

Schriftenreihe zu «Bildungssystem und Humanentwicklung»

Berichte aus dem Forschungsbereich
Schulqualität & Schulentwicklung

Schule wohin?

Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert

**Katharina Maag Merki
Anita Sandmeier
Patricia Schuler
Helmut Fend
(Hrsg.)**



Universität Zürich
Pädagogisches Institut Lehrstuhl Prof. Dr. H. Fend

FS&S
Forschungsbereich Schulqualität
& Schulentwicklung

FOKUS

Katharina Maag Merki	4
Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert. Problemzusammenhang und Forschungsdiesiderate	
Helmut Fend	15
Systemsteuerung im Bildungswesen – Anschlussfähigkeiten an die Schulwirklichkeit	
Hannelore Faulstich-Wieland	28
Geschlecht im Schulalltag – Plädoyer für eine Entdramatisierung	
Eckhard Klieme	40
Zur Bedeutung von Evaluation für die Schulentwicklung	
Silvia Grossenbacher	62
Versuch einer Zusammenschau: Bei der Übergabe des Steuers in voller Fahrt das Ziel nicht aus den Augen verlieren!	

„Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert“

Schriftenreihe zu „Bildungssystem und Humanentwicklung“: Berichte aus dem Forschungsbereich Schulqualität & Schulentwicklung/Maag Merki, K., Sandmeier, A., Schuler, P. & Fend, H. (Hrsg.). – Zürich: FS&S – Forschungsbereich Schulqualität & Schulentwicklung, Universität Zürich, 2005

ISBN-Nummer 978-3-905651-02-7

Nachdruck nur mit Genehmigung des Forschungsbereich
Schulqualität & Schulentwicklung, Universität Zürich

Verlag

FS&S – Forschungsbereich Schulqualität & Schulentwicklung
Pädagogisches Institut, Lehrstuhl Prof. Dr. H. Fend, Universität Zürich,
Freiestrasse 36, 8032 Zürich

Layout

Goetz Desktop GmbH, Ifangstrasse 1, 8153 Rümlang

Druck

Kaspar Schnelldruck AG, Birkenweg 2, 8304 Wallisellen

Am 24. September 2004 fand anlässlich des zehnjährigen Jubiläums des FS&S – Forschungsbereich Schulqualität & Schulentwicklung an der Universität Zürich ein Symposium zum Thema „Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert“ statt.

Ausgehend von den vor mehreren Jahren neu eingeführten Konzepten der Qualitäts- und Standardsicherung im Bildungswesen – die den Einzelschulen grössere Gestaltungs- und Handlungsfreiheit einräumen – und der Frage der Sicherung Qualitätsstandards durch die Implementation eines integralen Monitoringsystems, lautete das Kerninteresse des Symposiums:

Wie lässt sich die Beziehung zwischen Qualitätsmanagement, institutionellen Regelungen, Prozessen auf Schul- und Unterrichtsebene sowie den daran beteiligten Akteuren so gestalten, dass die Schule sowohl ihre Schülerinnen und Schüler individuell bestmöglich fördert als auch die Grundlagen eines demokratischen und leistungsorientierten Gemeinwesens sichert.

In diesem Band sind Beiträge vereinigt, die von den Referent/innen am Symposium vorgetragen wurden. Ergänzend dazu führt Katharina Maag Merki als ehemalige Leiterin des FS&S in die Thematik ein, indem sie in ihrem Beitrag den Erkenntniszusammenhang des Symposiums skizziert und die obige Fragestellung in den aktuellen Forschungskontext einbettet. Im Zentrum steht dabei das Konzept der „selbstständigen“, „teilautonomen“ oder „lernenden Schule“, welche den Auftrag hat, aufgrund systematischer Reflexion und Analysen Schulentwicklungsmassnahmen zu ergreifen und gezielt an der Qualität der Schule zu arbeiten. Im Beitrag wird der aktuelle Forschungsstand zu Funktions- und Wirkungsweisen von „lernenden Schulen“ aufgezeigt und daraus Forschungsdesiderate abgeleitet.

Im Beitrag von Helmut Fend wird der Fokus auf neue Formen der Systemsteuerung des Bildungswesens gelegt. Diese werden in ihrer historischen und gesellschaftlich bedingten Unterschiedlichkeit vorgestellt. Fend nennt dies die „Grammatik“ eines Bildungswesens, die analog zu sprachlichen Regelsystemen das Handeln der Akteure leitet. Die Grammatik des Schweizer Bildungswesens unterscheidet sich von jener des deutschen erheblich. Auch die neuen Formen der Systemsteuerung über Bildungsstandards, Tests, externe Evaluation und Stärkung

der Schulautonomie repräsentieren eine neue Grammatik. Ihre Auswirkungen auf die innerschulischen Prozesse und auf die Bildungswirkungen haben heute noch weitgehend die Gestalt von Wünschen und Erwartungen und nicht die von empirisch erhärteten Erkenntnissen. Dies ist nach Fend dringend änderungsbedürftig

Der Beitrag von Hannelore Faulstich-Wieland fokussiert die schulischen Prozesse unter einer geschlechtsspezifischen Fragestellung. Schulentwicklung wird dabei auf der Basis der Analyse von Ungleichheit oder Gleichheit zwischen den Geschlechtern beschrieben, wobei Ansätze einer „geschlechterbewussten Schule“ hinsichtlich ihrer Effekte analysiert werden. Zentraler Fokus bildet dabei die Frage, welche dieser Massnahmen zu Dramatisierungen oder Entdramatisierungen der Geschlechterdifferenzen im Schulalltag führen.

Im Beitrag von Eckhard Klieme geht es um die Frage, welchen Beitrag Schulevaluation als Instrument des Qualitätsmanagements für die Schulentwicklung leisten kann. Nachdem er den theoretischen Hintergrund mit Fokus auf Entscheidungsfelder und verschiedenen Paradigmen der Evaluation skizziert hat, zeigt er die Bedeutung von Schulevaluation als Teil eines systematischen Bildungsmonitorings auf, das als „integriertes System der Qualitätssicherung und Qualitätsentwicklung“ konzipiert werden soll. Anhand empirischer Beispiele legt er danach dar, welche Erkenntnisse das Schulsystem aus einem solchen System der Qualitätssicherung und -entwicklung gewinnen kann.

Im abschliessenden Beitrag stellt Silvia Grossenbacher als Diskutantin die am Symposium gehaltenen Referate zueinander in Beziehung¹. Sie verweist darauf, wie wichtig die Verzahnung der verschiedenen Ebenen schulischer Qualität und die Kooperation zwischen Forschung, Politik sowie Schulpraxis ist.

Wir danken allen Autor/innen für ihren Beitrag zum Gelingen dieses Bandes. Herrn Martin Nigg danken wir für das Setzen der Marginalien und das Korrekturlesen, Frau Evelyn Hug für das Korrekturlesen und ihr Mitdenken in allen Belangen rund ums Symposium und der Herausgabe dieses Bandes.

*Zürich, im Oktober 2005
Anita Sandmeier für die Herausgeberschaft*

¹ Sie bezieht sich dabei auf die gehaltenen Referate, die zum Teil von den hier publizierten Texten abweichen können. Vom Referat von Prof. Dr. J. Baumert, auf das sich Frau Dr. S. Grossenbacher bezieht, besteht kein Artikel.

Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert. Problemzusammenhang und Forschungsdesiderate

Katharina Maag Merki

Schulen haben die Pflicht, die gewonnene Teilautonomie zur Qualitätsentwicklung zu nutzen.

Vor knapp 20 Jahren zeichnete sich im Zusammenhang mit der Steuerung von Bildungssystemen eine Neuorientierung ab. Waren bis Ende der 70er-Jahre zur Regulierung der Schulqualität insbesondere Massnahmen auf der Makroebene, wie beispielsweise die Restrukturierung der Oberstufe oder die Einführung neuer Lehrpläne bedeutsam, so rückte seit Mitte der 80er-Jahre die Einzelschule als Gestaltungs- und Handlungseinheit ins Zentrum der Aufmerksamkeit (Fend, 1986, 1998; Rolff 1995). Nicht nur in der Schweiz, sondern auch international, wird den einzelnen Schulen in der Folge eine erhöhte Gestaltungsfreiheit zugestanden (Burkard & Eikenbusch, 2002). Im Rahmen des so genannten „Output-Steuerungsmodells“ haben die Schulen die Pflicht, diese „Teilautonomie“ im Sinne der Qualitätsentwicklung zu nutzen, indem sie über eine systematische Analyse ihrer Stärken und Schwächen auf der Basis von internen Evaluationen Schulentwicklungsmassnahmen einleiten und damit gezielt an der Qualität ihrer Schule arbeiten. Über ihre Arbeit haben sie zudem gegenüber der Aufsichtsbehörde Rechenschaft abzulegen. Periodische, externe Evaluationen zur Erfassung des „Outputs“ der Schulen ergänzen diese interne Sicht und sorgen dafür, dass die institutionellen Weiterentwicklungen der einzelnen Schulen dem gesellschaftlichen Auftrag entsprechen. In föderalistisch organisierten Bildungssystemen wie in der Schweiz oder in Deutschland ist dies in erster Linie Aufgabe des einzelnen Kantons oder des einzelnen Bundeslandes, in dem beispielsweise Leistungsvergleichsstudien durchgeführt und die Kompetenzen der Schüler/innen am Ende spezifischer Jahrgangsstufen erfasst werden. Zusätzlich zu diesen mehr regionalen Monitoringsstudien zeichnet sich im deutschsprachigen Raum über die Implementierung von Bildungsstandards oder der Festlegung eines Kerncurriculums auch die Entwicklung hin zu einem nationalen Monitoring ab.

Neues Konzept geht von Diskrepanz zwischen bildungspolitischen Absichten und Realisierung in der Praxis aus.

Dieses neue Konzept von Qualitäts- und Standardsicherung geht somit im Gegensatz zu den herkömmlichen Strategien davon aus, dass es eine systematische Diskrepanz zwischen bildungspolitischen Absichten und ihrer Realisierung in der Praxis geben kann, welche es zu reduzieren gilt. Die Frage stellt sich hingegen, ob sich durch diese Veränderung der „Grammatik eines Bildungswesens“ (Fend, in diesem Band) und der neuen Form der Standard- und Qualitätssicherung die Qualität und Funktionalität des Bildungssystems und der einzelnen Schule massgeblich verbessern kann und welche Massnahmen auf der Prozessebene als wichtige Bedingungsfaktoren für eine produktive Schulentwicklung interpretiert werden können. Zur Beantwortung dieser Fragestellung ist der Fokus dabei in zentraler Weise auf die Entwicklung der Kompetenzen der Schüler/innen zu legen, wobei das Regelsystem schulischen Handelns in Abhängigkeit des Ziels, nämlich der Entwicklung eines optimalen Lernkontextes für die Schüler/innen, zu überprüfen ist.

Diese Problematik fokussierend, lautete die zentrale Fragestellung des Symposiums, das am 24. September 2004 zum zehnjährigen Bestehen des FS&S – Forschungsbereich Schulqualität & Schulentwicklung an der Universität Zürich zum Thema

„Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert“ stattfand, folgendermassen:

Wie lässt sich die Beziehung zwischen Qualitätsmanagement, institutionellen Regelungen, Prozessen auf Schul- und Unterrichtsebene sowie den daran beteiligten Akteuren so gestalten, dass die Schule sowohl ihre Schülerinnen und Schüler individuell bestmöglich fördert als auch die Grundlagen eines demokratischen und leistungsorientierten Gemeinwesens sichert.

Zentrale Fragestellung des Symposiums:

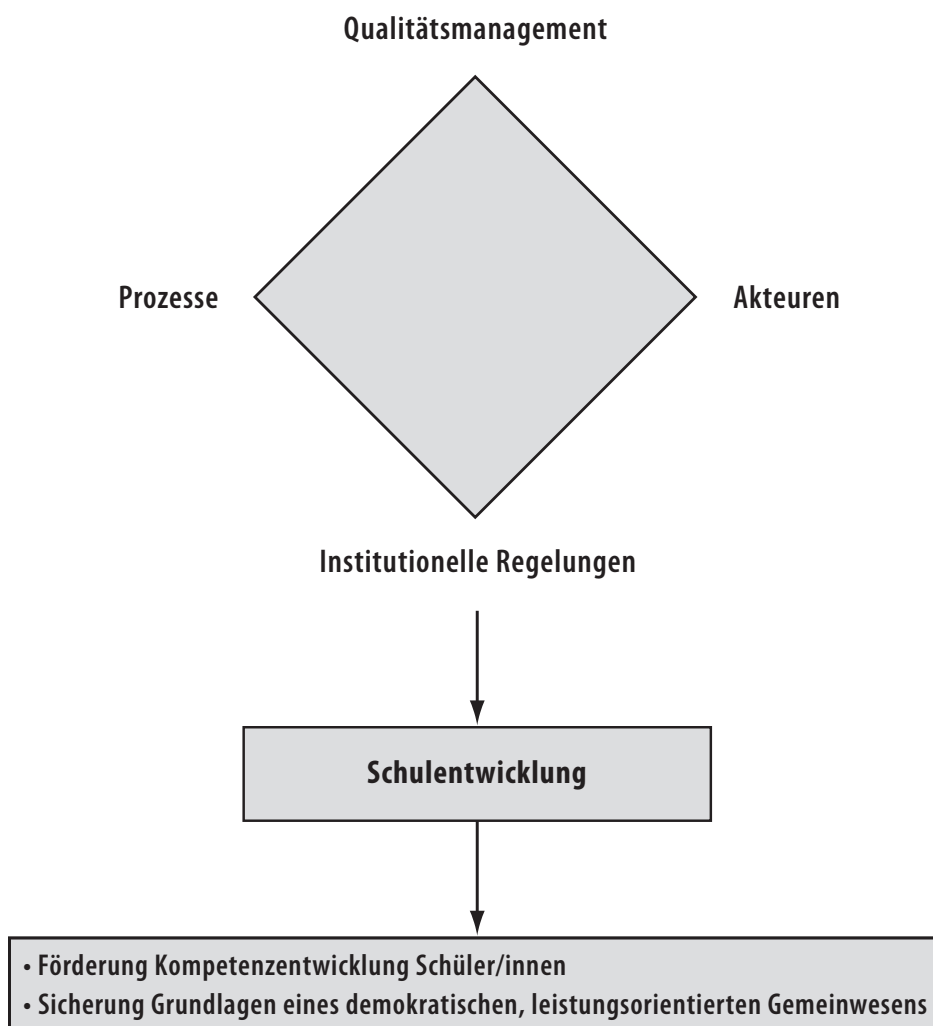


Abbildung 1: Zentrale Fragestellung des Symposiums

Im vorliegenden Artikel wird diese Fragestellung eingebettet in den aktuellen Forschungskontext, indem bisherige Studien fokussiert werden, die Hinweise über die Funktionalität des neuen Steuerungskonzeptes geben können. Aufgrund seiner Bedeutung als zentrale „Schaltstelle schulischer Qualität“ steht im Zentrum dieses Artikels das Konzept der „selbstständigen“, „teilautonomen“, „geleiteten“ oder „lernenden“ Schule, einer Schule also, die über die Zuweisung einer erhöhten Selbstständigkeit auf der Basis eines selbstreflexiven, bewusst gesteuerten und gestalteten Prozesses die eigene Arbeit in den Bereichen Organisation, Unterricht und Person professionalisiert.

Die Zuweisung erhöhter Gestaltungsfreiheit an die einzelne Schule basiert auf dem Konzept der „Lernenden Schule“

Schulen sind lose gekoppelte, in komplexe Handlungszusammenhänge eingebundene Systeme.

Die Zuweisung erhöhter Gestaltungsfreiheit an die einzelne Schule fokussiert die einzelne Schule als Einheit. Theoretische Basis bildet die organisationssoziologische Schulforschung. Es wird davon ausgegangen, dass Schulen lose gekoppelte, in komplexe Handlungszusammenhänge eingebundene Systeme sind, die einerseits über erheblichen Spielraum und somit Gestaltungsmöglichkeiten verfügen, andererseits aber Beschränkungen aufgrund externer und interner Faktoren erfahren (Dalin, 1986; Dalin, Rolff & Buchen, 1996). In Anlehnung an das Konzept einer lernenden Organisation (Argyris & Schön, 1999; Senge, 1996) und unter Berücksichtigung der Besonderheiten von Organisationen im Bildungsbereich (Fullan, 1999; Krainz-Dürr, 1999; Rolff, 1995) wird zwischen individuellem und organisationalem Lernen unterschieden. Diese Lernformen und ihre Kombination führen zur systematischen Weiterentwicklung und kontinuierlichen Professionalisierung von Individuum und Organisation und werden als funktional für die Zielerreichung des „Unternehmens Schule“ beschrieben.

Output-Steuerungsmodelle im Bildungsbereich nehmen Bezug auf dieses Konzept, indem nicht mehr nur die einzelne Lehrperson, sondern die Schule als Ganzes für die Qualität der Schule verantwortlich ist. Es wird also davon ausgegangen, dass die Schule als Organisation bzw. ihre Akteure grundsätzlich die Möglichkeiten haben, aufgrund der erhöhten Gestaltungsfreiheit eine kontinuierliche Verbesserung der schulischen Prozesse und damit der Leistungen der Schüler/innen zu erreichen.

Welche Erkenntnisse zeigen sich aufgrund empirischer Analysen?

Die Funktionalität des Konzepts der „Lernenden Schule“ ist als ambivalent einzuschätzen.

Eine Analyse der bestehenden Studien zeigt, dass die Funktionalität des Konzeptes zumindest als ambivalent, mit Sicherheit aber noch als unscharf eingeschätzt werden kann. Verschiedene Studien verweisen beispielsweise darauf, dass die Zuweisung erhöhter Gestaltungsfreiheiten alleine noch keine unmittelbaren Auswirkungen auf die Qualität der Schulen hat bzw. haben muss. So weist beispielsweise Ekholm (1997) in seiner Studie nach, dass sich trotz neuem Steuerungssystem zwar Einstellungsveränderungen bei den Lehrpersonen und Schulleiterinnen und Schulleitern ergeben haben, die innere Arbeitsroutine bzw. die Arbeitskultur aber unverändert fortbesteht. Wirkungen der strukturellen Veränderungen auf die Leistungen der Schüler/innen wurden nicht untersucht. Zu ähnlichem Ergebnis kommen Helsper et al. (Helsper, Böhme, Kramer & Lingkost, 1997, 2001) oder Munin (2001).

Leithwood & Menzies (1998) hatten in einer Metaanalyse von 83 empirischen Studien zwischen 1985 und 1995 gerade 11 Studien gefunden, in denen Effekte auf Schüler/innen untersucht worden sind. Die vorgefundenen Effekte waren in fünf Studien neutral oder negativ und einzig in drei Studien positiv. Hingegen konnten in 35 Studien Wirkungen auf die Lehrpersonen und die Schule bzw. die Gemeinschaft in der Schule festgestellt werden, wobei diese fast durchgehend positiv ausgefallen sind.

