte auf folgenden Parametern:

- Das Erscheinungsbild der Klassenzimmer,
- die Haltung der Lehrkräfte gegenüber den Kindern,
- die Praxis des Rezitierens,
- die Art der Beschäftigung der Schüler (busy work),
- Antworten der Lehrkräfte auf zwölf allgemeine Fragen,
- Versammlungen der Lehrerinnen und Lehrer,
- Anstrengungen der Lehrkräfte zur eigenen Fortbildung (ebd., S. 5).

Das Buch war ein Erfolg, weil der Autor kein Blatt vor den Mund nahm; Rice beschönigte nichts, sondern beschrieb schonungslos die Misstände, also:

- inadäquate Methoden des Unterrichts,
- Apathie der Schüler,
- Inkompetenz der Lehrkräfte,
- politische Patronage bei der Besetzung der Ämter
- und zahlreiche Fälle von Korruption.

Es gab noch weitere Resultate: Die schlechtesten Lehrkräfte wurden durchgehend in den unteren Klassen eingesetzt, weil die Schulen davon ausgingen, dass die jüngeren Schüler weniger Aufwand und so geringere Kompetenz erfordern. Durchgehend beschrieb Rice die Praxis des *Rote-Learnings*, also des ständigen Wiederholens und Nachsprechens von Mustersätzen, die auswendig gelernt wurden. Üblich war auch die Methode der *Group Recitations*, also des Nachsprechens der ganzen Klasse. Auf ein Signal des Lehrers musste die Klasse Merksätze wiederholen, so lange, bis sie „sassen“, wie man in der deutschen Lehrersprache des 19. Jahrhunderts sagte. Diese Praxis war verbreitet und vor den Studien von Rice aber kaum öffentlich.

Von den 36 Distrikten, die Rice besuchte, wurden nur vier als positiv und fortschrittlich beurteilt.⁵ Vor allem dieser Befund löste den Schock aus, weil damit das System der amerikanischen Schulen insgesamt als dringend reformbedürftig hingestellt wurde, wogegen die Lehrkräfte Sturm liefen. Sie wollten keineswegs die Verantwortung für die Resultate übernehmen, die im Übrigen auch nirgendwo genau beschrieben war. Die grössere Öffentlichkeit allerdings akzeptierte den Befund, der sich mit den Erfahrungen der Eltern deckte und so alles andere als unglaubwürdig war. Aber Rice zog sich den Zorn der Lehrkräfte zu, die ihm Auskunft gegeben und ihre Klassenzimmer für seine Untersuchung geöffnet hatten. Andere Lehrkräfte befürchteten, von ihm besucht zu werden. Aber das Prinzip der externen Visite oder der Evaluation unabhängig von der Schulaufsicht war etabliert.

⁵ Alle diese Schulsysteme waren im Mittleren Westen, in Indiana, Indianapolis, in La Porte, Indianapolis, die Lincoln Schule in Minneapolis sowie die Francis Parker Schule im Cook County.

Besser wurde die amerikanische Schule dadurch offenbar nicht, denn sonst hätte Levin achtzig Jahre später und in Unkenntnis dieser Geschichte nicht ihre schlechte Qualität beklagen und eine unabhängige Datenerhebung als Abhilfe vorschlagen können. Sein kybernetisches Modell umfasst vier Komponenten, die sich so bestimmen lassen:

- *Ziele*: Erreichbare Endzustände eines Prozesses.
- *Mittel*: Massnahmen, die sicherstellen, dass der begonnene Prozess in Richtung Ziele verläuft.
- *Feedbackschleifen*: Regelmässige Überprüfung der erreichten Qualität.
- *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 Kompetenzen der Lehrkräfte oder die Qualität der Schulen. Zugrunde gelegt wird sinngemäss oft das Modell von Levin. „Qualität“ wird mit Hilfe von Feedbackschleifen *im* Prozess beschrieben. Damit wird es theoretisch möglich, Lernen nicht nur auf Ziele auszurichten, sondern den fortlaufenden Prozess zu beeinflussen. Es ist im Sinne dieses Prozessmodells nicht zureichend, aus allgemeinen Lernzielen operative abzuleiten und dann dabei stehen zu bleiben; entscheidend ist, die tatsächlichen Erfahrungen einschätzen und gegebenenfalls auch korrigieren zu können. Die Schule übernimmt so Verantwortung für ihren eigenen Lernprozess.

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 zugute kommen oder in ihr Kosten verursachen.⁶ 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.⁷ Zu diesen Formen gehören vor allem Inspektionen, Evaluationen und Leistungsmessungen. Getestet wurde an amerikanischen Schulen schon vor der Einführung des Intelligenztests im Jahre 1920.⁸ Instrumente wie diese waren Jahrzehnte lang 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.

Seitdem ist das Thema politisch stark aufgewertet worden, ohne an dem grundlegenden Modell von Levin viel zu verändern. Es beschreibt, was Qualitätssicherung im Kern ausmacht und wozu sie in der Praxis dienlich ist.

- 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.
- 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.

⁶ „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).

⁷ 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.).

⁸ 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.

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 demokratiethoretisch gut begründetes Postulat. Die Frage ist nur, was „Resultate“ sind, wie sie zustande kommen und was davon in welcher Form die Öffentlichkeit erreicht. Das ist mein nächster Punkt.

3. *Qualitätssicherung und Kommunikation mit der Öffentlichkeit*

Im Modell von Levin fehlt eine Mengenvorstellung, es sagt nicht, wie *viele* 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 Ebene 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. Vor der Verbreitung des *worldwideweb* und der Gewöhnung an die damit verbundenen Vorteile wäre das Gesetz über die Information und den Datenschutz weder möglich noch sinnvoll gewesen.

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 dichter werdenden 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.

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 ausschlaggebend 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.

Wer Blogs und Leserbriefe zum Vorfall im Schulhaus Borrweg analysiert hat, stellt fest, dass sofort verallgemeinert wird und oft auch das Erziehungssystem pauschal unter Anklage gestellt ist, ohne den Fall als solchen zu beachten. Er alarmiert durch die Darstellung und ist das Beispiel, das zum allgemeinen Bild passt oder an dem die Tendenz abgelesen wird, die die Berichterstattung mindestens unterschwellig oft auch nahelegt. Einen erkenntnistheoretischen Vorbehalt gibt es nicht oder umso weniger, je schneller und zugespitzter die Kommunikation wird. Allerdings sind dann die Themen auch kurzlebig und wird schnell vergessen, was gerade noch hoch aktuell war.

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 Schuldvorwürfen 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.

4. Zwei unterschiedliche Fallbeispiele

Die Funktionsweise medialer Kommunikation lässt sich an einem der Fallbeispiele zeigen. Was als „ETH-Ranking“ nun schon fast Geschichte ist, hat für helle Aufregung gesorgt, 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. *Ratemyteachers.com* lautet die Webanschrift eines amerikanischen Anbieters, der Gelegenheit bietet, die Lehrkräfte zu bewerten und dann eine Rangfolge für jede Schule zu erstellen.⁹ 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*¹⁰ oder für die Hochschulen *meinprof.de*.¹¹ Hier werden einzelne Lehrkräfte beurteilt und miteinander verglichen, ohne dass irgendeine Behörde eingreifen könnte. Eltern können auf der Seite

⁹ <http://www.ratemyteachers.com>.

¹⁰ <http://www.spickmich.de> Die Seite wird als Hilfe für die Lehrer und als Orientierung für die Eltern angeboten.

¹¹ <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.

*schulradar.de*¹² Wertungen abgeben und so die Entwicklung in den Schulen ihrer Kinder kommentieren.¹³

Die Frage ist, wie Kunden auf diese Art Governance reagieren. Auf jeden Fall sind Rankings 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, kantonal 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 untergangene 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.

Solange keine wirklich politische Entscheidung gefallen ist, bleibt das Reizthema erhalten, und damit zugleich alle Erzählungen von Fortschritt und Untergang, die sich um das

¹² <http://www.schulradar.de> Es gibt Top-Listen mit den Schulen, die von den Eltern am besten bewertet wurden.

¹³ 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.

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).