Eine in der Schweiz durchgeführte Metaanalyse von kantonalen, evaluierten Schulreformprojekten weist nach, dass aufgrund struktureller Reformen zwar Auswirkungen in den Bereichen „Teamentwicklung“ und „Organisationskultur“, ansatzweise auch im Bereich „Unterricht“ ausgemacht werden können. Die Unterlagen der unter-

suchten Reformprojekte lassen aufgrund ihrer Evaluationsanlagen hingegen ebenfalls keine Aussagen in Bezug auf die Auswirkungen der erhöhten Gestaltungsfreiheit auf die Leistungen der Schüler/innen zu (Maag Merki & Büeler, 2002). Auch die im Kanton Luzern durchgeführte Studie zur Wirksamkeit der Reformstrategie „Schulen mit Profil“ erweist sich in verschiedensten Bereichen als effektiv, so beispielsweise in Bezug auf Organisations- und Teamentwicklung. Hingegen zeigen sich keine Einflüsse auf den Unterricht. Nicht untersucht wurden wiederum die Effekte der Schulreform auf die Leistungen der Schüler/innen (Büeler, Buholzer & Roos, 2005).

In einer Studie im Kanton Zürich zur Erfassung von fachspezifischen Leistungen der Schüler/innen der 5. Klasse wurde diese Auswertungsstrategie hingegen ansatzweise berücksichtigt, indem die Leistungen der Schüler/innen in Abhängigkeit von der Schulorganisationsform („Teilautonome Volksschulen“ vs. „Traditionelle Schulen“) analysiert worden sind. Dabei konnte festgestellt werden, dass die alleinige strukturelle Reform („TaV“) keine Auswirkungen auf die fachlichen Leistungen der Schüler/innen hat (Moser, Keller & Tresch, 2002).

In der PISA-Studie 2000 hingegen ergaben sich schwache positive Zusammenhänge zwischen verschiedenen Aspekten schulischer Autonomie und den Schüler/innenleistungen (vgl. OECD, 2001, S. 208), wenn auch die empirischen Ergebnisse keineswegs widerspruchsfrei sind (Fend, 2004b). Auch die Ergebnisse von PISA 2003 zeigen im internationalen Vergleich wiederum schwache positive Effekte (OECD, 2004). In der Schweiz erweist sich dieser Effekt allerdings konfundiert mit sprachregionalen Aspekten. „Der scheinbar positive Effekt der Schulautonomie auf die Mathematikleistung fällt zusammen mit dem sprachregional unterschiedlichen Schuleintrittsalter und erweist sich innerhalb der Schweiz als sprachregionaler Effekt“ (Moser & Berweger, 2004, S. 56).

Effekte auf die fachlichen Kompetenzen der Schüler/innen lassen sich hingegen bei Cheung & Cheng (2002) ausmachen. Die Autoren haben in ihren kontrastiven Fallanalysen von drei Primarschulen in Hong Kong den Zusammenhang zwischen „multilevel self-management“ und Schulleistungen untersucht, wobei diese einerseits über Leistungstests in Muttersprache, Englisch und Mathematik, andererseits über schulische Qualitätsindikatoren aus der Perspektive der Lehrpersonen und über Arbeitszufriedenheit der Lehrpersonen erfasst worden sind. Die Ergebnisse zeigen, dass sich die über eine Extremgruppenauswahl bestimmten Schulen hinsichtlich verschiedenster der erfassten Output-Kriterien unterscheiden. Bei den schulischen Leistungen unterscheidet sich insbesondere jene Schule mit dem am höchsten eingeschätzten „multilevel self-management“ von derjenigen Schule, die auf den drei erfassten Levels des „Self-Managements“ tiefe Werte erzielten. Allerdings zeigt die Studie auch, dass die Schüler/innen in Schulen, die nur in einzelnen Bereichen ein hohes Self-Management aufweisen, vergleichbar gute Leistungen erbringen wie die Schüler/innen in Schulen mit einem in allen Bereichen hohen Self-Management.

Australische Studien weisen auf die Bedeutung indirekter Effekte hin. So weisen Silins & Mulford (2004) nach, dass Schulen als Lernende Organisationen insbesondere die Zufriedenheit der Lehrpersonen und die Professionalität im Team positiv beeinflussen, welche wiederum einen positiven Einfluss auf die Partizipation und das Engagement der Schüler/innen im Unterricht haben. Lernende Schulen üben damit einen indirekten Effekt auf wichtige Outputdimensionen der Schüler/innen aus. „However, teacher leadership itself does not contribute to enhancing student's positive perception of teachers' work or their participation in or engagement with

Es lassen sich keine oder lediglich schwache positive Zusammenhänge zwischen strukturellen Reformen und Leistungen der Schüler/innen feststellen.

school“ (ebd., S. 461). Ähnliche Ergebnisse zeigt eine Metaanalyse dreier Studien in Australien, Neuseeland und England (Dempster, 2000). „Does school-based management lead to improve student learning outcomes? From the research results reported above, the short answer to this question is, no. However, there are indicative data which imply that aspects of school-based management related to planning and communication help shape some of the conditions which indirectly influence classroom practice“ (ebd., S. 56).

In den Analysen von TIMSS- und PISA-Daten, die Wössmann (2003, 2005) durchgeführt hat, ergeben sich wichtige Hinweise darauf, dass sich schulische Autonomie erst in Kombination mit einem output-orientierten Monitoringssystem – in diesem Fall zentrale Abschlussprüfungen – positiv in Bezug auf die Schüler/innenleistungen auswirkt, wobei sich diese Effekte hinsichtlich verschiedenster Autonomieaspekten nicht gleichermassen herausstellt. Es kann sogar sein, dass schulische Autonomie einen negativen Effekt auf die Schüler/innenleistungen aufweist, wenn das Konzept der erhöhten Gestaltungsmöglichkeiten nicht ergänzt wird mit zentralen Abschlussprüfungen. „Central exams help otherwise decentralised school systems to benefit from local knowledge leads without suffering from local opportunistic behaviour“ (Wössmann, 2005, S. 54).

Es fehlen fundierte Kenntnisse darüber, wie sich strukturell-organisatorische Reformen auf die Leistungen der Schüler/innen auswirken.

Sprechen diese Ergebnisse nun gegen die Absicht, den Schulen mehr Gestaltungsfreiheiten und mehr Verantwortung für den eigenen Qualitätssicherungsprozess zu übergeben? Aufgrund dieses kurzen Überblicks muss zuerst einmal konstatiert werden, dass sich aus empirischer Perspektive die Befundlage heterogen und in dem Sinn defizitär präsentiert, als Wirkungen organisatorisch-struktureller Reformen auf die fachlichen und überfachlichen Leistungen der Schüler/innen nur sehr beschränkt analysiert worden sind und fundierte Erkenntnisse darüber fehlen, inwiefern sich Veränderungen in der Organisation und Struktur der einzelnen Schule auf die Entwicklung von fachlichen und überfachlichen Kompetenzen der Schüler/innen auswirken. Einzelne neuere Studien untersuchen zwar den Zusammenhang zwischen strukturell-organisatorischen Reformen und Schüler/innenleistungen, so beispielsweise die Studie in Nordrhein-Westfalen (Projektdauer 2003-2008), in welcher das für die Volksschule (bis 9. Klasse) entwickelte Modellvorhaben „Selbständige Schulen“ aus unterschiedlichen Perspektiven mittels einer Längsschnittstudie evaluiert wird. Differenzierte Ergebnisse sind allerdings erst in den nächsten Jahren zu erwarten.

Nicht nur in der Schulentwicklungsforschung, sondern auch in der Schuleffektivitätsforschung kann allerdings ein zu Engführen der einzelnen Forschungstraditionen beobachtet werden (Teddlie & Reynolds, 2001). In den wenigsten Effektivitätsstudien, die zum Ziel haben, die Wirkungen von Schulen zu erfassen, werden die durchgeführten Leistungs- und Kompetenzmessungen systematisch entlang von schulspezifischen Prozessmerkmalen auf der Mesoebene (z.B. Organisationsklima, Feedbackkultur) analysiert. Bei Leistungsvergleichsstudien wie PISA ist das methodische Design (Stichprobenplan) zudem nicht auf einen Vergleich von Schulen und ihrer Funktionalität für die Förderung von Schüler/innenkompetenzen angelegt, da nicht systematisch ganze Klassen und die die Klassen unterrichtenden Lehrpersonen sowie Lehrer/innenteams untersucht worden sind. Weiter zeigt sich, trotz der Bedeutung von nicht-fachbezogenen Kompetenzen für das Lernen der Schüler/innen (Grob & Maag Merki, 2001), dass viele der durchgeführten Effektivitätsstudien einzig Fachleistungen im Bereich Mathematik, Naturwissenschaften und Sprache einbeziehen (Campbell, Kyriakides, Muijs & Robinson, 2003; Creemers & Kyriakides, 2005). In jüngster Zeit wird dieses Defizit insofern berücksichtigt, als bei Leistungsvergleichsstudien zunehmend

schulische Kontextfaktoren multiperspektivisch und die Wirkungen multidimensional erfasst werden. Die Ergebnisse dieser Studien werden in nächster Zeit publiziert, so dass zu den hier interessierenden Fragestellungen neuere Erkenntnisse erwartet werden können. Dies ist beispielsweise der Fall bei der internationalen Desi-Studie, in welcher die fachlichen Leistungen (Deutsch, Englisch) von Schülerinnen und Schülern u.a. in Bezug zu schulisch-organisatorischen Bedingungsfaktoren gesetzt werden. In dieser Studie werden die komplexen Bedingungsfaktoren der sprachlichen Leistungen von Schülerinnen und Schülern am Ende der obligatorischen Schulzeit mit Erhebungsinstrumenten für Schulleiter/innen, Lehrpersonen, Schüler/innen und Eltern untersucht, um den individuellen, familiären, unterrichtlichen und schulischen Hintergrund der Schülerleistungen zu erfassen (vgl. <http://www.dipf.de/desi/>). Bei einer Längsschnittstudie in den Gymnasien des Kantons Zürich wiederum werden die überfachlichen Kompetenzen der Schüler/innen in Abhängigkeit einer breiten Palette an schulischen Prozessfaktoren sowohl aus der Perspektive der Lehrpersonen wie auch aus der Perspektive der Schüler/innen erfasst (Maag Merki, 2005; Maag Merki, im Druck).

Auf der Basis der bisherigen Studien scheint es, dass organisatorisch-strukturelle Veränderungen im Sinn einer Teilautonomisierung in Bezug auf die Förderung der Leistungen der Schüler/innen für sich allein nur beschränkt wirkungsmächtig sind. Aufgrund verschiedener Befunde kann aber trotzdem davon ausgegangen werden, dass die zumindest indirekte Wirksamkeit dieser Strukturreform eine hohe Evidenz aufweist, insbesondere dann, wenn die Qualität der Umsetzung der Teilautonomie mitberücksichtigt wird, wenn der strukturellen eine didaktische Reform folgt (Maag Merki & Roos, 2005) oder wenn das Konzept der Teilautonomie kombiniert wird mit einem outputorientierten Bildungsmonitoring (Wössmann 2005).

So kann Organisationsentwicklung nach Rolff (1998) neben Personal- und Unterrichtsentwicklung einzig als *ein* Teil eines umfassenden Schulentwicklungsansatzes verstanden werden, wobei sich die verschiedenen Bereiche systematisch aufeinander zu beziehen haben. Schulentwicklungsmassnahmen werden von den beteiligten Lehrpersonen sodann auch erst dann als gut investierte Zeit mit einer hohen Effektivität betrachtet, wenn sich diese Vorhaben systematisch auf ihr Kerngeschäft, den Unterricht beziehen (Maag Merki, Imhasly & Leutwyler, 2005). Zudem zeigen Analysen von Schulen, die eine erhöhte Gestaltungsfreiheit zugewiesen erhalten haben, dass die Umsetzungsqualität dieser Gestaltungsfreiheit keinesfalls homogen ist. Vielmehr lassen sich zwischen den Schulen grosse Unterschiede in Bezug auf ihre prozessuale Gestaltung (Kuper, 2002; Pang, 2005) oder ihre Problemlösekompetenz erkennen (Cheung & Cheng, 2002; Steinert, Klieme, Schweizer & Döbrich, 2005), wobei sich diese auch auf wichtigen schulischen Dimensionen abbilden lassen (Maag Merki, Halbheer & Kunz, 2005).

In der internationalen Vergleichsstudie in Deutschland wird aus den vertiefenden Analysen aus PISA 2000 abgeleitet, dass die Entwicklung schulischer Leistungen wesentlich von einer durch die Einzelschule geprägten und beeinflussten Gestaltung von schulischen Rahmenbedingungen und von pädagogischen Prozessen abhängt (vgl. Arbeitsgruppe ‚Internationale Vergleichsstudie‘, 2005, S. 210). Zu ähnlichem Ergebnis kommen auch Larcher & Oelkers (2005) im Kontext der schweizerischen PISA-2000-Vertiefungsstudien, indem sie auf die hohe Bedeutung von starken Supportsystemen, Lernumgebungen mit hoher Qualität oder präventiven Massnahmen hinsichtlich Integration bzw. Segregation hinweisen. Fullan & Watson (2000) verweisen ebenfalls auf eine indirekte Wirksamkeit dieser Art von Strukturreformen, da

Es kann davon ausgegangen werden, dass Strukturreformen eine indirekte Wirksamkeit besitzen.

Organisationsentwicklung kann nur als *ein* Teil eines umfassenden Schulentwicklungsansatzes gesehen werden.

diese einen wirkungsvollen und notwendigen Rahmen darstellen, um professionelle Lehr- und Lernformen zu etablieren und dadurch die Unterrichtsqualität zu verbessern. Shachar & Sharan (1993) betonen, dass zwischen der Unterrichtsform und der Organisationsform eine Interdependenz besteht und die Implementierung von neuen Lehr- und Lernmethoden somit erst dann effektiv realisiert werden kann, wenn bei der Schulorganisation entsprechende Rahmenbedingungen nach dem Modell der „Organisation offener Systeme“ bestehen.

Organisationssoziologische Analysen geben zudem Hinweise darauf, warum die Umsetzung eines Konzeptes wie jenes der „Lernenden Organisation“ nicht linear entsprechend den vorgegebenen Eckpunkten erfolgreich verlaufen kann bzw. muss. So lässt sich auf der Basis des soziologischen Neo-Institutionalismus aufzeigen (vgl. Schaefers, 2002), dass der Umgang von Organisationen mit Anforderungen aus der Umwelt vermehrt geprägt sein kann durch den Wunsch nach Legitimitätssicherung und weniger durch den Wunsch nach Erhöhung der Effektivität. Als Folge davon entwickelt sich eine „pro forma compliance“ (Mitchell, 1996, S. 175). Zudem hängt der Erfolg von Reformprozessen von den organisationsintern etablierten Deutungs-, Problemlösungs- und Handlungsmustern ab. Insbesondere in sogenannten schwachen (Schwier 2005) oder belasteten Schulen (Fend 1998) erweisen sich diese als besonders resistent gegenüber Veränderungen.

Schulentwicklungsdiskussionen unterstellen den beteiligten Personen ein gemeinsames Interesse.

Schönig und Fees wiederum zeigen, dass die gegenwärtige Schulentwicklungsdiskussion den am Entwicklungsprozess der Einzelschule beteiligten Personen einen gemeinsamen Interessepfad unterstellt. Dadurch werden aber Interessen Einzelner oder Gruppen und deren spezifische Abwehrkonstellationen, welche einer gemeinsamen Zielerreichung zuwiderlaufen können, nicht hinreichend berücksichtigt (vgl. Fees, 2004; vgl. Schönig, 2002, S. 817). Die interaktiven Prozesse zwischen den verschiedensten Akteursgruppen innerhalb einer Schule und den beteiligten Personen erscheinen sodann als potenzielle Stolpersteine, aber auch als Gelingensfaktoren, die den Umsetzungsprozess zentral beeinflussen, was ein sorgfältiges Planen und Umsetzen der Teilautonomie in der einzelnen Schule erforderlich macht.

In welchen Bereichen zeichnen sich Forschungslücken ab?

Aufgrund der Erkenntnisse bisheriger Schulqualitäts- und Schuleffektivitätsstudien wird zwar davon ausgegangen, dass sich Lernprozesse und Kompetenzentwicklung bei den Schülerinnen und Schülern über die Gestaltung makropolitischer Strukturen nicht zielgenau herstellen lassen. Bei der aktuellen Fokussierung von Lernenden Schulen im Rahmen der neuen Output-Steuerungsmodelle muss allerdings aufgrund des heutigen Forschungsstandes gefragt werden, ob dieses Technologisierungsdefizit nicht zu „blind“ an die Akteure der Mesoebene delegiert wird, welche den Auftrag erhalten, über die Konzipierung einer lernförderlichen Umgebung in der Einzelschule die gesellschaftlich formulierten Ziele bei ihren Schülerinnen und Schülern zu erreichen, die wiederum über Output-Steuerungsmechanismen regelmässig und systematisch überprüft werden sollen. Denn: Wie und ob die Schulen die Vorgaben erreichen können, ist noch nicht hinreichend bekannt. Dem analytischen Aufeinanderbeziehen von Organisation und Kompetenz der einzelnen Schule muss somit aus wissenschaftlicher und bildungspolitischer Perspektive in Zukunft verstärktes Gewicht zukommen. Weiterführende Analysen sind notwendig, um zu verstehen, wie Schulen funktionieren, wie sie sich entwickeln und wie sie die Lehr- und Lernprozesse der Schüler/innen möglichst optimal beeinflussen können (Klieme, 2003). Dies

Organisation und Kompetenz der einzelnen Schule müssen verstärkt analytisch aufeinander bezogen werden.

ist insbesondere auch aus Sicht der Schulpraxis wichtig, da die Implementierung von Schulen mit erweiterten Gestaltungsmöglichkeiten eine für alle Beteiligten aufwändige und kostspielige Angelegenheit ist.

Insbesondere muss die Wirksamkeit der „Lernenden Organisation“ in einer differenzierteren Art und Weise erforscht werden, als dies bis anhin gemacht worden ist. So ist nicht einzig der – letztlich makropolitische – Umstand, ob Teilautonomie gewährt wird oder nicht, für die Analyse der Effektivität dieses Konzeptes relevant, sondern vor allem der Professionalisierungsgrad der einzelnen „Lernenden Schulen“. Dies bedingt eine Unterscheidung zwischen unterschiedlichen Kompetenzniveaus in Bezug auf die Implementierung einer „Lernenden Schule“, wie dies beispielsweise Steinert & Klieme (2005) oder Cheung & Cheng (2002) aus empirischer Perspektive gemacht haben. Steinert & Klieme unterscheiden fünf verschiedene Kooperationslevels, die die Dichte der Kooperationen zwischen Lehrpersonen in unterschiedlichen Themenbereichen beschreiben, währenddem Cheung & Cheng verschiedene Ausprägungsgrade im „multilevel self-management“ unterscheiden.

Die Überprüfung der Funktionalität des Konzeptes sollte zudem in Abhängigkeit eines breiten Sets an Wirkungsdimensionen aus unterschiedlichen Perspektiven (Schulleitungen, Lehrpersonen, Schüler/innen, Eltern, Schulbehörde) erforscht werden. Dazu sind Längsschnittstudien für die Analyse schulischer Prozesse und der Kompetenzentwicklung der Schüler/innen notwendig, wobei die Kompetenzen breit zu fassen sind und sich nicht nur auf fachspezifische Aspekte beschränken sollten. Insbesondere erlaubt ein dynamisches Analysemodell, wie dies Creemers und Kyriakides (2005) vorschlagen, die Beschreibung der komplexen Struktur von Prozess und Wirkung unter Berücksichtigung der unterschiedlichen Analyselevels. Dabei muss in Betracht gezogen werden, dass die Beziehungen zwischen Prozess- und Wirkungsfaktoren nicht linear, sondern curvilinear sein können. Dies impliziert die Suche nach der „optimalen Zone“ schulischen Handelns in Abhängigkeit spezifischer Wirkungsfaktoren. Erst diese umfassenden Analysen erlauben ein annähernd differenziertes Bild über Funktions- und Wirkungsweisen von „Lernenden Schulen“.

Die Analysen von Wössmann (2005) zeigen zudem auf, dass Effekte unter einer differenziellen Perspektive zu untersuchen sind. Insbesondere bleibt zu analysieren, in welchem Ausmass sich Effekte bei unterschiedlichen Gruppen von Schülerinnen und Schülern etablieren, so beispielsweise aufgrund ihres sozio-demographischen Hintergrundes, aufgrund ihrer Leistungsfähigkeit oder aufgrund spezifischer Persönlichkeitsmerkmale.

Der Professionalisierungsgrad der Einzelschule ist bei der Analyse zu berücksichtigen.

Die Funktionalität des Konzeptes sollte mit einem dynamischen Analysemodell überprüft werden.

Ziel ist es, die „optimale Zone“ schulischen Handelns unter differenzieller Perspektive zu finden.



Prof. Dr. Katharina Maag Merki
Johann Wolfgang Goethe-Universität
Institut für Allgemeine Erziehungswissenschaft
Robert-Mayer-Str. 1
D-60054 Frankfurt am Main

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Schloßstraße 29
D-60486 Frankfurt am Main

K.Maag.Merki@em.uni-frankfurt.de

Literatur

- Arbeitsgruppe ‚Internationale Vergleichsstudie‘ (2005). *Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten*. Berlin, Frankfurt a.M.: Bundesministerium für Bildung und Forschung, Deutsches Institut für Internationale Pädagogische Forschung.
- Argyris, C. & Schön, D.A. (1999). *Die lernende Organisation. Grundlagen, Methode, Praxis*. Stuttgart: Klett-Cotta.
- Büeler, X., Buholzer, A. & Roos, M. (2005). *Schulen mit Profil. Forschungsergebnisse - Brennpunkte - Zukunftsperspektiven*. Innsbruck: Studien Verlag.
- Burkard, C. & Eikenbusch, G. (2002). Schulentwicklung international - eine Bilanz. Was können wir von den Erfahrungen anderer Länder lernen? *Pädagogik*, 54(11), 44-49.
- Campbell, R.J., Kyriakides, L., Muijs, R.D. & Robinson, W. (2005). Differential teacher effectiveness: towards a model for research and teacher appraisal. *Oxford Review of Education*, 29(3), 547-562.
- Cheung, F.W.M. & Cheng, Y.C. (2002). An Outlier Study of Multilevel Self-Management and School Performance. *School Effectiveness and School Improvement*, 13(3), 253-290.
- Creemers, B.P.M. & Kyriakides, L. (2005). Establishing links between educational effectiveness research and improvement practices through the development of a dynamic model of educational effectiveness. Paper presented at the AERA Conference in Montreal, April 2005.
- Dalin, P. (1986). *Organisationsentwicklung als Beitrag zur Schulentwicklung. Innovationsstrategien für die Schule*. Paderborn: Ferdinand Schöningh.
- Dalin, P., Rolff, H.-G. & Buchen, H. (1996). *Institutioneller Schulentwicklungsprozess. Ein Handbuch*. Soest: Verlag für Schule und Weiterbildung.
- Dempster, N. (2000). Guilty or not: the impact and effects of site-based management on schools. *Journal of Educational Administration*, 38(1), 47-65.
- Ekhholm, M. (1997). Steuerungsmodelle für Schulen in Europa. Schwedische Erfahrungen mit alternativen Ordnungsmodellen. *Zeitschrift für Pädagogik*, 43(4), 597-610.
- Fees, K. (2004). Schule als „Lernende Organisation“. Zur Problematik eines Theorieimportes. *Die Deutsche Schule*, 96(1), 10-22.
- Fend, H. (1986). „Gute Schulen – Schlechte Schulen“. Die einzelne Schule als pädagogische Handlungseinheit. *Deutsche Schule*(3), 275-295.
- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fend, H. (2004a). *Gestaltung des Bildungswesens zwischen Euphorie und Resignation. Auf dem Weg zu konsistenten Steuerungskonzepten? Referat gehalten am Symposium „Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert“ an der Universität Zürich, 24. September 2004*.
- Fend, H. (2004b). Was stimmt mit den deutschen Bildungssystemen nicht? Wege zur Erklärung von Leistungsunterschiede zwischen Bildungssystemen. In: G. Schümer, K.-J. Tillmann & M. Weiss (Hrsg.), *Die Institution Schule und die Lebenswelt der Schüler* (S. 15-38). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fullan, M. (1999). *Die Schule als lernendes Unternehmen. Konzepte für eine neue Kultur in der Pädagogik*. Stuttgart: Klett-Cotta.
- Fullan, M. & Watson, N. (2000). School-Based Management: Reconceptualizing to Improve Learning Outcomes. *School Effectiveness and School Improvement*, 11(4), 455-473.
- Grob, U. & Maag Merki, K. (2001). *Überfachliche Kompetenzen. Theoretische Grundlegung und empirische Erprobung eines Indikatorensystems*. Bern: Peter Lang.
- Helsper, W., Böhme, J., Kramer, R.-T. & Lingkost, A. (1997). Schulkultur und Schulmythos – Skizzen und Thesen zur Entwicklung gymnasialer Schulkultur in den Neuen Bundesländern. In: H.-H. Krüger & J.-H. Olbertz (Hrsg.), *Bildung zwischen Staat und Markt* (S. 583-595). Opladen: Leske+Budrich.
- Helsper, W., Böhme, J., Kramer, R.-T. & Lingkost, A. (2001). *Schulkultur und Schulmythos. Rekonstruktionen zur Schulkultur I*. Opladen: Leske+Budrich.
- Klieme, E. (2005). *Zur Rolle der empirischen Forschung bei der Qualitätsentwicklung im Bildungswesen: Theoriearbeit, kritische Instanz oder Service für Politik und Öffentlichkeit? Referat gehalten an der Frühjahrstagung der Sektion „Empirische Bildungsforschung“ der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE), 17. März 2003*. Frankfurt am Main.
- Krainz-Dürr, M. (1999). *Wie kommt Lernen in die Schule? Zur Lernfähigkeit der Schule als Organisation*. Innsbruck: Studien Verlag.

- Kuper, H. (2002). Entscheidungsstrukturen in Schulen. Eine differentielle Analyse der Schulorganisation. *Zeitschrift für Pädagogik*, 48(6), 856-878.
- Larcher, S. & Oelkers, J. (Hrsg.). (2005). *Die besten Ausbildungssysteme. Thematischer Bericht der Erhebung PISA 2000*. Neuchâtel: BFS/EDK.
- Leithwood, K. & Menzies, T. (1998). Forms and effects of school-based management: A review. *Educational Policy*, 12(3), 525-546.
- Maag Merki, K. (2005). Die Qualität eines kantonalen Bildungssystems – erste Ergebnisse einer externen Evaluation. In: K. Maag Merki & P. Schuler (Hrsg.), *Überfachliche Kompetenzen. Berichte aus dem Forschungsbereich Schulqualität & Schulentwicklung* (S. 56-79). Zürich: Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.
- Maag Merki, K. (Hrsg.) (im Druck). *Lernort Gymnasium. Individuelle Entwicklungsverläufe und Schulerfahrungen*. Bern: Haupt Verlag.
- Maag Merki, K. & Büeler, X. (2002). Schulautonomie in der Schweiz. Eine Bilanz auf empirischer Basis. In: H.-G. Rolff, H.G. Holtappels, K. Klemm, H. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung, Band 12. Daten, Beispiele und Perspektiven* (S. 131-161). Weinheim, München: Juventa.
- Maag Merki, K., Halbheer, U. & Kunz, A. (2005). „Pädagogische Entwicklungsbilanzen (PEB)“ - Profile von Zürcher Mittelschulen. Zürich: Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.
- Maag Merki, K., Imhasly, M.-T. & Leutwyler, B. (2005). „Spieglein, Spieglein an der Wand, wer ist die Schönste im ganzen Land?“ oder Fördert Selbstevaluation die Qualität der Schule? *Journal für Schulentwicklung*, 2, 33-41.
- Maag Merki, K. & Roos, M. (2005). Erfolgsfaktoren auf dem Weg zur fördernden Schule. Erfolgreiche Change-Prozesse brauchen strukturelle und didaktische Reformen. In: H. Buchen, L. Horster & H.-G. Rolff (Hrsg.), *Schulleitung und Schulentwicklung. Erfahrungen, Konzepte, Strategien* (S. B 2.7 (1-18)). Stuttgart: RAABE.
- Mitchell, D.E. (1996). Institutional Theory and the Social Structure of Education. In: R.L. Crowson, W.L. Boyd & H.B. Mawhinney (Hrsg.), *The Politics of Education and the New Institutionalism: Reinventing the American School. The 1995 Yearbook of the Politics of Education Association* (S. 176-188). Washington, DC: Falmer.
- Moser, U. & Berweger, S. (2004). Einflüsse des Bildungssystems und der Schulen auf die Mathematikleistung. In: Bundesamt für Statistik (BFS) & Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (Hrsg.), *PISA 2003: Kompetenzen für die Zukunft. Erster nationaler Bericht* (S. 45-60). Neuchâtel, Bern: BFS/EDK.
- Moser, U., Keller, F. & Tresch, S. (2002). *Evaluation der 3. Primarschulklassen. Schlussbericht zuhanden der Bildungsdirektion des Kantons Zürich*. Zürich: KBL, Universität Zürich.
- Munin, H. (2001). *Schulautonomie. Diskurse, Massnahmen und Effekte im internationalen Vergleich, insbesondere in Deutschland*. Weinheim: Beltz.
- OECD (2001). *Lernen fürs Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000*. Paris: OECD.
- OECD (2004). *Learning for Tomorrow's World - First Results from PISA 2003*. Paris: OECD.
- Pang, N.S.-K. (2005). Binding Forces and TeachersSchool Life: A Recursive Model. *School Effectiveness and School Improvement*, 14(3), 293-520.
- Rolff, H.-G. (1995). *Wandel durch Selbstorganisation. Theoretische Grundlagen und praktische Hinweise für eine bessere Schule*. Weinheim, München: Juventa.
- Rolff, H.-G. (1998). Entwicklung von Einzelschulen: Viel Praxis, wenig Theorie und kaum Forschung - Ein Versuch, Schulentwicklung zu systematisieren. In: H.-G. Rolff, K.-O. Bauer, K. Klemm & H. Pfeiffer (Hrsg.), *Jahrbuch der Schulentwicklung, Band 10. Daten, Beispiele und Perspektiven* (S. 295-526). Weinheim, München: Juventa.
- Schaefers, C. (2002). Der Neo-Institutionalismus. Eine organisationstheoretische Analyse- und Forschungsperspektive auf schulische Organisationen. *Zeitschrift für Pädagogik*, 48(6), 855-855.
- Schönig, W. (2002). Organisationskultur der Schule als Schlüsselkonzept der Schulentwicklung. *Zeitschrift für Pädagogik*, 48(6), 815-854.
- Schwier, B. (2005). Ein blinder Fleck: Schulentwicklung als Schulverbesserung von „schwachen“ Schulen in England und Wales. Probleme und Entwicklungsansätze. *Zeitschrift für Pädagogik*, 51(3), 580-596.
- Senge, P.M. (1996). *Die fünfte Disziplin*. Stuttgart: Klett-Cotta.
- Shachar, H. & Sharan, S. (1995). Schulorganisation und kooperatives Lernen im Klassenzimmer: Eine Interdependenz. In: G.L. Huber (Hrsg.), *Neue Perspektiven der Kooperation: ausgewählte Beiträge der Internationalen Konferenz 1992 über Kooperatives Lernen* (S. 54-70). Baltmannsweiler: Schneider-Verlag Hohengehren.

- Silins, H. & Mulford, B. (2004). Schools as Learning Organisations - Effects on Teacher Leadership and Student Outcomes. *School Effectiveness and School Improvement*, 15(3-4), 443-466.
- Steinert, B., Klieme, E., Schweizer, K. & Döbrich, P. (2005). *Stufen der Lehrerverbesserung: Ein kriteriumsorientierter Ansatz zur Evaluation von Schulentwicklung. Referat gehalten an der Frühjahrstagung der Sektion „Empirische Bildungsforschung“ der Deutschen Gesellschaft für Erziehungswissenschaften (DGfE). 19. März 2003.* Unpublished manuscript.
- Teddlie, C. & Reynolds, D. (Hrsg.). (2001). *The International Handbook of School Effectiveness Research*. London: Routledge Falmer.
- Wössmann, L. (2005). Central Exams as the „Currency“ of School Systems: International Evidence on the Complementarity of School Autonomy and Central Exams. *DICE Report - Journal for Institutional Comparisons*, 1(4), 46-56.
- Wössmann, L. (2005). The Effect Heterogeneity of Central Examinations: Evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, 13(2), 143-169.

Systemsteuerung im Bildungswesen – Anschlussfähigkeiten an die Schulwirklichkeit

Helmut Fend

1. Erwartungen an die Systemsteuerung des Bildungswesens

Wir leben, was die Bildungspolitik betrifft, in staunenswürdigen Zeiten. In wenigen Jahren scheint von Seiten der Bildungspolitik in ungeahnter Geschwindigkeit alles übernommen zu werden, was die empirische Bildungsforschung im Stillen, teils unbeachtet, teils angefochten, in den letzten dreissig Jahren entwickelt hat. Nach Zeiten des Stillstandes ist eine Reformbereitschaft zu beobachten, die das Bildungswesen in rascher Zeit modernisieren könnte. Dies gilt z.B. für die Bereitschaft, sich an international vergleichenden Leistungsmessungen zu beteiligen, aber auch für die Entwicklungsarbeit, die sich an Standards der Zielerreichung orientiert. Ein weiteres Beispiel ist die Inhaltsreform mit der Einführung des Frühenglisch und der Neugestaltung des mathematisch-naturwissenschaftlichen Unterrichts. Ein drittes Kernelement repräsentiert die Bereitschaft zu grundlegenden Verwaltungsreformen, die auf eine grössere Autonomie der Einzelschule und auf die Stärkung dieser Handlungsebene zielen.

Diese neuen Steuerungsinstrumente stehen jedoch nicht isoliert, sie werden vielmehr in einem Zusammenhang gesehen, der wie folgt artikuliert wird: Auf bildungspolitischer Ebene wird ein nationaler Konsens über die *Kompetenzen und Standards* angestrebt, die ein Bildungswesen in seinen verschiedenen Teilsektoren bei Kindern und Jugendlichen erreichen soll. Sie operationalisieren theoriegeleitet das kulturelle Programm der Schule. Die damit verknüpften Erwartungen werden in neu gegründeten Schulinstitutionen operationalisiert und in *Testaufgaben* überführt. Diese Aufgaben werden in zu vereinbarenden Abständen in den Schulklassen eingesetzt und über die moderne Infrastruktur der Informationsverarbeitung bis in die Schulklassen hinein zurückgemeldet. Schulzweige, Schulen, Lehrerinnen und Lehrer, Schülerinnen und Schüler können so ihre Leistungen in einem nationalen und internationalen Bezugsrahmen einschätzen.

Dieser Evaluation auf Systemebene entspricht eine grössere *Autonomie der Einzelschule*. Sie wird aus der dichten verwaltungsorientierten Aufsicht entlassen und in ihren Handlungsmöglichkeiten gestärkt.

Die klassische Systemsteuerung manifestierte sich traditionell in der Polarität von Einzellehrkraft und politisch-administrativer Kontrolle. Diese Polarität beginnt sich sukzessive aufzulösen. Dabei wird die intermediäre Ebene der Schule entdeckt und zu einem zentralen Ort der Entwicklungsarbeit. Die Entwicklung einer grossen Zahl von Instrumenten zielt darauf, die pädagogische Qualität der Einzelschule zu stärken und sie zu einem teilautonomen Ort führungsintensiver pädagogischer Gestaltung zu ma-

Nach Zeiten des Stillstands ist in der Bildungspolitik eine grosse Reformbereitschaft.

Systemebene:
Nationaler Konsens über Kompetenzen und Standards.

Mesebene:
Die Einzelschule wird gestärkt.

chen. *Selbstevaluation* und *Fremdevaluation* sowie Schulentwicklungsarbeit sollen Hand in Hand die Unternehmen „Schule“ qualitativ stärken. Die neue Rechenschaftslegung führt dazu, dass sie ihre neue Autonomie pädagogisch verantwortungsvoll nutzen.

**Mikroebene:
Lehrerinnen und Lehrer bauen
ihre fachdidaktischen
Kompetenzen aus.**

Die Lehrerinnen und Lehrer als operative Akteure sind in diesem Rahmen Ort der permanenten Personalentwicklung. Sie erarbeiten über neue Medien und gegenseitige Unterrichtsbeobachtungen insbesondere ihre *fachdidaktischen Fähigkeiten* – den wohl wichtigsten neuen Schwerpunkt der Professionalisierung. Ihre Kompetenzen als Fachleute für die Lehre werden von der Fachaufsicht, den Kollegen und Kolleginnen, der Elternschaft und Schülerschaft regelmässig evaluiert und über Zielvereinbarungen weiter entwickelt.

Die Verfahren der Standardsicherung sind so begleitet von permanenten Entwicklungsarbeiten auf allen Ebenen des Bildungswesens. Die Makrosteuerung entwickelt die Standards, die Lehrmittel und die dazugehörigen Testaufgaben weiter und die Schulen optimieren ihre pädagogische Qualität über Verfahren der Selbstevaluationen, der Schulprogrammentwicklung und der externen Evaluation. Über Leistungsvereinbarungen mit den Lehrpersonen und Schulen wird die für notwendig erachtete Schulentwicklung eingeleitet und die Kooperation gesichert.

**Die neue Systemsteuerung
ist am Output orientiert.**

War die alte Systemsteuerung am Rechtssystem orientiert, also an der Schaffung von Recht und an der Überprüfung der Praxis auf ihre Rechtskonformität, so ist die neue auf Ergebnisse ausgerichtet, auf die Schaffung bestmöglicher Bedingungen zur Erfüllung von Aufträgen. Sie sieht sich selbst als Input und richtet diesen an einem bestmöglichen Output aus, wie er im *New Public Management* (Schedler, 1995) konzipiert ist. Die Effizienzerwartungen der Wirtschaft stehen hier eher Pate als die Rechtsförderungsstandards der Justiz und Politik.

Überfällige Modernisierung?

Viele der neuen Massnahmen haben eine hohe Plausibilität.