Auch hier kann ich mich auf ein Fallbeispiel beziehen, nämlich die 2004 in Auftrag gegebene Langzeitstudie über die Entwicklung des Lernstandes im Kanton Zürich. Wie Sie wissen, sind die Schülerinnen und Schüler bei Schuleintritt untersucht worden und nachfolgend am Ende dritten Klasse; sie werden derzeit in der sechsten Klasse getestet und dann abschliessend nochmals in der neunten Klasse. Die Ergebnisse der ersten beiden Studien (Moser/Stamm/Hollenweger 2005; Moser/Keller/Tresch 2008; Moser/Hollenweger 2008) sind bekannt und sogar vergleichsweise breit kommuniziert worden, auch weil die Studie tatsächlich sehr aufschlussreich ist.

Ich erinnere an einige der Ergebnisse:

- 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. Damit bin ich bei meinen fünf Fragen, die ich nach diesen Vorbemerkungen jetzt wohl beantworten kann.

5. Antwort auf die Fragen

Eigentlich verdichten sich die Fragen zu einer einzigen: Wie kommuniziert man als öffentliches Organ mit der Öffentlichkeit, wenn diese nicht einfach aus Bürgerinnen und Bürgern besteht, sondern medial vermittelt wird? Die allgemeine Antwort lautet: so transparent und offen wie möglich, aber in Einschätzung der Risiken, die die heutige Medienlandschaft mit sich bringt. Es ist ein Unterschied, ob sich Bürgerinnen und Bürger über die Arbeit einer Behörde informieren wollen oder die Medien ein Thema gestalten, das die öffentliche Aufmerksamkeit fixieren soll.

Die erste Frage geht eher ins Allgemeine; sie lautete, nochmals gesagt, so:

Was bedeutet das IDG, also der Grundsatz der Öffentlichkeit aller (nicht personalisierten) Bildungsdaten, für eine gelingende Qualitätspolitik?

Die Bildungsdirektion ist gemäss Gesetz ein öffentliches Organ des Kantons Zürich. Alle öffentlichen Organe, so das Gesetz, gestalten den Umgang mit Informationen so, dass sie „rasch, umfassend und sachlich informieren“ (IDG §4). „Informationen“ sind alle „Aufzeichnungen, welche die Erfüllung einer öffentlichen Aufgabe betreffen, unabhängig von ihrer Darstellungsform und ihrem Informationsträger“ (IDG §3). Und das öffentliche Organ „verwaltet seine Informationen so, dass das Verwaltungshandeln nachvollziehbar und die Rechenschaftsfähigkeit gewährleistet ist“ (IDG §5). Die Qualitätspolitik im Bereich der Bildung ist eine öffentliche Aufgabe, insofern - sagt mein juristischer Laienverstand - gelten die Bestimmungen des Gesetzes.

Alle abgeschlossenen Studien, die mit Qualitätssicherung zu tun haben, sind auch bisher schon veröffentlicht worden, oft als wissenschaftliche Publikation und immer als Download, also in der Form, die den schnellsten Zugriff gestattet. Das geschieht offenbar „rasch, umfassend und sachlich“. Die Ergebnisse sind kein Geheimnis, sondern dienen gerade der Information der Öffentlichkeit. Damit ist natürlich nichts darüber ausgesagt, wie die Daten verwendet werden, sie sind einfach nur für jeden Interessenten und jede Interessentin zugänglich. Eine Zielgruppe gibt es nicht, gefragt wird nicht, wer diese Studien liest, sie sind einfach nur zugänglich.

Das Gesetz will erreichen, dass das Handeln der öffentlichen Organe „transparent“ gestaltet wird, um so die Kontrolle des staatlichen Handelns durch die Bürgerinnen und Bürger zu „erleichtern“ (IDG §1). Dagegen ist aus demokratiethoretischer Sicht überhaupt nichts einzuwenden, im Gegenteil, Transparenz im Verwaltungshandeln ist im Grundsatz sehr zu begrüßen. Das gilt im Kern auch für die Regelung, dass dabei „nach Möglichkeit auf die Bedürfnisse der Medien“ Rücksicht zu nehmen ist (IDG §15). Unabhängige und kritische Medien sind für demokratische Meinungsbildung unverzichtbar. Daher lautet die Antwort auf die Frage: *Das Prinzip der Öffentlichkeit unterstützt die Qualitätspolitik; sie ist auf Öffentlichkeit im weiteren Sinne angewiesen.*

Die zweite Frage, die hier anschliesst, lautet:

Was bedeutet dieses Öffentlichkeitsprinzip auch für die Durchführung von Datenerhebungen?