**Schaffung von mehr Transparenz
in den Leistungen des
Bildungswesens**

1. Die *Schaffung von mehr Transparenz* in den Leistungen des Bildungswesens ist überfällig. Die neuen Evaluationsverfahren erzeugen Wissen über das System, das die rein internen Rückmeldemechanismen durch die Schulaufsicht ergänzt. Die neuen Verfahren der Systemevaluation setzen eine Festigung des Zielwissens über die Entwicklung von Standards voraus, sie erfordern die Umsetzung in verschiedene Anspruchsniveaus und schliesslich die Operationalisierung durch Prüfungsaufgaben und Tests.

Stärkung der Autonomie

2. Auch die *Stärkung der Autonomie* und Professionalität der einzelnen Schule, die Stärkung der Schule als professionelle Handlungseinheit, ist plausibel angesichts der Heterogenität der Probleme vor Ort, die weder aus der Ferne noch individuell und einzelkämpferisch bewältigt werden können.

**Revision des schulischen
Inhaltsprogramms**

3. Die *Revision des schulischen Inhaltsprogramms*, die Stärkung der Muttersprache, der Naturwissenschaften und der fremdsprachlichen Kompetenzen macht ebenso Sinn wie das Überdenken der lebensgeschichtlichen Platzierung von Schwerpunkten. Dass die Lernmöglichkeiten der Kindheit unterschätzt wurden, ist wissenschaftlich gut fundiert und findet in der Frühförderung und dem frühen Erlernen einer Fremdsprache eine plausible bildungspolitische Umsetzung.

4. Die Konzepte zur *Entwicklung der Lehrprofessionalität*, wie sie z.B. die neuen Programme zum mathematischen oder naturwissenschaftlichen Unterricht kennzeichnen, die insbesondere die fachdidaktischen Fähigkeiten und die professionelle Selbstreflexivität stärken, sind überzeugende Wege in die Zukunft.

Entwicklung der
Lehrprofessionalität

Wer aber die pädagogischen und bildungspolitischen Entwicklungen der letzten Jahre verfolgen konnte oder dies sogar in der eigenen Wissenschaftsbiographie erleben durfte, der ist von der skeptischen Erfahrung und von der Verwunderung geprägt, dass die heutigen Forderungen teilweise bereits vor dreissig Jahren erhoben und entsprechende Entwicklungen eingeleitet wurden. Dies gilt für die lernzielorientierten Tests, die kriterienbezogene Leistungsmessung (Flehsig & Haller, 1973) ebenso wie für die Systemevaluation, die Reform der Verwaltung in Richtung autonomer Schulen (Deutscher Bildungsrat, 1975b) und in Bezug auf die Stärkung der fachdidaktischen Professionalität der Lehrer über eine praxisnahe Curriculumentwicklung (Deutscher Bildungsrat, 1975a). Keiner dieser Bemühungen war ein dauerhafter Erfolg beschert. Deshalb stellt sich die Frage, die nicht zuletzt durch das Verschwinden der Reforminitiativen der 70-er Jahre inspiriert ist, welches Ausmass an Veränderung in der alltäglichen Praxis verkraftet wird und welche Massnahmen durch die alltäglich funktionierende Bildungsrealität neutralisiert werden. Wird die bestehende Praxis die Neuerungen abtossen oder wird sie produktiv mit ihnen umgehen? Werden wir in absehbarer Zeit wieder auf das Zerplatzen von Reformblasen zurückblicken? Diese Fragen stehen im Hintergrund der folgenden Ausführungen.

Die heutigen Forderungen
sind nicht neu.

2. Qualitätssicherung in historisch-vergleichender Perspektive

Die Feststellung, dass jede Bildungsreform auf ein nach benennbaren Regeln funktionierendes Bildungswesen trifft, wird wohl kaum nennenswerten Widerspruch erfahren. Dennoch hat diese schlichte These eine weit reichende Bedeutung. Sie kann implizieren, dass je nach der Beschaffenheit der Reformmassnahmen diese «geschluckt», an die bestehende Funktionsweise anverwandelt und damit in der Praxis neutralisiert und unwirksam gemacht werden.

Die Systemtheorie hat für diese Problematik einen einleuchtenden Begriff gefunden, der zumindest hilft, das Problem zu benennen: Jede Reform müsse an das bestehende System *anschlussfähig* sein. Folgt man dieser These, dann muss herausgearbeitet werden, wie das bestehende Bildungswesen funktioniert, an das Reformen anchlussfähig sein müssen, wenn sie wirksam sein wollen.

Reform muss anchlussfähig an
bestehendes System sein

Um die historisch entstandene Funktionsweise des Bildungswesens zu erarbeiten, sei von einer Beschreibung der herkömmlichen Programmsteuerung im Bildungswesen ausgegangen, die jüngst Jürgen Baumert (2001) vorgenommen hat. Er beschreibt die existierende Makrosteuerung des Bildungswesens als *regulative Programmsteuerung (Stundentafeln, Lehrpläne), Mittelallokation und Ausführungskontrolle durch die Schulaufsicht*. Die Übereinstimmung von Vorgabe und Ergebnis wird durch die Schulaufsicht geprüft. In der Folge wird von einer selbstverständlichen *Übereinstimmung zwischen Absicht und Realisation* ausgegangen. Dass die Schule ihre Ziele tatsächlich erreicht, wird als naturgegeben unterstellt. Diese Annahme schützt die Schule vor einer Rechenschaftslegung. „In diesem System ist die Thematisierung von Bildungser-

Aktuelle Makrosteuerung im Bil-
dungswesen beinhaltet regulative
Programmsteuerung, Mittelal-
lokation und Ausführungskontrolle

gebissen nicht nur überflüssig, sondern sogar ein Fremdkörper. Allein die Frage nach ihnen ist schon eine implizite *Verwaltungsbeschwerde ...*“ (Baumert, 2001, S. 14).

Die Mittelallokation baut auf der Gleichheit der institutionellen Bedingungen auf.

Es kommt jedoch eine wichtige Besonderheit hinzu. Die Steuerung des Bildungswesens ist nicht nur von der Aufsicht, sondern auch von der Mittelallokation so gestaltet, dass von einer *Gleichheit der institutionellen Bedingungen* für alle Schüler/innen ausgegangen wird. Alle Schulen erhalten gleich viel Geld, die Lehrkräfte werden ihnen zentral zugeteilt, das Inhaltsangebot ist gleichförmig, ebenso die Ausbildung der Lehrkräfte und damit die unterstellte Qualität der Lehrenden. Baumert resümiert: „*Schulunterschiede* sind Webfehler im System, die, wenn sie bekannt werden, im Grunde ein Eingreifen der Schulaufsicht verlangen“ (Baumert, 2001, S. 15).

Leistungsunterschiede werden auf unterschiedliche Nutzung zurückgeführt

Wenn damit von der Fiktion ausgegangen wird, dass alle Schülerinnen und Schüler ein qualitativ gleichwertiges Angebot erhalten und darüber die Chancengleichheit als realisiert gesehen wird, dann wird die Verantwortung für unterschiedliche Leistungen, die natürlich unübersehbar sind, der individuellen Nutzung durch Schülerinnen und Schüler und dem Elternhaus zugeschrieben. Sie tragen damit auch die alleinige Verantwortung für die erbrachten Leistungen.

Qualitätsunterschiede der Angebotsseite

Dass die *Qualität des Angebotes* systematisch variieren kann und deshalb auch auf der Angebotsseite Verantwortung für die Lernergebnisse und Erziehungswirkungen des Bildungswesens anzusiedeln ist, ist hier *systemfremd*. Die Kultusministerien in Deutschland haben sich konsequenterweise über Jahre gegen eine vergleichende Evaluation der Leistungen des Systems gewehrt.

Baumert hat sich in seiner Diagnostik auf die *deutschen Bildungssysteme* bezogen. Steuerungsprobleme sind in anderen Bildungssystemen möglicherweise ganz anders gelagert. Meine Erfahrungen in österreichischen, deutschen und schweizerischen Bildungssystemen haben mich für Unterschiede in Steuerungsformen sensibilisiert und zugleich die Überzeugung gestärkt, dass man die nach Ländern variierenden Strategien der Qualitätssicherung nur durch ihre historische Rekonstruktion verstehen kann.

In allen Ländern, die ich aus eigenem Erleben kenne, scheint es eine Gemeinsamkeit zu geben: Qualitätssicherung bedeutet überall, dass die *Obrigkeit* ein waches Auge auf das Geschehen wirft und überprüft, ob alles in Ordnung ist. Wer die Obrigkeit ist und wie diese Überprüfung stattfindet, unterscheidet sich in den genannten Ländern. Ich möchte im Folgenden diese unterschiedlichen Verfahren der Qualitätssicherung beschreiben. Ich komme zuerst zum Bildungssystem des Kantons Zürich.

2.1. Fallstudie Zürich

Die Qualitätssicherung im Kanton Zürich bereits über 200-jährig

Die Geschichte des Bildungswesens des Kantons Zürich offenbart die Überraschung, dass Qualitätssicherung nicht neu ist. Schon im 17. Jahrhundert sind im Kanton Zürich die Pastoren beauftragt worden, in allen Familien nachzuforschen, wie der Bildungsstand ist: Wer wie gut lesen kann und welche Bücher gelesen werden. 1774 beobachten wir eine erste grosse Enquête in Zürich, später eine bundesweite durch Philipp Albert Stapfer (1766-1840). Die Idee der flächendeckenden Überprüfung der Qualität des Bildungswesens durch dafür ausgewählte Verantwortliche wurde dann gleichzeitig mit der Einführung der Volksschule 1831 realisiert.

In mehrere Jahrzehnte andauernden Auseinandersetzungen wurde schliesslich eine säkularisierte Form der Aufsicht über das Volksschulwesen etabliert, die vor allem auf der Gemeindeebene angesiedelt war. Gewählte Vertreter der Gemeinde waren und sind bis heute das Aufsichtsorgan für die Lehrarbeit. Lehrerinnen und Lehrer wurden und werden immer noch mehrmals im Jahr besucht und beurteilt. Am Ende des Schuljahres fand früher eine öffentliche Abschlussprüfung statt, in der die Lehrperson zeigen sollte, was ihre Schüler/innen gelernt hatten. Die regulativen Prozesse und die Festlegung der Lerninhalte erfolgten auf der kantonalen Ebene durch den Erziehungsrat.

Gewählte Gemeindevertreter sind bis heute das Aufsichtsorgan für die Lehrarbeit.

Man sucht deshalb bis heute in Volksschulen vergebens nach dem Schulleiter. Entscheidungen werden von der lokalen Schulpflege gefällt. Die Einführung geleiteter Schulen ist vor diesem Hintergrund ein neuer Weg der Qualitätssicherung, der zu einer Stärkung der Schule als teilautonomer Organisation führen kann, aber neue Abstimmungsprozesse mit der lokalen und kantonalen Schulaufsicht erfordert, welche durch eine starke Organisationseinheit Schule eher geschwächt werden.

Wie wird eine Lehrperson in einem solchen gemeindenahen Bewertungsrahmen seine Arbeitssituation optimieren? Sie kann sich offensichtlich keine permanenten Fehlleistungen erlauben, da diese im engen Umfeld der Gemeinde durch die häufigen Besuche und durch die Rückmeldungen der Schüler/innen schnell rüchbar werden. Sie muss sich aber auch mit den lokalen Schlüsselfiguren arrangieren, um deren Wohlwollen zu sichern. Dass dies bei Kindern, deren Eltern in den Gemeinden nicht zu den Honoratioren gehörten, zu schmerzhaften Erfahrungen führen konnte, belegen Schulerinnerungen um die Wende vom 19. zum 20. Jahrhundert (Schohaus, 1950).

Mit dem Ausbau der weiterführenden Bildungswege in Gymnasien und in Sekundarschulen entwickelte sich im 20. Jahrhundert ein ausgedehntes *Prüfungswesen*. Wer in Gymnasien oder Sekundarschulen wollte, der hatte dort Aufnahmeprüfungen zu absolvieren. Dies ist bis heute so. So waren im Kanton Zürich in den letzten Jahren nur ca. 50% der Bewerber für die Langzeitgymnasien nach dem 6. Schuljahr (nach dem 9. Schuljahr ergibt sich eine zweite Chance) in den Aufnahmeprüfungen sowie dem folgenden Probejahr erfolgreich.

Eine Prüfung regelt den Übergang von der Primar- zur Sekundarschule, bzw. zum Gymnasium.

Diese *Übergangsprüfungen* schaffen höchst wirksame Richtpunkte für die Volksschullehrerinnen und Volksschullehrer. Sie werden nun daran gemessen, ob es ihnen gelingt, Kinder erfolgreich auf diese Übergänge vorzubereiten. Beurteilen sie ihre Kinder als gut und reüssieren diese bei Aufnahmeprüfungen und im Probejahr des Gymnasiums nicht, so hat dies Rückwirkungen auf die Beurteilung der Lehrerinnen und Lehrer in der Gemeinde. Die hohen Schwellen für den Eintritt ins Gymnasium führten jedoch zu einer zweiten Entwicklung. Um Anforderungen zu bewältigen, haben insbesondere wohlhabende Eltern ihre Kinder zur Vorbereitung der Prüfung in *Zweitschulen* geschickt, in die so genannten Lernstudios. Ähnlich wie in Japan führt damit eine integrierte Beschulung bis zur 6. Schulstufe zusammen mit einer strengen Auslese für den Weg ins Gymnasium über Aufnahmeprüfungen neben den obligatorischen staatlichen Einrichtungen zu „Zweitschulen“.

Zweitschulen etablieren sich neben den staatlichen Einrichtungen.

Wie erfolgt nun Qualitätssicherung und Aufsicht in den Mittelschulen? Kennzeichnend für Zürich ist hier das Fehlen einer flächendeckenden Kontrolle. Die einzelnen Schulen sind relativ autonom. Sie werden reguliert durch schulspezifische Aufsichtskommissionen. Schulautonomie ist hier also nicht das Problem, eher schon die fehlende schulübergreifende Koordination. Man sucht in den Mittelschulen vergebens nach einem *gemeinsamen und für alle gültigen Lehrplan*. Stattdessen findet man kur-

An Mittelschulen fehlt ein gemeinsamer und für alle gültiger Lehrplan.

ze Rahmenregelungen in den jeweiligen Schulen, die bis vor wenigen Jahren nur wenige Seiten umfasst haben. Die Eidgenössische Maturitäts- anerkennungsverordnung, die in den letzten Jahren auf Bundesebene erarbeitet wurde, hat hier eine neue Orientierung geschaffen.

Wie können auf dieser Basis eine Qualitätssicherung und eine Standardsicherung zustande kommen?

Diese Frage haben sich nicht nur Aussenbeobachter immer wieder gestellt, sondern auch in der Schweiz wurde darüber oft diskutiert. Die Vermutung, dass die historischen Formen der Evaluation ungenügend seien und sehr disparate, lokal bestimmte Bildungsverhältnisse erzeugen könnten, wurde immer wieder vorgebracht. Die Heterogenität wurde nicht auf kantonsinterne Verhältnisse begrenzt gesehen, sondern noch ausgeprägter für die ganze Schweiz erwartet. Nicht zuletzt dieser kantonspezifische Charakter der Bildungssysteme hat im 19. Jahrhundert dazu geführt, dass der Militärdienst als Basis für umfassende Bildungsevaluation genutzt wurde. Da alle männlichen 20-Jährigen zu so genannten Rekrutenschulen eingezogen wurden, bot sich die Gelegenheit, deren Bildungsniveau zu prüfen. Sie wurden in diesem Rahmen seit 1875 bis zum Beginn des Ersten Weltkrieges systematisch auf ihre Fähigkeiten hin getestet. Die Ergebnisse wurden auf Kantonsebene hochgerechnet, aber auch in das Dienstbuch der Rekruten eingetragen, das sie durch ihr ganzes Leben begleitete.

Militärdienst wird als Basis für eine umfassende Bildungsevaluation genutzt

**Ergebnisse der Rekrutenprüfungen.
1875-1882.**

Kantone.	75.	76.	77.	78.	79.	80.	81.	82.	Durchschnitt.
Jürich	4	5	4	3	4	2	4	3	3
Zern	15	21	13	15	15	17	18	20	19
Fuzen	12	8	10	12	9	18	19	21	15
Uri	19	19	24	23	21	24	24	22	23
Schwyz	23	17	22	21	20	19	21	18	21
Obwalden	21	16	11	9	6	10	14	6	11
Nidwalden	24	23	15	19	13	22	23	16	20
Glarus	17	22	20	18	17	6	13	7	16
Sug	14	12	8	8	7	13	12	9	8
Freiburg	20	18	21	22	24	21	20	24	22
Solothurn	9	7	7	7	12	15	10	8	7
Baselst.	1	2	1	1	3	1	1	1	1
Baselst.	10	13	16	10	14	16	18	19	13
Schaffhausen	6	6	6	4	3	5	5	2	5
Appenzel.	11	14	12	20	22	14	17	14	17
Appenzel.	25	24	23	25	25	25	25	23	25
St. Gallen	8	11	9	16	18	9	15	13	12
Graubünden	13	15	17	13	16	7	11	15	14
Argau	16	9	14	17	10	8	6	12	10
Zürich	3	1	3	5	8	4	3	4	4
Zürich	16	40	19	11	19	20	7	17	18
Basel	5	4	5	5	5	11	8	11	6
Soloth.	23	25	25	24	23	23	22	25	24
St. Gallen	7	10	13	14	11	12	9	10	9
Gené	2	3	2	2	1	3	2	5	2

Kantonsrangliste für 1875 - 1882

Abb. 1: Kantonsvergleiche auf der Grundlage der Rekrutenprüfungen

Quelle: (Lustenberger, 1996, S. 93)

Die Ranglisten lösten heftige Reaktionen aus.

Die auf dieser Grundlage entstandenen Kantonsranglisten haben damals ähnlich heftige Reaktionen hervorgerufen, wie dies die heutigen Ländervergleiche im Rahmen des „Programms of International Student Assessments“ (PISA) tun (Lustenberger, 1996). Abb. 1 macht sichtbar, wie die Kantonsvergleiche ausgesehen haben. Die Daten wurden bis auf einzelne Gemeinden, ja bis auf persönlich genannte – und gelegentlich von den Gemeinden besonders ausgezeichnete – Einzelpersonen spezifiziert. Natürlich war damit identifizierbar, welche Lehrperson Schüler mit besonders guten oder schwachen Leistungen hatte.

Diese Evaluationen haben über Jahre die öffentliche Diskussion zur Entwicklung des Bildungswesens bestimmt und auch geholfen, Reformen, wie z.B. die Einrichtung von Fortbildungsschulen, zu befördern. Aber auch Berichte über die Folgen für die Lehrerschaft sind lehrreich: „Es war kein guter Tag, wenn die Zeitungen die Übersichten brachten! Ich fühlte dann meinen Vater zusammensinken, an sich und seiner Arbeit irre werden und den grossen, alles niederbrechenden Verleider bekommen, der eine fruchtbare Lehrerarbeit so unendlich schwer macht.“ (Lustenberger, 1996, S. 112). Lustenberger berichtet sehr aktuell anmutende Initiativen im Umkreis dieser Evaluationen. Durch die personenbezogenen Rückmeldungen kamen die Schulaufsicht, Gemeinden, die Lehrer und nicht zuletzt die Jugendlichen selber unter permanenten Druck, der von vielen als unerträglich empfunden wurde. Die provisorische Abschaffung dieser Rekrutenprüfungen im Ersten Weltkrieg hat denn auch dazu geführt, dass sie nach Kriegsende nicht mehr aufgenommen wurden.