Prinzipiell gilt: Die Datenerhebungen müssen so anlegt sein, dass sie mit dem Öffentlichkeitsprinzip verträglich sind. Daten, die zur Bearbeitung öffentlicher Aufgaben erhoben wurden, sind selber öffentlich, sofern nicht besonders geschützte Personendaten betroffen sind. So gesehen, gibt es nicht mehr, was früher „Ressortforschung“ genannt wurde, also eine Datenerhebung rein für den internen Gebrauch, wenn diese, was aber eigentlich immer der Fall ist, eine öffentliche Aufgabe tangiert.

Im Detail stellen sich aber ungleich schwierigere Fragen als im Grundsätzlichen. Eine habe ich bereits angesprochen: Wenn jede Leistungsmessung und jede vergleichende Schulevaluation ein Ranking impliziert, weil letztlich Abstände erfasst werden, heisst das, die Ergebnisse auch in der Form eines Rankings beschreiben zu müssen? Hier kommt es zunächst natürlich auf den Auftrag an, aber es dürfte zunehmend heikel werden, Rankings als Darstellungsform auszuschliessen, wenn die Ergebnisse sie enthalten. Letztlich ist die Entscheidung politisch, aber es sollte klar sein, dass es keinen empirischen Vergleich geben kann ohne die Darstellung von Abständen. Die Folgen wollen bedacht sein.

Bei Gymnasien herrscht im Kanton Zürich die freie Schulwahl, die praktisch bislang vermutlich noch nicht viel verändert hat, weil alle Schulen die gleiche Berechtigung vergeben. Würde man den Hochschulzugang auch an die Qualität der einzelnen Schule binden und würden die Unterschiede in Form eines Rankings dargestellt, wäre das sofort anders. Rankings im Bereich der obligatorischen Schule würde das Thema Freie Schulwahl enorm beschleunigen. Wir sind nicht in Schweden, wo ein Gesetz vorschreibt, dass die Gemeinden über gleiche Ressourcen für die Bildung verfügen müssen, was nur mit einer harten Umverteilungspolitik durchgesetzt werden konnte, die in einer direkten Demokratie - zum Glück - nicht möglich ist. Aber wenn die besten Schulen, öffentlich gemacht, in den Gemeinden stehen, die über die meisten Ressourcen verfügen, haben wir ein Problem, verursacht durch Datenerhebung.

Eine weitere Frage ist die Interpretation der Daten. Die öffentliche Wahrnehmung der Ergebnisse und der mitgelieferten Erklärungen ist selektiv und Interessen gesteuert: Resultate werden weit mehr zur Bestätigung einer bereits bestehenden Meinung verwendet als zu deren Korrektur; und es gibt so etwas wie eine Perfektionserwartung, die keine schlechten Daten verträgt und sie doch ständig vermutet. Das „Wording“ in vielen Evaluationsberichten reagiert genau darauf, es dosiert die Zumutungen. Tatsächlich ist die Sprache der Berichte ausschlaggebend für die Akzeptanz. Es macht einen Unterschied, ob Defizite beschrieben werden, oder - daran anschliessend - künftige Potenziale, und es ist nicht egal, wie klein oder gross die Defizite erscheinen.

Für die Beschreibung der Resultate ist eine Form anzuraten, die Auskunft gibt nicht nur über die Methode der Datenerhebung, sondern auch über die Reichweite der Befunde und Aussagen. Auch wenn die Medien das ignorieren, es sichert das staatliche Handeln. Was also die jeweilige Studie aussagt und vor allem auch, was *nicht*, sollte genau bestimmt werden. Das sollte explizit zum Auftrag gehören und die Behörden müssen sich darauf verlassen können. Jede Studie hat damit zu kämpfen, dass mehr hineingelesen wird als die Daten hergeben, und so kommt es darauf an, den Frame of Reference zu bestimmen und genau darauf zu achten, wie die Ergebnisse beschrieben werden, denn nur die erreichen die Öffentlichkeit. Oft geschieht das lediglich in Form von ausgewählten plakativen Zahlen, auf die also besonders geachtet werden muss. Die Antwort auf die zweite Frage lautet knapp so: *Die Kommunikation der Ergebnisse müssen auf die Bedürfnisse der Bürgerinnen und Bürger sowie die Anforderungen der Medien eingestellt sein.*

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.