Personenbezogene Rückmeldungen erzeugten Druck

Zurück zur Gegenwart: Angesichts der heute noch bestehenden *schwachen Zentralsteuerung der Standards* auf kantonaler Ebene und auf Bundesebene stellt sich verschärft die Frage, wie trotzdem hohe Erwartungen und hohe Anstrengung gesichert werden und auch beim Fehlen der Rekrutenprüfungen ab 1914 gesichert werden konnten.

Die erste Antwort ist oben enthalten: Die Standardsicherung erfolgt heute immer noch primär lokal, in den Schulpflegen der Gemeinden und durch die *Schulaufsichtskommissionen* bei Gymnasien.

Als Kernelement der Steuerung muss dabei die Ausrichtung auf strenge *Aufnahmeprüfungen* gesehen werden. Die zweite Steuerungsfunktion übernehmen die *Lehrmittel*. In ihnen sind die Standards enthalten und sie steuern über ihre verpflichtende Einführung das Erwartungsniveau. Dennoch liegt ein Kernproblem, das wohl bis heute besteht, in der Herstellung von Vergleichbarkeit. Es ist hier also ein Weg von der Vielfalt zur Koordination notwendig, der sich auch in den letzten Jahren abgezeichnet hat, sei es in Lehrplänen für die Volksschulen, in Maturitätsanerkennungsverordnungen und heute im so genannten *HarmoS-Projekt* (Harmonisierung der obligatorischen Schule) der Eidgenössischen Konferenz der kantonalen Erziehungsdirektoren, die eine Standardisierung der Anforderungen auf Bundesebene anstrebt.

Kernelemente der Steuerung sind strenge Aufnahmeprüfungen und Lehrmittel.

2.2. Fallstudie Deutschland

Die *Standardsicherung* hat in Deutschland im Vergleich zur Schweiz andere Wege eingeschlagen. Der wichtigste Unterschied ist der, dass bis ins 20. Jahrhundert die Volksschulen unter *kirchlicher Schulaufsicht* standen, während in der Schweiz die kommunale und weltliche Aufsicht etwa ab der Mitte des 19. Jahrhunderts die Oberhand gewann (Frey, 1955). Der Volksschullehrer hatte dadurch z.B. in Baden und Württemberg über das gesamte 19. Jahrhundert die inferiore Stellung des so genannten *Messnerlehrers*. Um sich aus der kirchlichen Abhängigkeit zu emanzipieren, hat die Volksschullehrerschaft eine starke Aufsicht durch den Staat begrüsst und sich in dessen Aufsichtsorganen und den zentralen Regelungen des Staates beheimatet. Damit wird die *grosse Bedeutung der Dienstaufsicht durch den Staat* verständlich und die *starke Rolle, die bis heute in Deutschland zentrale staatliche Regelungen* haben, wobei „Staat“ heute „Bundesland“ bedeutet. Die enge Verbindung von Schulaufsicht, zentral verantworteter Bildungspolitik, Beamtenstatus der Lehrkräfte und zentraler, gleichheitsorientierter Inputsteuerung charakterisieren die derzeitige Qualitätssicherung im deutschen Bildungswesen.

Standardsicherung in Deutschland ist geprägt durch kirchliche Schulaufsicht.

Die staatliche Dienstaufsicht erhält grosse Bedeutung.

Zentrale Abiturprüfungen bilden den Kern der Standardsicherung.

Für die Gymnasien ist eine andere Entwicklung kennzeichnend. Die Stärke des Gymnasiums resultiert nicht so sehr aus der neuhumanistischen Kulturbewegung, sondern aus der Standardsicherung des Zugangs zu Universitäten, die sich etwa von 1780 bis 1850 entwickelte. Dadurch wurde die höchst unterschiedliche Qualität lokaler Gymnasien in einen gemeinsamen Anforderungshorizont einbezogen, der zusammen mit der staatlichen Schulaufsicht zum Kerninstrument der gymnasialen Qualitätssicherung wurde. Prüfungssysteme, heute in der Gestalt zentraler Abiturprüfungen, bilden bis heute den Kern der Standardsicherung. Die Qualitätssicherung des Inputs, also der Ressourcenausstattung und der Professionalität des Unterrichts, erfolgt über eine staatliche, mehr oder weniger öffentlich verantwortete Bildungspolitik und Bildungsverwaltung. Im Mittelpunkt der Standardsicherung stehen aber externe und interne Prüfungen, deren Regelwerk bis in den Unterrichtsprozess hineinwirkt. Das Besondere der deutschen, aber auch der Schweizer Prüfungssysteme besteht in ihrem überwiegend *terminalen, Abschlüsse vermittelnden Charakter*. Die Matura bzw. das Abitur vermittelt die Berechtigung für den Besuch von Hochschulen auf der Grundlage jahrelanger, klasseninterner Prüfungen durch die Lehrpersonen.

2.3. Fazit: Institutionelle Regeln und die Grammatik der Schule

Grammatik der Schulregulierung prägt Lehrhandeln und Handeln der Schüler/innen

Das Handeln der Lehrkräfte wird von den oben formulierten Rahmenregelungen zu Prüfungen, Schullaufbahnen und Abschlüssen tagtäglich geleitet. Sie repräsentieren ein Regelwerk ähnlich einer Grammatik der Sprache. Wie letztere das Sprachhandeln reguliert, leiten die institutionellen Regelungen des Bildungswesens das Handeln (das „Sprechen“) der Lehrkräfte. Dabei ist zu beachten, dass dieses Regelwerk eine lebbare Konsistenz hat und zu einer Gewohnheit des „Sprechens“ führt. Jeder versteht den anderen, wenn er sich regelkonform verhält. Diese Grammatik schafft Erwartungssicherheit, aber auch bestimmte Formen des „schulischen Sprechens“ und eine sie begleitende Semantik, die anderen, die nach einer anderen Grammatik der Schulregulierung leben, fremdartig erscheinen mag.

Werden die Regeln geändert, wird also die Grammatik neu konzipiert, dann verändern sich auch die Formen des Sprechens, dann verändert sich das Lehrhandeln und das Handeln der Schülerinnen und Schüler.

Lehrperson ist Lehrende und Richtende

Nach der herkömmlichen Grammatik ist die Lehrperson tagtäglich in ein Muster von individueller Unterrichtsvorbereitung, lehrpersonengeleitetem Lektionenhalten im Klassenverband und Prüfungsabfolgen im Anschluss an Lektionen eingespannt. Sie trägt die Verantwortung für die Qualität der Stoffpräsentation und das Niveau der Anforderungen. Sie ist damit gleichzeitig Lehrende und Richtende.

Nach einer anderen Grammatik könnten diese beiden Funktionen getrennt sein. Wenn die Prüfungen extern geplant und durchgeführt werden, wie dies in vielen erfolgreichen Ländern geschieht, dann verändert sich die Regulierung. Die Lehrperson ist vornehmlich verantwortlich für die bestmögliche Unterrichtsgestaltung im Interesse des Schülers, der sich dann ausserhalb der Schule zu bewähren hat.

Veränderungen von Steuerungsformen implizieren nach dieser Metapher der Grammatik Änderungen der Regeln für das Handeln der Lehrpersonen sowie für die Allokation von Entscheidungsmöglichkeiten. Wenn andere Länder in Leistungstests besser abschneiden als das eigene, dann liegt es nahe, dort nach der Grammatik der Schulregulierung zu suchen, die offensichtlich erfolgreich ist. Es ist menschlich, dort dann die Regeln zu suchen, die einem gut gefallen und sie in die eigene Grammatik integrieren

zu wollen. Dabei besteht die Gefahr, dass die Konsistenz der Regeln und ihre gegenseitigen Ergänzungsbeziehungen nicht gesehen werden, so dass – wenn nur die eine Regel und nicht die dazu gehörigen anderen übernommen wird – kontraproduktive Wirkungen im Rahmen der bestehenden Grammatik der Schulregulierung entstehen. Im eigenen Land möchte man nicht das ganze Paket der Regeln übernehmen, sondern nur jene, die bei einem Vergleich mit dem eigenen Regelsystem vorteilhaft erscheinen. Die Übernahme der gemeinsamen Schule für alle bis zum 9. Schuljahr ist dafür ein markantes Beispiel. Sie funktioniert nur unter anderen Regeln wie schulexternen Leistungsprüfungen und Aufnahmeprüfungen von Absolventen. Das ganze „Bundle“ wird selten gesehen und noch weniger häufig als ganzes akzeptiert.

Bei der Übernahme einzelner Regeln aus erfolgreichen Steuerungssystemen besteht die Gefahr kontraproduktiver Wirkungen.

3. Aufgaben der wissenschaftlichen Begleitung von Reformprozessen

Die Bildungsforschung hat über die PISA-Studien in den letzten Jahren eine Beschreibung der Effektivität von Bildungssystemen geleistet, die nicht mehr administrativ abgewehrt und ausgegrenzt werden kann. Sie hat deutlich machen können, dass auch die Qualität des Angebotes etwas mit dem Leistungsausweis des Bildungswesens zu tun hat und die Verantwortung nicht völlig auf die Nutzungsseite geschoben werden kann.

Hat sie damit ihren Dienst getan und steht sie vor einer vollendeten Aufgabe, die gleichzeitig einen kontinuierlichen Bedeutungsverlust von Bildungsforschung einleitet?

In meinen Augen ist das Gegenteil der Fall. Die reformbegleitende Forschung der letzten Jahre, und viele Projekte im FS&S sind dazu zu zählen, hat sichtbar gemacht, dass es ein neuer Rückfall in die Annahme der Identität von Plan und Praxis wäre, auf prozessbegleitende Studien zu verzichten.

Reformbegleitende Forschung ist wichtig.

Neben der akzeptanzorientierten Überprüfung der Zielerreichung hat die reformbegleitende Forschung vor allem auf die unbeabsichtigten Nebenwirkungen zu achten. Sie aber lediglich als Abfallprodukte der Zielerreichungsbemühungen zu sehen, würde zu kurz greifen. Ihnen ist vielmehr ein systematischer Stellenwert zuzuweisen, den wir durch den Rekurs auf die Metapher der Grammatik der Schulregulierung und der sie begleitenden Semantik präzisieren können. Wenn Lehrpersonen im Rahmen einer Grammatik der Schule zu einem bestimmten regelgeleiteten Handeln tendieren, dann wäre systematisch zu fragen, welches die praktischen Folgen von bestehenden Regulierungen sind und wie neue Regelungen in diese praktischen Handlungsorientierungen eingreifen. Wir müssten uns also der weitgehend unerforschten Frage zuwenden, wie Lehrkräfte unter den verschiedenen Verfahren der Standardsicherung funktionieren und welche Folgen mit der damit zusammenhängenden unterschiedlichen Praxis verbunden sind. Dabei ist davon auszugehen, dass Lehrpersonen intelligent, selbstwertdienlich, mit Defensivität begabt und überlebensinteressiert mit Vorgaben umgehen, die sie jeweils auf ihre alltäglichen Handlungsanforderungen rekontextualisieren. Diese Übersetzungsleistungen der Lehrerschaft gilt es zu beobachten und zu erforschen. Die Kernfrage liegt auf der Hand: Wie verändern neue Formen der Standardsicherung über Bildungsstandards, Vergleichsarbeiten, zentrale Tests und zentrale Abschlussarbeiten das Unterrichtsgeschehen in Schulen?

Es muss dabei auf die unbeabsichtigten Nebenwirkungen geachtet werden.

Standardsicherung und *Autonomie*, dies sind die Kernpunkte der derzeitigen bildungspolitischen Gestaltungsbemühungen. Im Rahmen der obigen Überlegungen ergeben sich daraus auch zwei grosse Bereiche der Bildungsforschung.

**Begleitforschung muss
Folgen der neuen
Informationsmöglichkeiten
beobachten.**

1. Die empirische Begleitforschung muss sich vor dem obigen Hintergrund vor allem um die Beobachtung der Folgen bemühen, die durch die neuen Informationsmöglichkeiten über den Leistungsstand in Ländern, Kantonen, Schulen und Klassenzimmern entstehen können. Qualitätsbezogen wäre zu fragen: Wie können die Informationen, die über standardbasierte Testaufgaben erhalten werden, genutzt werden, um die Qualität des Bildungswesens zu verbessern?

Diese nur empirisch beantwortbare Frage kann jedoch auf viele Erfahrungen zur standardisierten Leistungsmessung in anderen Ländern, insbesondere in den USA und in England, zurückgreifen. Dort stehen sie auch im Rahmen einer kulturspezifischen Grammatik und Semantik, die bei der Übertragung ins eigene Land zu berücksichtigen ist.

**Gewinnung von Informationen
über Zielerreichung
und Problemzonen**

Die Initiative zur standardorientierten Leistungsmessung im Bildungssystem wird vor allem ein Ergebnis haben: Es werden neue Informationen zur fachspezifischen, schulformspezifischen und schulspezifischen Zielerreichung bzw. über Problemzonen geschaffen. Die zu untersuchende Kernfrage lautet in der Folge: Welche Verwendung finden Informationen über die Leistungsqualität des Bildungswesens auf verschiedenem Aggregationsniveaus: auf Systemebene, in den einzelnen Schulen, bei den einzelnen Lehrpersonen und bei den Schülerinnen und Schülern?

So können diese Informationen genutzt werden, um eine objektive Rückmeldung über den Stand der Leistungen auf den verschiedenen Ebenen des Bildungswesens zu erhalten. Sie würden auf dieser Stufe der Verwertung lediglich der *Orientierung* der Bildungspolitik und der Lehrerschaft dienen. Massnahmen blieben den Initiativen der Politik und Verwaltung vorbehalten.

**Erweiterte Nutzung der Infor-
mationen für Entlohnung und
Ressourcenzuweisung**

Eine erweiterte Nutzung könnte darin bestehen, sie zum Ausgangspunkt für die Entlohnung und Ressourcenzuweisung für die einzelnen Schulen zu machen. Dies geschieht in England, wenn die schulbasierten Leistungsergebnisse für diese Zwecke genutzt werden. Dies könnte bis hin zur leistungsabhängigen Beurteilung einzelner Lehrerinnen und Lehrer geschehen.

**Nutzung der Informationen für
individuelle Beurteilung**

Eine dritte Verwendungsweise würde die Leistungsergebnisse zur individuellen Beurteilung der Schülerinnen und Schüler einsetzen (High Stake Testing). Die schulübergreifend eingesetzten Leistungstests wären dann Teil des gesamten, individuumbezogenen Prüfungssystems.

**Nutzungsprozesse von
Leistungsrückmeldungen
müssen untersucht werden**

Es wäre sehr wichtig, dass Nutzungsprozesse von Leistungsrückmeldungen beobachtet werden und dass international studiert wird, welche Wirkungen mit ihnen verbunden sind (siehe für vertiefende Analysen: Arbeitsgruppe Internationale Vergleichsstudie, 2005). Ohne Verwertungsprozesse wäre das Monitoring selbstverständlich sinnlos. Welche es aber sein sollen, gilt es praktisch zu erproben und im günstigen Falle empirisch zu begleiten. Monitoring kann sehr Unterschiedliches bewirken und zwar je nach den Konsequenzen, die mit ihm verbunden werden, sei dies ein öffentliches Schulranking, eine leistungsabhängige Ressourcenausstattung der Schule oder gar eine leistungsabhängige Belohnung von Lehrkräften.

Nach meiner Einschätzung, die es natürlich zu überprüfen gälte, wäre eine primär informative und orientierende Verwendung von schulübergreifenden Standards mit gekoppelten Testaufgaben sinnvoll. Sie würde ein Qualitätsbewusstsein wach halten, die Selbstevaluation anregen und die bestehenden Standardsicherungen intakt lassen, sie wäre also am ehesten an die bestehende Praxis der Inputsteuerung, der Prüfungen und Aufsichtsverfahren anschlussfähig.

Eine primär informative und orientierende Verwendung von schulübergreifenden Standards wäre sinnvoll.

Für eine Akzeptanz fördernde Rückmeldekultur erscheint es mir ferner bedeutsam zu beobachten, wie gut Leistungsrückmeldungen auf Verantwortungen abgestimmt sind. Grundsätzlich gilt, dass man nur für etwas folgenreich evaluiert werden kann, wofür man auch verantwortlich ist. Wenn die Leistungen von zwei Lehrpersonen einer Schule als Durchschnittswerte erscheinen, jede Lehrperson aber für ihren Unterricht verantwortlich ist, dann wäre es inadäquat, über die Leistungen beider Lehrkräfte einen Durchschnitt zu bilden und zur Grundlage für eine leistungsabhängige Entlohnung zu machen.

Leistungsrückmeldungen müssen auf Verantwortungen abgestimmt werden

2. Ein zweites Beobachtungsfeld sollte sich der Frage nach den positiven oder negativen Auswirkungen der Schulautonomie zuwenden. Wie verändert sich die Pädagogik der Einzelschule und die Pädagogik in den Klassenzimmern durch die Stärkung der Schulebene im Sinne einer grösseren Autonomie bzw. der Einrichtung geleiteter Schulen? Wie verändert sich der Arbeitsplatz Schule für die Lehrperson und der Lernort Schule für die Schülerinnen und Schüler? Führt die Pflicht zur Rechenschaftslegung zu einer Stärkung der Qualität von Lehrkollegien und von Erziehungsgemeinschaften zwischen Elternhaus und Schule? Lässt sich eine kluge Verbindung von Selbst- und Fremdevaluation finden, die das Bildungswesen spürbar voranbringt?

Begleitforschung muss als zweites Beobachtungsfeld die Auswirkungen der Schulautonomie prüfen

Dies sind Fragen, die nicht rein theoretisch am runden Tisch beantwortet werden können. Es wäre durchaus möglich, dass standardorientierte Steuerungsmassnahmen nicht den Kern der Qualitätsentwicklung treffen und dass wir diesen in der Entwicklung der Professionalität der Lehrkräfte und ihrer fachdidaktischen Schulung für guten Unterricht zu suchen haben. Wahrscheinlicher ist, dass eine förderliche Schulentwicklung in der Verbindung von Massnahmen auf Makroebene und Mikroebene zu suchen ist und nicht in den in der Vergangenheit oft praktizierten Polarisierungen von pädagogikfeindlichen Verwaltungen und kinderfreundlichen Lehrpersonen.

Förderung von Schulentwicklung in Verbindung von Massnahmen auf Makro- und Mikroebene

4. Ausblick: Steuerungskonzepte und Menschenbild

Auch wenn wir empirisch nach Wirksamkeit, nach Prozessen und Zielerreichung schulischer Bildungsprozesse fragen, dürfen wir nicht unterstellen, damit ausreichend gerüstet zu sein, um klug und verantwortlich «Menschengestaltung» über das Bildungswesen zu betreiben. Wir haben einerseits die Frage nach der Erreichbarkeit von Zielen zu stellen und empirisch zu belegen, und andererseits normativ zu fragen, was wir wollen und wie wir dieses erreichen wollen. Auch wenn z.B. optimale Wirksamkeit erreicht würde, indem Lehrpersonen sowie Schülerinnen und Schüler permanent und detailgenau kontrolliert und evaluiert werden, müsste immer noch die Frage gestellt werden, ob dies mit unserem Menschenbild vereinbar ist. Wir haben in unserer Kultur eine Vorstellung von Verantwortung, von Freiheit und Pflicht

Die Frage nach der Vereinbarkeit von Regelsystem mit Menschenbild muss immer gestellt werden.

entwickelt, welche die Lehrpersonen als selbständig Denkende und verantwortlich Handelnde versteht. Ihre Instrumentalisierung für Zwecke einer Institution hat normative Grenzen. Gleiches gilt für die Schülerinnen und Schüler, die nicht nur in Instrumentalitäten und Effektivitäten eingeübt werden sollen, sondern in eine Gestalt selbstverantwortlichen und mündigen Menschseins.

**Kinder lassen sich nicht
intentionsgetreu „machen“**

Viele Konzepte der gegenwärtigen Schulreform erwecken zumindest von der Semantik her den Eindruck, Bildungswirkungen liessen sich in gleicher Weise zielgenau planen und realisieren wie die Herstellung technischer Produkte. In der so genannten Technologie-Debatte der Pädagogik glaubte man, diese Vorstellung überwunden zu haben (Dreeben, 1970; Luhmann & Schorr, 1982). Kinder lassen sich nicht intentionsgetreu „machen“, es können ihnen nicht gleichförmig in abmessbaren Zeiträumen Kompetenzen eingepflanzt werden. Kant hatte schon die Problematik des Einwirkens auf Freiheit formuliert (Kant, 1803, 1964) und damit das philosophisch formuliert, was die moderne Psychologie unter selbstreferentiellen und eigenintentionalen Personen versteht. Die schulischen Lernangebote treffen auf kindliche Nutzungsprozesse, die freudige Zuwendung aber auch defensive Abwehr erzeugen können. Vor allem: Die Erfolge der intentional formulierten Angebote gewinnen hier eine bedeutsame Unschärfe. Ob man darauf mit einem Quasi-Kausalitätssatz von Prüfungszwängen oder mit der Schaffung eines pädagogischen Lernklimas reagiert, dürfte normativ nicht unerheblich sein. Aber nicht nur von ethischen Überlegungen, sondern auch von praktisch-pädagogischen her, sind die Vorstellungen, wie auf Kinder eingewirkt werden kann und soll, bedeutsam. Wer einem schlichten Kontroll- und Herstellungsdenken widersagt, der muss auf andere Wege setzen: Er muss mehr von Motivation und Überzeugung, von Anreizen und bestmöglichem Unterricht, von Förderung und Ressourceneinsatz sprechen. Standardsicherung ist in diesem Fall nur ein Rahmen, der durch pädagogische Förderung auszufüllen ist. Standardformulierungen können dann keinen Präzisionsgrad haben, der fachlichen und insbesondere fachübergreifenden Wirkungen wesensfremd ist.

**Prüfungszwänge versus
pädagogisches Lernklima?**

Diese Überzeugung bildet auch den Hintergrund dafür, warum im Pädagogischen Institut und im FS&S bewusst eine Synthese von Philosophie und Geschichte der Pädagogik, von institutionellen und personalen Betrachtungsweisen pädagogischer Menschenbildung sowie von empirischen Studien zur Realität von Erziehung und Unterricht in modernen Gemeinwesen gepflegt wird.



Prof. Dr. Helmut Fend
Pädagogisches Institut
Universität Zürich
Freiestrasse 36
8032 Zürich
fend@paed.unizh.ch

Literatur

- Arbeitsgruppe Internationale Vergleichsstudie. (2005). *Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten*. Bonn: Bundesministerium für Bildung und Forschung.
- Baumert, J. (2001). Vergleichende Leistungsmessung im Bildungsvergleich. *Zeitschrift für Pädagogik*, 43. Beiheft: Zukunftsfragen der Bildung, 15-56.
- Deutscher Bildungsrat. (1975a). *Empfehlungen der Bildungskommission: Zur Förderung praxisnaher Curriculumentwicklung*. Stuttgart: Ernst Klett Verlag.
- Deutscher Bildungsrat. (1975b). *Empfehlungen der Bildungskommission zur Reform von Organisation und Verwaltung im Bildungswesen, Teil I: Verstärkte Selbständigkeit der Schule und Partizipation der Lehrer, Schüler und Eltern*. Stuttgart: Klett.
- Dreeben, R. (1970). *The nature of teaching. Schools and the work of teachers*. Glenview, Illinois: Scott, Foresman and Company.
- Flehsig, K.-H. & Haller, H.-D. (1975). *Entscheidungsprozesse in der Curriculumentwicklung* (Vol. 24). Stuttgart: Ernst Klett Verlag.
- Frey, P. (1953). *Die zürcherische Volksschulgesetzgebung 1831-1951*. Zürich: Dr. J. Weiss, Affoltern a. A.
- Kant, I. (1803). *Immanuel Kant über Pädagogik. Herausgegeben von D. Friedrich Theodor Rink*. Königsberg: Friedrich Nicolovius.
- Kant, I. (1964). *Schriften zur Anthropologie, Geschichtsphilosophie und Pädagogik. (Werke in sechs Bänden, Band 6)*. Darmstadt: Wissenschaftliche Buchgemeinschaft.
- Luhmann, N. & Schorr, K. E. (Eds.). (1982). *Zwischen Technologie und Selbstreferenz*. Frankfurt a.M.: Suhrkamp.
- Lustenberger, W. (1996). *Pädagogische Rekrutenprüfungen. Ein Beitrag zur Schweizer Schulgeschichte*. Chur: Verlag Rüegger.
- Schedler, K. (1995). *Ansätze einer wirkungsorientierten Verwaltungsführung. Von der Idee des New Public Management (NPM) zum konkreten Gestaltungsmodell: Fallbeispiel Schweiz*. Bern: Verlag Paul Haupt.
- Schohaus, W. (1930). *Schatten über der Schule*. Zürich: Schweizer-Spiegel-Verlag.

Geschlecht im Schulalltag – Plädoyer für eine Entdramatisierung

Hannelore Faulstich-Wieland

Vortrag am Symposium „Schule wohin? Schulentwicklung im 21. Jahrhundert“ des Forschungsbereichs Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.

Einleitung

Gender Mainstreaming lautet das neue Stichwort für alle Politikbereiche – nicht nur in der Europäischen Union. Als Politikstrategie bedeutet dies, jede Massnahme auf ihre potenzielle Wirkung für beide Geschlechter zu durchforsten und nur dann zu realisieren, wenn sie zur Gleichstellung der Geschlechter beiträgt. Schulentwicklung ist eine Form von Bildungspolitik, so dass die Bedeutung von Geschlecht eine relevante Frage für mögliche Veränderungen ist.

Zunächst sollte man prüfen, inwieweit Ungleichheiten zwischen den Geschlechtern im Bildungssystem vorliegen. Dies will ich in einem ersten Schritt an der Teilhabe von Frauen an höheren Bildungsgängen sowie an den Leistungsdifferenzen zwischen Mädchen und Jungen in der Lesekompetenz, in Mathematik und in den Naturwissenschaften tun. Es wird sich zeigen, dass es in der Schweiz, ähnlich wie in Deutschland, sehr wohl Ungleichheiten gibt, diese jedoch nicht primär zu Lasten von Mädchen gehen.

Schulentwicklung dient dazu, erkannte Missstände oder erwünschte Veränderungen in eine Prozessplanung umzusetzen, um eine verbesserte Schulqualität zu erreichen. Der Abbau von Geschlechterungleichheiten ist im Sinne des *Gender Mainstreaming* ein wichtiges Kriterium für Schulqualität. Es gibt auch durchaus einige Schulen, die geschlechterbezogene Aktivitäten explizit als einen Schwerpunkt ihrer Arbeit ausweisen. Als Stichworte dafür findet man: geschlechterbewusste oder geschlechtergerechte Schule, Mädchen- und Jungenförderung, genderorientierte Pädagogik oder auch reflexive Koedukation.

Barbara Koch-Priewe (2002) hat in einem Sammelband zur „Mädchen- und Jungenförderung“ Beispiele für Schulprogramme von „geschlechterbewussten“ Schulen – so der Untertitel – zusammengestellt.

Auch in der Schweiz gibt es Richtlinien, wie eine „gleichwertige Förderung von Mädchen und Jungen“ in der Schule zu gewährleisten ist (vgl. z.B. Ryter & Schafroth, 2001).

In einem zweiten Schritt will ich deshalb kurz skizzieren, welche Ansätze sich unter diesen Stichworten finden. Es zeigt sich, dass viele Massnahmen nach wie vor auf getrennte Angebote für Mädchen bzw. für Jungen setzen und eine Dramatisierung – verstanden im Sinne Erving Goffmans (Goffman, 1996) als eine explizite Fokussierung oder Thematisierung von Geschlecht – darstellen.

Ausgesprochen oder unausgesprochen liegen diesen Massnahmen Gendertheorien zugrunde, die von grundlegenden Differenzen zwischen den Geschlechtern und einem hierarchischen Verhältnis ausgehen. Sie stehen damit in der Gefahr, Stereotypen festzuschreiben.

Gegen ein solches Verständnis soll in einem dritten Schritt eine konstruktivistische Position skizziert werden, nach der Geschlecht, und damit Geschlechterdifferenzen, ganz wesentlich durch interaktionelle Praktiken aller Beteiligten immer wieder hergestellt werden. Ausgehend von einem solchen Ansatz des *doing gender* soll viertens das Geschehen in der Schule etwas genauer beleuchtet werden. An einigen Massnahmen zur Mädchen- bzw. Jungenförderung wird aufzuzeigen sein, wie durch die Dramatisierung von Geschlecht Paradoxien entstehen, die den Absichten der Pädagoginnen und Pädagogen zuwiderlaufen.

Abschliessend sollen einige Konsequenzen für den Umgang mit Geschlecht im Schulalltag gezogen werden. Diese verstehe ich als Entdramatisierung von Geschlecht – oder eigentlich richtiger, als die Herstellung einer Balance von Dramatisierung und Entdramatisierung.

1. Ungleichheiten im Schweizer Bildungssystem

Eine Auswertung der Daten des Bundesamtes für Statistik von May Lévy und Elisabeth Pastor zur „Stellung der Frauen in Wissenschaft und Technologie in der Schweiz“ zeigt eindrücklich ein stetiges Anwachsen der Teilhabe von Frauen an universitärer Bildung. Seit 1994 liegt der Anteil der jungen Frauen mit Matur/Abitur über 50%, 2002 erreichten sie auch bei den Erstsemestern die 50%-Marke (Lévy & Pastor, 2005, S. 5). 2001 betrug die Maturitätsquote bei den Frauen 16,8% bezogen auf die 19jährige Wohnbevölkerung, bei den Männern mit 14,6% mehr als zwei Prozentpunkte weniger (www.zahlenspiegel.ch). Ähnlich wie in Deutschland sind die Fachwahlen in den Schulen wie an den Universitäten bei den Geschlechtern sehr ungleich – mit einem klaren Überhang der Frauen in den Geistes- und Sozialwissenschaften und einem ebenso klaren Überhang der Männer in den Naturwissenschaften und der Technik. Auch die PISA-Ergebnisse (Deutsches PISA-Konsortium, 2001, S. 252) zeigen für die 15-Jährigen im Jahr 2000 signifikant bessere Mathematikleistungen bei den Jungen. In den Naturwissenschaften schneiden die Jungen ebenfalls besser ab, wenngleich nicht signifikant. In den Leseleistungen allerdings sind die Mädchen deutlich und signifikant besser. Hier ist das geradezu erschreckende Ergebnis, dass weit mehr als ein Fünftel der Jungen auf oder unter der niedrigsten Kompetenzstufe liegen. Über ein Viertel der Jungen gehören hinsichtlich der Fähigkeiten „reflektieren und beurteilen“ zur Risikogruppe, nur etwa 6% sind hier sehr gut (Moser, 2001, S. 22). Diese wenigen Daten mögen bereits genügen, um zu zeigen, dass wir Geschlechterdifferenzen finden, die es zu verändern gilt, weil sie dem Qualitätskriterium der bestmöglichen Förderung aller widersprechen.

Die Teilhabe von Frauen an der universitären Bildung wächst stetig.

Die Fachwahlen sowie die Leistungen unterscheiden sich nach den Geschlechtern.

2. Schulprogramme zur Mädchen- und Jungenförderung: Dramatisierungen durch getrennte Angebote

Geschlechterbewusste Arbeit zeichnet sich vor allem durch geschlechtergetrennte Angebote aus und hat ihren Ausgangspunkt in der Mädchenarbeit.

Der Sammelband von Barbara Koch-Priewe (2002) beinhaltet dreizehn Beispiele, wie in Schulen geschlechterbewusste Arbeit in die Schulprogramme bzw. in die Schulentwicklung eingeflossen ist. Mit wenigen Ausnahmen handelt es sich dabei um die Realisierung von geschlechtergetrennten Angeboten, insbesondere im Informatikbereich, in Berufswahlkursen bzw. Berufsorientierungsprojekten oder im Sexualkundeunterricht. Zum Teil werden Mädchentage veranstaltet oder ein Mädchenraum bereitgestellt. Für alle Schulen gilt, dass die geschlechtsbewussten Ansätze ihren Ausgangspunkt in der Arbeit mit Mädchen und in einer mädchenparteilichen Sicht haben. Zwar ist in mehreren Schulen inzwischen die Erkenntnis gewachsen, dass es auch nötig sei, parallele Jungenarbeit anzubieten. Konzepte dafür werden jedoch kaum entfaltet, vielmehr liest man in mehreren Beiträgen, dass die Jungenarbeit nicht recht funktioniere. Als Ursache dafür wird in der Regel das Fehlen engagierter Kollegen ausgemacht, verbunden mit der Behauptung, Frauen könnten keine angemessene Jungenarbeit durchführen.

Als Grund für die Geschlechtertrennung werden häufig fundamentale Differenzen zwischen den Geschlechtern angenommen.

Nicht immer wird eine Begründung für die Geschlechtertrennung gegeben. So heisst es beispielsweise im Bericht von Hans-Peter Neumann über die Gesamtschule des Ennepe-Ruhr-Kreises in Sprockhövel, der getrennte Informatikunterricht würde „geschlechtsspezifisch durchgeführt“ (Neumann, 2002, S. 122). Die zugrunde liegende Annahme ist, es gäbe fundamentale Differenzen zwischen den Geschlechtern. In Bezug auf den Umgang mit Computern und damit für den Informatikunterricht gelten Mädchen entweder als weniger oder an anderen Anwendungen interessiert oder auch als weniger erfahren als Jungen. Empirisch ist es in vielen Bereichen ohne weiteres möglich, solche Differenzen zu finden. Man kann z.B. die eben schon angesprochenen signifikanten Leistungsunterschiede in Mathematik aus der PISA-Studie heranziehen.

Durch diese Dramatisierung der Differenz findet eine Dichotomisierung statt.

Als Problem für die praktische Umsetzung stellt sich jedoch in solchen Dramatisierungen der Differenz die mit ihnen einhergehende Dichotomisierung: *die* Mädchen haben keine Erfahrung, kein Interesse oder sind nur über Anwendungsbeispiele aus dem Alltag oder aus der Medizin zu interessieren; *die* Jungen dagegen haben Erfahrung, Interesse oder sind vor allem über Abstraktionen zu begeistern. Tatsächlich finden wir aber deutliche Überschneidungen zwischen den Gruppen der Mädchen und der Jungen, d.h. es gibt Mädchen mit viel Erfahrung, grossem Interesse und Begeisterung an mathematischen Zugängen und es gibt Jungen mit wenig Erfahrung, geringem Interesse und Lust auf konkrete Anwendungen des Computers. Ein „geschlechtsspezifischer“ Unterricht blendet möglicherweise gerade diese Schülerinnen und Schüler aus. Oder er macht sie besonders sichtbar als „geschlechtsuntypisch“ und verweist sie gerade damit zurück auf die Stereotype.

Eine vor allem in den Anfängen der Mädchenarbeit, aber auch in einigen der aktuellen Beschreibungen vorzufindende Begründung, warum es getrennten Fachunterricht geben sollte, bezieht sich nicht allein auf Differenzen, sondern auf Geschlechterhierarchien:

So schreibt die Lehrerin Brigitte Klostermann: „Im Schuljahr 98/99 stellte ich dann in einer 8. Klasse Folgendes fest: Die gesamte Klasse hatte offensichtlich Sexualkundeunterricht gehabt. Die Jungen kamen ‚gestärkt‘ mit unangenehm sexualisierter Sprache, die bis zur Diskriminierung der Mädchen ging, in meinen Kunstunterricht. Selbst die sonst starken Mädchen schwiegen und die ganz stillen saßen mit gesenkten Köpfen an den Tischen. Nach Gesprächen mit Kolleginnen haben wir beschlossen: Das muss sich ändern!“ Klostermann nimmt dann zufällig an einer Biologiefachkonferenz teil und thematisiert ihre Beobachtungen. Und: „In dieser Teilkonferenz wurde für die 8. Klasse mit den KlassenlehrerInnen sowie den BiologielehrerInnen verbindlich beschlossen, Jungen und Mädchen während des ‚Sexualkundeunterrichts‘ getrennt unterrichten zu lassen.“ (Klostermann, 2002, S. 59). Karin Michel begründet die Einrichtung getrennter Informatikkurse an der Lise-Meitner-Gesamtschule in Köln mit der Notwendigkeit, „darauf zu achten, dass die Mädchen in diesem Bereich nicht an die Seite gedrängt werden.“ (Michel, 2002, S. 112).

Diesen Begründungen liegt vor allem die Annahme zugrunde, Jungen dominierten das Unterrichtsgeschehen und Mädchen könnten sich in der koedukativen Situation dagegen nicht wehren. Koedukativer Unterricht würde die Jungen in ihrem Selbstbewusstsein stärken, den Mädchen aber sogar bereits vorhandenes Selbstvertrauen zerstören. Ein Unterricht ohne Jungen böte ihnen folglich die Chance, ihre Fragen und Interessen zu realisieren. Auch hier haben wir es mit einer pauschalen Dichotomisierung zu tun, welche die Jungen unter Generalverdacht stellt und für die Mädchen Protektion fordert.

Man kann also als erstes Zwischenfazit festhalten, dass eine Reihe von Versuchen durch ihr Ansetzen an Geschlechterdifferenzen eine Dramatisierung von Geschlecht vornehmen. Genderorientierte Arbeit in der Schule in dieser Form zu realisieren steht aber in der Gefahr, Geschlechterstereotype zu verfestigen statt sie abzubauen. Dies soll im Weiteren gezeigt werden. Dazu will ich vorab eine Gendertheorie skizzieren, die von der sozialen Konstruktion von Geschlecht ausgeht und damit Ansatzpunkte für Veränderungen bietet.

3. Doing gender

Alltägliches Verhalten unterliegt einem ständigen Bewertungsprozess, durch den Interaktionspartner wechselseitig Einfluss aufeinander nehmen und die soziale Ordnung ebenso wie die individuelle Entwicklung produzieren und reproduzieren. Erving Goffman hat solche Interaktionsrituale minutiös beobachtet und beschrieben (Goffman, 1986). Candace West und Don Zimmerman, die als erste vom *doing gender*, nämlich der Herstellung und ständigen interaktiven Reproduktion von geschlechtsangemessenem Verhalten, gesprochen haben, sehen in der Bewertung – also in der Frage nach der Geschlechtsangemessenheit – das entscheidende Moment. Allerdings geht es durchaus nicht darum, sich normativ korrekt zu verhalten – so wie es sich für eine Frau oder einen Mann idealerweise „gehört“, sondern das Verhalten eines Menschen *kann* immer vor der Folie der Geschlechtszugehörigkeit beurteilt werden. *Doing gender* „is to engage in behavior at the risk of gender assessment“ (West & Zimmerman, 1991, S. 25).