Damit bin ich bei der Bildungsphilosophie angelangt und kann so auf die dritte Frage eingehen:

Welche Leitideen und Strategien einer „Good Educational Governance“ im Umgang mit Bildungsdaten schlagen Sie für den Kanton Zürich vor?

Forschung hat kein Wahrheitsmonopol und kann nicht von sich aus eine bestimmte Politik legitimieren. Auch und gerade im Auftrag der Politik muss die Forschung ihre Unabhängigkeit bewahren. Die Ergebnisse der Studien müssen vollständig öffentlich werden, auch dann, wenn sie aus Sicht des Auftraggebers nicht positiv sind oder heikel erscheinen. Anders ist Glaubwürdigkeit nicht zu erreichen und nur mit einer glaubwürdigen Forschung macht die Projektvergabe überhaupt Sinn.

„Governance“ meint mehr als Regierungshandeln im formalen Sinne, der Ausdruck impliziert auch eine Klugheitsregel und so die Frage, ob alles gemacht werden muss, was gemacht wird. Daher ist ein überschaubares Mass an Audits anzustreben, es muss nicht alles,

was irgendwie strittig ist, evaluiert werden, obwohl genau das den Streit beruhigt - bis die Daten vorliegen, die ja interpretiert werden müssen. Es ist naiv, sich vorzustellen, dass Forschung politischen Streit entscheiden könnte, bei dem ja jede Seite eigene Experten hat oder haben kann. Genau aus diesem Grunde muss bezogen auf Good Governance die Unabhängigkeit gewahrt sein.

Im Blick auf Forschungsaufträge muss zudem abgewogen werden, welchem Zweck die Datenerhebung dient und was damit am Ende mit den Resultaten angefangen werden soll. Behörden sind keine Forschungsinstitute, sie können nicht beliebig viele Daten vertragen, sondern müssen den Strom bündeln und für das jeweilige Geschäft nutzbar machen. Nur dann sind Daten mehr als *nice to have*. Die Umsetzung ist in einem Bereich, der hundertfünfzig Jahre gut ohne empirische Datenerhebung ausgekommen ist, durchaus problematisch. Daher sollte immer auch die Zielgruppe vor Augen stehen; die Akteure in den Schulen reagieren auf empirische Befunde anders als die Medienvertreter, die Schulaufsicht, die Behörde selbst oder die wissenschaftliche Öffentlichkeit. Daher lautet die Antwort auf die Frage: *Good Governance heisst bei hoher Transparenz kluge Datennutzung*.

„Bündeln“ meint auch, die Ergebnisse mit einer Interpretation versehen, die für die Öffentlichkeit verständlich ist. Damit bin ich bei der nächsten Frage.

Wo sehen Sie Regelungsbedarf?

Wahrscheinlich ist es ziemlich nebensächlich, wo ich Regelungsbedarf sehe, entscheidend ist, dass sich die Behörde darauf verständigen kann. Ich glaube tatsächlich an Regelungen. Forschungsdaten - besonders in den Bereichen von Leistungsmessung, Schulevaluation und Lernstandserhebung - sind sensibel, weil, wie gesagt, immer Vergleiche und Abstände impliziert sind. Daher sollte die Reichweite der Aussagen bestimmt sein, bevor die Öffentlichkeit informiert wird. Bei Nachfragen muss es eine Sprachregelung geben, die Missverständnisse ausschliesst und mit den Ergebnissen übereinstimmt. Das lässt sich regeln und wird am Ende der Tagung sicher auch geregelt sein.