Zentral an diesem Ansatz ist die Annahme, dass nicht eine biologische oder natürliche Anlage das Verhalten steuert, sondern die Interaktionen und der soziale Kontext entscheidend dafür sind, ob sich eine Person als „weiblich“ oder „männlich“ darstellt und/oder so wahrgenommen wird. Der Glaube an eine natürliche biologische Steue-

Eine weitere Begründung für getrennten Fachunterricht bezieht sich auf Geschlechterhierarchien.

Die pauschale Dichotomisierung stellt Jungen unter Generalverdacht und fordert für Mädchen Protektion.

Es besteht die Gefahr, dass durch die Dramatisierung von Geschlecht die Geschlechterstereotype verfestigt, statt abgebaut werden.

Doing gender gilt als die Herstellung und ständige interaktive Reproduktion von geschlechtsangemessenem Verhalten.

nung des Geschlechts – der es uns so schwer macht, *doing gender* als interaktionelle Leistung zu begreifen – wird durch drei „axiomatische Basisannahmen“ (Hirschauer, 1989, 1993, 1994) gespeist:

Doing gender als interaktionelle Leistungen zu begreifen, wird durch drei „axiomatische Basisannahmen“ erschwert.

- Die Annahme der Konstanz: Wir unterstellen, dass ein Mensch, der uns heute als Frau begegnet, nicht morgen als Mann auftritt, wir gehen also von einer lebenslangen Gültigkeit der Geschlechtszugehörigkeit aus.
- Die Annahme einer Naturhaftigkeit von Geschlecht: Geschlecht wird an körperlichen Merkmalen festgemacht, insbesondere an den äusseren Geschlechtsmerkmalen. Dabei sehen wir davon ab, dass diese Merkmale in der alltäglichen Interaktion normalerweise nur bedingt erkennbar sind.
- Die Annahme der Dichotomizität: Wir kennen nur eine polare Zugehörigkeit, man ist entweder weiblich oder männlich, jedoch nichts dazwischen oder jenseits davon.

Wir kategorisieren nach dem „if-can“-Test.

Diese Annahmen erleichtern uns die Darstellung der Geschlechtszugehörigkeit, denn im alltäglichen Umgang vertrauen wir der „moral certainty“ (West & Zimmerman, 1991, S. 19) und kategorisieren Menschen nach dem Geschlecht gemäss einem „if-can“-Test: „This test stipulates that *if people can be seen as members of relevant categories, then categorize them that way*“ (ebd., S. 20). Solange wir nicht Grund für ernsthafte Zweifel haben, genügt der Augenschein der ersten Zuordnung zum weiblichen oder männlichen Geschlecht, um jemand als Frau oder Mann zu akzeptieren. Wir sorgen jedoch auch dafür, uns so zu geben, dass ein Erkennen unserer Geschlechtszugehörigkeit „sofort“ möglich ist und wir erwarten gleiches von unserem Gegenüber. In der Interaktion halten wir diese Balance, wir benehmen uns also im Rahmen der Bewertungsmaßstäbe, die unser Handeln als angemessen erscheinen lassen. Unstimmigkeiten treten dabei „natürlich“ auf, weil die Verhaltensspielräume durchaus gross und eben nicht strikt vorgegeben sind. Wir verfügen aber über die Fähigkeiten, mit diesen Spielräumen gekonnt umzugehen. Wir können auch sehr wohl bewusst gegen Zuschreibungen verstossen, bewegen uns dann aber auf dem schmalen Grat zwischen der Anerkennung als „aussergewöhnliche Frau“ und der Diskriminierung als „Mann-Weib“.

In unserem alltäglichen Verhalten aktualisieren wir die Geschlechterdifferenz ständig, trotzdem verändern sich die Formen von *gender*.

Man kann also festhalten: In unserem alltäglichen Verhalten präsentieren und aktualisieren wir die Geschlechterdifferenz. Insofern kann man die von West & Zimmerman aufgestellte Omnirelevanz-Annahme bestätigen: Geschlecht spielt immer eine Rolle, ob wir das wollen oder nicht. Trotzdem verändern sich die kulturellen Formen von *gender*. So gibt es je nach historischer und gesellschaftlicher Situation höchst unterschiedliche Bestimmungen dessen, was *accountable* – geschlechtsangemessen – ist (vgl. Robak, 1988; Wetterer, 1995).

„Institutionelle Reflexivität“ sorgt zusätzlich für eine überdauernde Geschlechterzuschreibung.

Neben den genannten „Basisannahmen“ gibt es weitere Aspekte des Arrangements der Geschlechter, die für überdauernde Zuschreibungen sorgen und *doing gender* erleichtern. Erving Goffman hat sie als „institutionelle Reflexivität“ bezeichnet. Die Etablierung paralleler Organisationen stellt beispielsweise eine solche institutionelle Reflexivität dar, welche die Herstellung von Differenzen besonders gut erlaubt:

„Wie bei den parallelen Organisationen, die sich an anderen binären sozialen Klassifizierungen festmachen – Schwarze–Weisse, Erwachsene–Kinder, Offiziere–Rekruten und so weiter –, bietet die auf dem Geschlecht basierende parallele Organisation einen leicht handhabbaren Ausgangspunkt für die Etablierung einer unterschiedlichen Behandlung der Geschlechter“ (Goffman, 1994, S. 114).

Goffman bringt als Beispiel aus dem schulischen Raum die Praxis, nach der Kinder sich nach Mädchen und Jungen getrennt aufstellen müssen, bevor sie die Klassenräume betreten dürfen.

„Wenn Kinder einmal dazu gebracht wurden, nach Geschlechtern getrennte Schlangen zu bilden, dann kann auch leicht veranlasst werden, dass die weibliche vor der männlichen Schlange ins Haus geht, vermutlich um dem ‚zarteren‘ Geschlecht beim Verlassen der rauen Aussenwelt den Vortritt zu geben und umso beiden Geschlechtern eine kleine Lektion über die korrekte Rücksichtnahme auf das soziale Geschlecht zu erteilen“ (ebd.).

Über solche Dramatisierungen der Differenz lassen sich folglich unterlegte Glaubensvorstellungen bzw. Stereotype reproduzieren. Wir haben es hier mit dem Wechselspiel von Strukturen und ihren Interpretationen zu tun.

Symbolischen Repräsentationen kommen dabei, aber auch unabhängig von institutionellen Reflexivitäten, eine wichtige Bedeutung für die Konstruktion von Geschlecht zu. Dinge und Tätigkeiten werden mit männlich oder weiblich assoziiert – beispielsweise runde Gegenstände oder warme Farben mit weiblich, eckige Gegenstände oder kalte Farben mit männlich. Entscheidend für Symbolisierungen von Geschlecht ist, dass alle Vergeschlechtlichungen mit jeweils binären und oppositionellen Charakterisierungen arbeiten. Was einzeln betrachtet eine willkürliche Einteilung darstellt, erhält durch die Einreihung in ein System homologer Gegensätze (hoch/tief, oben/unten, vorne/hinten, gerade/krumm, trocken/feucht usw.) eine objektiv und subjektiv notwendige Wahrnehmung. Pierre Bourdieu spricht von einem unerschöpflichen Spiel von Umschreibungen und Metaphern, die eine gegenseitige Stützung dieser Denkschemata bewirken. So entsteht eine „Nötigung durch Systematizität“, da eine quasi natürliche Bestätigung die Bedeutungen verdoppelt und verstärkt (Bourdieu, 1997a; 1997b). Bourdieu fordert für Veränderungen folglich auch eine symbolische Revolution, d.h. nicht nur eine Veränderung der Ordnung der Dinge, der materiellen Strukturen, sondern auch einen mentalen Umbruch, „eine Transformation der Kategorien der Wahrnehmung, die uns dazu bringen, dass wir bei der Perpetuierung der bestehenden Gesellschaftsordnung mitspielen“ (Bourdieu, 1997b, S. 98). Eine Entkoppelung der Mathematik und der Naturwissenschaften mit *hart* und damit mit *männlich* wäre z.B. ein Schritt in diese Richtung.

Die soziale Konstruktion von Geschlecht verweist also zum einen auf die kulturelle und gesellschaftliche Gemachtheit dessen, was mit Geschlecht gemeint ist (vgl. Faulstich-Wieland, 2004a). *Doing gender* macht zugleich klar, dass jede und jeder an der Konstruktion beteiligt ist, sie interaktionell immer wieder reproduziert. Schulentwicklung muss diese komplexen Zusammenhänge von individuellem Verhalten, sozialen Interaktionen und institutionell-organisatorischen Massnahmen berücksichtigen, wenn Geschlechtergerechtigkeit erreicht werden soll.

4. Doing gender im heutigen Schulalltag

Ein solcher gendertheoretischer Ansatz erlaubt es, Massnahmen, die mit einer Dramatisierung von Geschlecht arbeiten, auf ihre Wirkungen für eine Veränderung des Geschlechterverhältnisses hin zu analysieren – wenngleich hier noch erheblicher Forschungsbedarf besteht, da die meisten Genderforschungen im Schulbereich sich an Differenzkonzepten orientieren.

Alle Vergeschlechtlichungen arbeiten mit binären und oppositionellen Charakterisierungen.

Nicht nur eine Veränderung der Ordnung der Dinge, sondern auch ein mentaler Umbruch ist gefordert.

Schulentwicklung muss komplexe Zusammenhänge berücksichtigen, um Geschlechtergerechtigkeit zu erreichen.

Ein geschlechtersensibler Anspruch führt nicht zwingend zu mehr Geschlechtergerechtigkeit.

**Beispiele für Dramatisierung von
Geschlecht im Unterricht**

An einigen Beispielen aus unserem gerade abgeschlossenen Forschungsprojekt zur sozialen Konstruktion von Geschlecht (vgl. Faulstich-Wieland, Weber & Willems, 2004) möchte ich aufzeigen, wie ein geschlechtersensibler Anspruch seitens der Lehrkräfte nicht zu einem Mehr an Geschlechtergerechtigkeit führt, weil stereotype Annahmen von Geschlecht zugrunde gelegt werden. In dem Forschungsprojekt haben wir drei Gymnasialklassen über drei Schuljahre, nämlich das 7. bis 9. bzw. das 8. bis 10., ethnografisch begleitet. Eine dieser Klassen war mit zwei Dritteln Jungen und einem Drittel Mädchen jungendominant, eine zweite mit dem umgekehrten Verhältnis mädchendominant, die dritte ausgewogen zusammengesetzt.

Wir finden in den Unterrichtsbeobachtungen einige Interaktionen, in denen Lehrkräfte explizit Differenzen zwischen den Geschlechtern ansprechen und damit zugleich vermitteln, was angemessen ist.

„In dem Tumult sagt Frau Ferreira zu Kurt mit spassigem Unterton, nachdem dieser verkündet, dass er die Hausaufgaben vergessen habe: ‚Du versaut das Bild!‘ Darauf kommt jedoch keine Reaktion, ich denke die Bemerkung geht unter bzw. wird nicht verstanden. Ich denke mir, dass sie darauf anspielt, dass Kurt der einzige Junge ist, der die Hausaufgaben nicht gemacht hat.

Diese Vermutung bestätigt sich kurz danach. Frau Ferreira bittet darum, dass sich alle melden sollen, die ihre Hausaufgaben nicht gemacht haben. Sieben bis acht Finger gehen in die Höhe. ‚So, jetzt sagt mir Henning, warum Kurt das Bild versaut‘, fordert sie Henning lachend auf. Dieser antwortet flugs, dass Kurt der einzige Junge sei. ‚Genau‘, erwidert die Lehrerin. ‚Sonst habe man das oft genau andersherum, dass die Mädchen die Hausaufgaben hätten und die Jungen nicht, da sie es verschlampt, vergessen, oder so, hätten. Sie lächelt bei dieser Ausführung ihrer Beobachtungen“ (Ad01127v).

Im Deutschunterricht der 10. Klasse haben ein Viertel keine Hausaufgaben, darunter nur ein Junge. Die Lehrerin dramatisiert Geschlecht, indem sie mit scherzhaftem Ton herausarbeiten lässt, wie ungewöhnlich es sei, dass fast alle Jungen die Hausaufgaben gemacht hätten. Normalerweise – und hier bestärkt die Lehrerin, was als geschlechtsangemessen gilt – wären die Mädchen ordentlich, während die Jungen schlampig und vergesslich seien.

Als Reaktion auf Forschungsergebnisse wird empfohlen, Mädchen und Jungen abwechselnd aufzurufen, um ungleiche Aufmerksamkeitsverteilungen zugunsten der Jungen zu verhindern. Diese Regel galt auch in den von uns beobachteten Klassen. Da die Zusammensetzung jedoch höchst ungleich nach Geschlecht war, bewirkt eine solche Massnahme verstärkte Aufmerksamkeit für die jeweilige Minderheit. Die Jugendlichen gingen relativ souverän damit um, indem sie die Lehrkräfte immer dann an die Regel erinnerten, wenn sie gerne drankommen wollten oder einer Mitschülerin oder einem Mitschüler dazu verhelfen wollten:

„Nun wird weitergelesen. Jolanda meldet sich und möchte lesen. Der Lehrer fragt, wer denn vorher gelesen habe und als er ‚Nina‘ hört, sagt er auf Jolandas Frage: ‚Dann kommt jetzt ein Junge dran.‘ Er ruft Joe auf.

Wenig später möchte Erik lesen, L lehnt nun ab und sagt: ‚Nee, jetzt liest Jolanda.‘ Als Jolanda fertig ist, sagt L: ‚Almut liest!‘ Nun rufen einige Ss empört über den ‚Regelverstoss‘: ‚Nee, jetzt kommt ’nen Junge!‘ L grinst, er fühlt sich wohl ertappt und sagt: ‚Stimmt.‘ Es liest Erik“ (Bb81027n).

Es bleibt bei dieser Praxis jedoch als mögliches Problem, dass Antworten primär als

Mädchen- bzw. Jungenantworten wahrgenommen werden, weil sie häufig explizit so von der Lehrkraft eingeführt werden:

„Lehrerin: ‚Was meint ihr, Mädels, es sind nur drei, die können aber auch mal schätzen‘.
Schülerin (Nicole?): ‚Zwei Millionen?‘
L sagt die richtige Lösung und schreibt an: 4,5 – 5 Millionen pro Kubikmillimeter“
(Ab80630d).

Nicoles Antwort auf die Frage nach der Anzahl von roten Blutkörpern in einem Kubikmillimeter wird als „Mädchenantwort“ aufgerufen. Sie wird zwar nicht als solche kommentiert, verbleibt aber möglicherweise in Erinnerung der Lehrerin nicht nur als Nicoles Fehler, sondern als falsche Mädchenantwort.

Dramatisierungen von Geschlecht bieten den Jungen die Möglichkeit, sich gegen Massnahmen zu wehren oder sich als ungerecht behandelt zu fühlen. Es bleibt ihnen nicht verborgen, dass der Anspruch von Geschlechtergerechtigkeit häufig mit negativen Charakterisierungen des Jungenverhaltens einhergeht. Während in Schulprogrammen oder Projekten als Ziele für Mädchen genannt wird, dass sie ein höheres Selbstbewusstsein entwickeln sollen, wird von den Jungen erwartet, sich weniger dominant zu verhalten. D.h. die Zielsetzungen für Mädchen sind positiv formuliert und besetzt, die für Jungen dagegen gehen mit Kritik an ihnen einher. Folgendes Beispiel zeigt, wie Mädchenparteilichkeit durch die Lehrkräfte Widerstand bei den Jungen produziert.

Die jugenddominante Klasse sollte im 8. Jahrgang über einige Wochen eine Einführung in die Arbeit mit dem Computer erhalten. Der Unterricht sollte mit der Hälfte der Klasse durchgeführt werden, während die andere Hälfte zur gleichen Zeit Deutschunterricht erhielt. Kurz vor Beginn der Unterrichtseinheit entstand folgende Protokollaufzeichnung:

„Die eine Gruppe soll für zwei Stunden in der Woche ein Vierteljahr lang informationstechnische Grundbildung haben und die andere Deutsch im Wechsel. Achmed fragt, wie die Gruppen eingeteilt werden. Nathalie wirft ein: ‚Nach Alphabet‘. Sven meint: ‚Jungen und Mädchen getrennt‘. Frau Böttcher geht nicht darauf ein und führt aus: ‚Das haben wir schon beschlossen‘ (Sie meint, sie und die Klassenlehrerin, Frau Storm, hätten die Zusammensetzung schon festgelegt). Sie erläutert dann ausführlich den Wechselmodus etc., ohne auf die Gruppenzusammensetzung einzugehen. Sie sagt, es ginge um das Thema Sexualität. Siegfried signalisiert Begeisterung. Sie sagt: ‚Moment! Es geht um sexuellen Missbrauch‘. Eine Weile später fragt Achmed noch mal danach, wie die Gruppenaufteilung aussieht. Die Lehrerin sagt: ‚Die Mädchen wollen zusammenbleiben und das werden sie auch‘. Sven meint süffisant: ‚Die Jungen auch‘. Die Lehrerin sagt entschieden: ‚Nee‘. und stützt sich wohl darauf, dass dies bei der hohen Anzahl der Jungen nicht möglich sei. Sören: ‚Veith und Joachim kommen noch zu den Mädchen‘. Sven sagt schimpfend etwas vor sich hin, von dem ich ‚Männerfeindlichkeit‘ und ‚Rassismus‘ aufschnappen kann.“ (Ad80902d).

Die Schülerinnen und Schüler wissen bereits, dass der Unterricht in Gruppen stattfinden wird. Auf die Frage von Achmed, wie die Gruppen aussehen werden, schlägt Nathalie vor, nach dem Alphabet zu gehen. Sven dagegen glaubt, Mädchen und Jungen würden getrennt. Da die Schule in einer Reihe von Projekten sich der Geschlechterfrage u.a. durch Geschlechtertrennung annimmt, kommt Svens Vermutung wahrscheinlich aus der Kenntnis dieser Praxis. Die Lehrerin beantwortet zunächst weder Achmeds Frage noch geht sie auf die Vorschläge und Mutmassungen ein, erklärt aber, der Teilungsmodus sei bereits beschlossene Sache. Stattdessen gibt sie bekannt, dass in

Antworten einzelner Kinder werden primär als Mädchen- bzw. Jungenantworten wahrgenommen.

Zielsetzungen für Mädchen sind positiv formuliert und besetzt, diejenigen für Jungen gehen mit Kritik an ihnen einher.

Geschlecht wird oft als Kriterium der Gruppeneinteilung herangezogen.