Angesichts der Qualität der meisten Studien sollte die Strategie grundsätzlich immer offensiv sein; die Bildungsdirektion hat nichts zu verstecken. Im Gegenteil, es ist ein Qualitätsnachweis, wenn wichtige Entscheide datengestützt gefällt werden, genau das ist ja auch ein öffentliches Anliegen. Behörden handeln in einem politischen Umfeld, aber dies vor dem Hintergrund und unter Nutzung von einschlägigen Forschungen, wie es ja auch bewährte Praxis ist. Die Antwort auf die Frage lautet: *Regelungsbedarf besteht vor allem in der innerbehördlichen Abstimmung sowie in der Kommunikation nach aussen*.

Bleibt die letzte Frage:

Was würde dies für den Umgang mit den Bildungsdaten der eingegebenen Fallbeispiele bedeuten?

Natürlich überlasse ich die Antwort auf diese Frage den Arbeitsgruppen. Von meiner Seite nur so viel: Die Fallbeispiele sind hochgradig verschieden. Was für die Befragung der Absolventinnen und Absolventen der Zürcher Gymnasien als Problem gilt, nämlich die Unterschreitung des Limits der Zufriedenheit um einen Punkt,¹⁴ hat mit Klassencockpit oder

¹⁴ In der Umfrage 2006 stimmten 74% der Ehemaligen der Frage zu, ob sie durch den gymnasialen Unterricht gut auf das Studium vorbereitet wurden. Die Quote liegt leicht unter dem von Kantonsrat vorgegeben Limit von 75% (Befragung 2006).

Stellwerk nichts zu tun, und die Fragen der Bildungsstatistik sind auf Quims nur so weit anwendbar, wie die Statistik reicht. Allerdings sind alle diese Beispiele immer auch Kommunikationsprobleme, am wenigsten wohl die gut geführte Bildungsstatistik; aber Quims etwa steht unter öffentlicher Beobachtung und Daten, die das Projekt betreffen, und müssen überzeugend kommuniziert werden, auch wenn vielleicht manche nur auf schlechte Daten warten. Daher lautet die Antwort allgemein: *Alle erhobenen System- und Leistungsdaten werden öffentlich, grössere Projekte brauchen neben der Begleitung eine Kommunikationsstrategie.*

Nehmen wir Stellwerk. Der Test ist bekanntlich 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. In einer Evaluation 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). Das lässt sich durchaus als Erfolg darstellen, der die Daten für sich hat.

Mein Credo für die Schulentwicklung, ganz unabhängig von dem Gesetz über Information und Datenschutz, lautet so: Es nutzt niemandem, wenn Evaluationsdaten woher sie auch stammen, verschwinden und nicht zur Weiterentwicklung der Schule Verwendung finden. „Schubladiesieren“ nennt man das in der Schweiz, „Verschlussache“ in der deutschen Behördensprache. Das Gegenteil ist bei einer evidenzbasierten Schulentwicklung angesagt, nämlich Transparenz und Offenheit. Das gilt auch und ganz besonders für schlechte Daten. Zu guten Daten sich zu bekennen, verlangt nicht viel Mut; schlechte Daten offen zu kommunizieren, verlangt Mut, aber nur dann kann die betreffende Schule vorankommen. Schlechte Daten haben immer Ursachen, wer die Ursachen nicht thematisiert, verschenkt die Entwicklung. Das Risiko ist nur dann zu gross, wenn es einen öffentlichen Pranger gibt und persönliche Konsequenzen drohen.

Literatur

A Nation at Risk: The Imperative for Educational Reform. A Report to the Nation and the Secretary of Education United States Department of Education by The National Commission on Excellence in Education. April 1983. Washington, DC: U.S. Department of Education 1983.

Ayres, L-P.: Measuring Educational Processes through Educational Results. In: School Review Vol. 20 (1912), S. 300-309.

Befragung ehemaliger Zürcher Mittelschülerinnen und Mittelschüler. Erstellt vom Statistischen Amt des Kantons Zürich. Zürich: Statistisches Amt des Kantons Zürich 2006.

Bell, T.H.: The Thirteenth Man. A Reagan Cabinet Memoir. New York/London: The Free Press, Collier-Macmillan Publishers 1988.

Gesetz über die Information und den Datenschutz des Kantons Zürich vom 12. Februar 2007.

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./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.
- Rice, J.M.: The Public-School System of the United States. New York: The Century Co. 1893.