Den Mädchen wird das Bedürfnis nach Zusammenbleiben unterstellt.

dem parallelen Deutschunterricht das Thema Sexualität behandelt würde. Siegfrieds Begeisterung darüber wird von ihr sofort gedämpft, in dem sie präzisiert, es ginge um sexuellen Missbrauch. Warum sie dies unter das Thema Sexualität subsumiert, bleibt unklar. Nachdem Achmed nach einer Weile erneut darauf insistiert, zu wissen, wie die Gruppen aussehen werden, bestätigt die Lehrerin, was Sven vermutet hatte, nämlich dass es zu einer weitgehenden Geschlechtertrennung kommen würde, indem eine Jungengruppe und eine gemischte Gruppe entstehen sollten. Sie gibt an, dies sei der Wunsch der Mädchen: „Die Mädchen wollen zusammenbleiben und das werden sie auch.“ Angesichts von Nathalies Vorschlag, die Gruppen nach Alphabet aufzuteilen, kann die Geschlechtertrennung wohl nicht der artikulierte Wunsch der Mädchen gewesen sein, sondern gibt ein ihnen von den Lehrerinnen *unterstelltes* Bedürfnis wieder. Die Jungen reagieren auf diese Begründung mit Widerstand und fordern nun ebenfalls, zusammen bleiben zu wollen – was angesichts ihrer Überzahl nicht möglich ist. In der Konsequenz grenzen sie zwei Aussenseiter – Veith und Joachim – durch symbolische Verweiblichung aus, markieren das Ganze zugleich als „männerfeindlich“ bzw. „rassistisch“, wodurch sie ihren hegemonialen Anspruch aufrecht zu erhalten versuchen.

Viele Mädchen leisten gegen die Protektion Widerstand.

Auch viele Mädchen leisten gegen die Protektion, die ihnen durch eine Geschlechtertrennung angedient wird, Widerstand. Dies zeigt sich insbesondere in einem Forschungsprojekt, in dem wir in einem österreichischen Gymnasium einen 7. Jahrgang über zwei Jahre begleiten (vgl. Faulstich-Wieland, 2004b, 2004c). Von den drei Klassen dieses Jahrgangs wurde eine als Mädchenklasse geführt. Auch in den koedukativen Klassen findet zum Teil geschlechtergetrennter Unterricht beispielsweise im Werken statt. Von einer Schülerin wird dies im Interview massiv kritisiert:

Viola: „...ja und sie haben nicht gemeint jetzt, ja bildet zwei Gruppen oder so, sondern, ja, wir trennen jetzt Buben Mädchen. Ohne Grund, eigentlich, nein! Sie haben gemeint (leicht verächtlich), weil dann vielleicht Mädchen vor Buben andere Themen nehmen würden und sich manche Themen nicht trauen würden, aber ich mein – stimmt nicht wirklich. Also für mich zumindest nicht.“ (Interview mit Viola, Sabine, Lydia, Falco März 2005).

Viola demonstriert ihr Unverständnis einer geschlechtergetrennten Einteilung, für die sie keinen einleuchtenden Grund erkennen kann. Die Begründung, die wohl von den Lehrkräften gegeben wurde, Mädchen würden sich im getrennten Unterricht mehr zutrauen, weist sie zurück. Sie habe keine Protektion nötig. Ähnlich argumentieren auch ihre Mitschülerinnen gegenüber der Einrichtung einer Mädchenklasse:

„M1: Also, ich bin froh, dass ich in keiner Mädchenklasse bin!

M2: Ja!

M1: Also, ich weiss nicht, weil ich mein, Buben sind manchmal schon ziemlich deppert, oder so, aber in unserem Alter kommen sie in die Pubertät (leises Lachen) und da ist es wohl auch besser, wenn man mit Buben umgehen kann, oder so?!

M2: Ja! (leises Zustimmung)

M1: Die Frage ist ja, es ist ziemlich langweilig in einer Mädchenklasse!

...

M1: Und der Vorteil an der Mädchenklasse, hab ich, glaube ich, gehört, ja, dass die Mädchen nicht so zurückhaltend sind und zu den Buben, dass diese kein Platz lassen, hab ich gehört, dass es deswegen ist!

Und das ist ja bei uns nicht so, dass die Buben die Chefs sind (verhaltenes Lachen, leise Zustimmung aus dem Hintergrund).“ (Interview mit Nicole, Loretta, Mareike (5A) und Manja (3C) im März 2005).

Zunächst äussern die Schülerinnen sich froh darüber, nicht in die Mädchenklasse gekommen zu sein. Sie sehen die Jungen zwar durchaus kritisch, halten es aber gerade in der Pubertät für wichtig zu lernen, wie man mit ihnen klar kommen kann. Schliesslich mutmassen die Schülerinnen, was der Grund für die Einrichtung einer Mädchenklasse war. Auch sie vermuten, dass Protektionismus dahinter stehe und lehnen diesen für sich ab. Sie reklamieren für sich sehr wohl eine selbstbewusst gleichberechtigte Position gegenüber den Jungen.

Administrative Trennungen, die nicht explizit auf die Bearbeitung des Geschlechterthemas selbst zielen – wie z.B. bei Fragen der Sexualität – oder die als Projekte freiwillig und nach Interesse gewählt werden können, sondern die den „regulären“ Unterricht betreffen, stossen überwiegend bei Mädchen wie Jungen auf Ablehnung. Birgit Klein-Uerlings, Moderatorin für Koedukation an Schulen in Nordrhein-Westfalen, sieht den Grund dafür in der mit solchen Massnahmen einhergehenden Missachtung der Jugendlichen als Subjekte:

„Während in früheren Jahren die Thematisierung der Geschlechterverhältnisse von Schülerinnen (und Schülern), insbesondere von Studentinnen (und Studenten) eingefordert wurde, wird sie jetzt durch entsprechende Erlasse und Curricula oktroyiert. Vor allem Schülerinnen reagieren nicht selten mit Bemerkungen wie: ‚Das ist einfach nicht mehr unser Thema. Wir haben kein Problem mehr damit.‘ Schülerinnen fühlen sich offensichtlich in einen Objektstatus verbannt – sie erleben sich als Objekte der Forschung und, noch schlimmer, als Objekte pädagogischer Bemühung“ (Klein-Uerlings, 2002, S. 174f.).

Administrative Trennungen der Geschlechter stossen bei Mädchen wie Jungen auf Ablehnung.

5. Gegen die Omnipräsenz von Geschlecht – Plädoyer für Entdramatisierung

Von einem konstruktivistischen Standpunkt aus gilt es, den aktiven Part der Jugendlichen beim *doing gender* zu beachten. Es gibt viele Arten, wie Schülerinnen und Schüler von Dramatisierungen des Geschlechts profitieren. Bisher gilt, dass hegemoniale Männlichkeit als Ressource für viele Jungen nach wie vor ein erfolgreiches Konzept ihrer Absichten darstellt. Das Hausaufgabenbeispiel verweist ja darauf, dass ihnen leicht nachgesehen wird, wenn sie schlampig oder vergesslich sind. Auch die Mädchen profitieren noch immer von Weiblichkeitsinszenierungen, in denen sie sich als unsicher und schützenswert darstellen.

Nicht alle Handlungen müssen jedoch als eine Inszenierung von Geschlecht begriffen werden. Man kann andere Formen der sozialen Konstruktion benennen, beispielsweise das *doing adult* als Versuch, nicht kindlich, sondern erwachsen zu wirken. Oder das *doing student*: Darunter verstehen wir Strategien von Kindern und Jugendlichen, strukturelle Bedingungen der Schule zu handhaben, die durch institutionalisierte Kontrolle gekennzeichnet sind. Dies können Solidarierungen gegen die Kontrolle der Lehrkraft sein – etwa das Vorsagen oder Abschreiben oder auch die Bestätigung bei Unsicherheit bezüglich einer richtigen Antwort. Rückzugsstrategien aus dem Unterricht, aber auch der Umgang mit Konkurrenz und Leistungsanforderungen sind Ausdruck von *doing student*.

Lenkt man den Blick auf solche anderen Formen des *doing differences*, wie Candace West und Sarah Fenstermaker (1995) sie bezeichnet haben, dann lässt sich die These der Omnirelevanz der Kategorie Geschlecht relativieren. Stattdessen zeigt sich ein Spannungsfeld zwischen den Polen einer Dramatisierung und einer Entdramatisie-

Es gilt, den aktiven Part der Jugendlichen beim *doing gender* zu beachten.

Der Blick auf andere Formen des *doing differences* relativiert die Omnirelevanz der Kategorie Geschlecht.

nung von Geschlecht. Gerade hinsichtlich des *doing student* lassen sich alle Varianten von Dramatisierung und Entdramatisierung zeigen. Es gibt einerseits Strategien, bei denen die Geschlechtszugehörigkeit eine Rolle spielt, weil Mädchen und Jungen auf je spezifische Strategien zurückgreifen oder die Strategien unterschiedlich handhaben. Andererseits finden sich aber auch gleiche Strategien bei beiden Geschlechtern oder gleiche Handhabungen: Jungen zeigen sich konkurrenzorientierter als Mädchen und legen ein Dominanzverhalten an den Tag, Mädchen zeigen mehr Strategien der Solidarität und Kooperation miteinander. Aber solche Stereotypen werden andererseits auch wieder durchbrochen, indem etwa die Ausgrenzung von „Strebern“ weder aktiv noch passiv an ein Geschlecht gebunden ist; oder die Schüler/innensolidarität über die Geschlechtergrenzen hinweg praktiziert wird. In solchen Verhaltensweisen tritt die soziale Kategorie Geschlecht in den Hintergrund.

Dramatisierungen der Differenzen erschweren durch die klare Unterscheidung der Geschlechter zum einen die Wahrnehmung der Differenzierungen innerhalb der Gendergruppen. Zum anderen erzwingen sie mindestens teilweise ein stereotypes *doing gender* durch die Schülerinnen und Schüler.

Die Forderung nach Entdramatisierungen kann nun allerdings nicht bedeuten, zur vermeintlichen Geschlechtsneutralität zurückzugehen. Jene Lehrkräfte, die glauben, keine Geschlechterunterschiede zu machen, alle gleich zu behandeln, sind sehr wohl in die alltäglichen *doing gender* - Prozesse involviert, ebenso wie die Schülerinnen und Schüler. Sie reproduzieren auf diese Weise durchaus die „normalen“ Geschlechterbilder. Pädagogisches Handeln wäre deshalb rückzubinden an Selbstreflexionen des eigenen *doing gender*. Solche Selbstreflexionen sind ohne Genderkompetenz – also ein Wissen um die strukturellen Ungleichheiten, um sie stützende Geschlechterstereotype, um die symbolischen Repräsentationen, die sich immer wieder selbst bestätigen (vgl. Rose, 2005) – nicht zu leisten. Gefragt ist dann aber eine Balance zwischen Dramatisierung und Entdramatisierung von Geschlecht. Für den pädagogischen Umgang mit Jungen wäre es wichtig, Schule derart zu gestalten, dass Jungen in ihrer individuellen Suche jenseits von Defizitorientierungen oder Remaskulinisierungen unterstützt werden. Für den pädagogischen Umgang mit Mädchen halten wir vor allem den Verzicht auf Protektionismus für wichtig, da dieser in den Aporien bisheriger Mädchenparteilichkeit verstrickt bleibt.

Wenn man Mädchen und Frauen, ebenso wie Jungen und Männer, als Subjekt ihrer Handlungen akzeptiert, dann geht dies nur durch ein Ernstnehmen ihrer Einstellungen und Wünsche und durch eine Auseinandersetzung damit. Gemeinsame Gestaltungen wären u.E. ein Weg in diese Richtung. Schulentwicklungsprozesse, die Schülerinnen und Schüler miteinbeziehen ohne diese primär als Repräsentanten ihres Geschlechts zu sehen, wäre ein Weg, der sich lohnen würde, um eine geschlechtergerechte Schumatmosphäre zu schaffen. Gegen die bisherigen dramatisierenden Massnahmen plädiere ich dabei für eine deutlich stärkere Entdramatisierung.

Genderkompetenz ist notwendig für selbstreflexives, pädagogisches Handeln, wobei eine Balance zwischen Dramatisierung und Entdramatisierung gefragt ist.



Prof. Dr. Hannelore Faulstich-Wieland
 Institut für Schulpädagogik und Pädagogische Psychologie
 Universität Hamburg
 Von Melle Park 8
 20146 Hamburg
 Faulstich-Wieland@erzwiss.uni-hamburg.de

Literatur

- Bourdieu, P. (1997a). Die männliche Herrschaft. In: I. Dölling & B. Kraus (Hrsg.), *Ein alltägliches Spiel: Geschlechtskonstruktionen in der sozialen Praxis* (S. 153-217). Frankfurt a. M.: Suhrkamp.
- Bourdieu, P. (1997b). Männliche Herrschaft revisited. In: *Feministische Studien*, 15 (2), 88-99.
- Deutsches PISA-Konsortium (Hrsg.) (2001). *PISA 2000*. Opladen.
- Faulstich-Wieland, H. (2004a). Doing gender: Konstruktivistische Beiträge. In: E. Glaser, D. Klika & A. Prengel (Hrsg.), *Handbuch Gender und Erziehungswissenschaft* (S. 175-191). Bad Heilbrunn: Klinkhardt.
- Faulstich-Wieland, H. (2004b). Das Geschlechterthema in einer österreichischen Schule. In: S. Buchen, C. Helfferich & M. S. Maier (Hrsg.), *Gender methodologisch. Empirische Forschung in der Informationsgesellschaft vor neuen Herausforderungen* (S. 231-247). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Faulstich-Wieland, H. (2004c). Wem nützt die Einrichtung einer Mädchenklasse? In: *Zeitschrift für Frauenforschung & Geschlechterstudien* 22 (4), 39-57.
- Faulstich-Wieland, H., Weber, M. & Willems, K. (2004). *Doing Gender im heutigen Schulalltag. Empirische Studien zur sozialen Konstruktion von Geschlecht in schulischen Interaktionen*. Weinheim: Juventa.
- Goffman, E. (1986). *Interaktionsrituale*. Frankfurt a. M.: Suhrkamp.
- Goffman, E. (1994). *Interaktion und Geschlecht*. Frankfurt a. M.: Campus.
- Goffman, E. (1996). *Wir alle spielen Theater. Die Selbstdarstellung im Alltag*. München: Piper.
- Hirschauer, S. (1989). Die interaktive Konstruktion von Geschlechtszugehörigkeit. *Zeitschrift für Soziologie*, 18 (2), 100-118.
- Hirschauer, S. (1995). Dekonstruktion und Rekonstruktion. Plädoyer für die Erforschung des Bekannten. *Feministische Studien*, 11, 59-66.
- Hirschauer, S. (1994). Die Soziale Fortpflanzung der Zweigeschlechtlichkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 46 (4), 668-692.
- Klein-Uerlings, B. (2002). Im Dschungel der Widerstände. Vom Umgang mit Schwierigkeiten bei der Verankerung von Jungen- und Mädchenarbeit. In B. Koch-Priewe (Hrsg.), *Schulprogramme zur Mädchen- und Jungenförderung. Die geschlechterbewusste Schule* (S. 170-181). Weinheim: Beltz.
- Klostermann, B. (2002). Ein Sexualmord als Anlass für geschlechterbewusste Schulentwicklung. Selbstbehauptung und Stärkung von Jungen und Mädchen. In B. Koch-Priewe (Hrsg.), *Schulprogramme zur Mädchen- und Jungenförderung. Die geschlechterbewusste Schule* (S. 58-68). Weinheim: Beltz.
- Koch-Priewe, B. (2002). (Hrsg.), *Schulprogramme zur Mädchen- und Jungenförderung. Die geschlechterbewusste Schule*. Weinheim: Beltz.
- Lévy, M., Pastor, E. (2003). Stellung der Frauen in Wissenschaft und Technologie in der Schweiz. *BFS aktuell*, 15, 5. (www.statistik.admin.ch/stat_ch/ber15/publ-st/pdf/frauen-st.pdf).
- Michel, K. (2002). Chancengleichheit von Mädchen und Jungen im Schulprogramm. Ein erster Erfolg ist an der Evaluation ablesbar. In B. Koch-Priewe (Hrsg.), *Schulprogramme zur Mädchen- und Jungenförderung. Die geschlechterbewusste Schule* (S. 108-120). Weinheim: Beltz.
- Moser, U. (2001). Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen. *Kurzfassung des nationalen Berichts PISA 2000*. BFS Neuchâtel 2001 (Hrsg.), S. 22 (www.statistik.admin.ch/stat_ch/ber15/pisa-pres/pisa-d.pdf).
- Neumann, H.-P. (2002). Geschlechterbewusste Schule als System von Unterricht, Ganztags, Projekten, Beratung und Erziehung. In B. Koch-Priewe (Hrsg.), *Schulprogramme zur Mädchen- und Jungenförderung. Die geschlechterbewusste Schule* (S. 121-150). Weinheim: Beltz.
- Robak, B. (1988). *Auf der Suche nach der weiblichen Facharbeit. Das Beispiel Schriftsetzerin*. Kassel: IAG GhK.
- Rose, L. (2005). Genderwissen und Genderforschung als Qualifizierungsimpuls für die Jugendarbeit. *deutsche jugend*, 51 (11), 467-475.
- Ryter, A. & Schafroth, K. (2001). *Gleichwertige Förderung von Mädchen und Knaben in der Volksschule im Kanton Zürich*. Bericht der Fachstelle für Gleichberechtigungsfragen des Kantons Zürich. Zürich 2001.
- West, C., & Fenstermaker, S. (1995). Doing Difference. *Gender & Society*, 9 (1), 8-37.
- West, C. & Zimmerman, D. H. (1991). Doing Gender. In: J. Lorber & S. A. Farrell (Hrsg.), *The Social Construction of Gender* (S. 13-37). Newbury Park/London: Sage.
- Wetterer, A. (1995). (Hrsg.), *Die soziale Konstruktion von Geschlecht in Professionalisierungsprozessen*. Frankfurt a. M.: Campus.
- www.zahlenspiegel.ch (Bundesamt für Statistik).

Zur Bedeutung von Evaluation für die Schulentwicklung

Eckhard Klieme

1. Schulqualität im Kontext neuer Steuerungsmodelle

Vor allem die innere Schulkultur ist verantwortlich für die Qualität von Schule.

Die Qualität von Schule und Unterricht ist im deutschsprachigen Raum seit langem Gegenstand der empirischen Forschung. Hierzu haben vor allem die Evaluationsstudien an Gesamtschulen beigetragen, die in den siebziger Jahren durch Helmut Fend durchgeführt wurden (vgl. dazu Fend, 1998). Als Ergebnis dieser empirischen Untersuchungen stellte Fend die These auf, dass weniger die Schulform, also strukturelle Merkmale, als die Qualität der Einzelschule, also vor allem die innere Schulkultur, verantwortlich sei für die Qualität von Schule. Nicht das System und die administrativen Vorgaben, sondern die Einzelschule mit ihrer „individuellen“ Werte-, Kooperations- und Führungskultur, ihrem Schulklima, ihren Partizipationsmöglichkeiten, ihren curricularen Schwerpunkten und spezifischen Unterrichtsangeboten gilt es in den Blick zu nehmen, wenn Kindern und Jugendlichen bessere Lernchancen in besseren Schulen gegeben werden sollen. Diese Erkenntnis war in den 1980er und 90er-Jahren Ausgangspunkt vielfältiger Schulentwicklungsbemühungen und Forschungsvorhaben. Die Grundthese wurde durch die quantitative Schulwirkungsforschung wie auch die qualitative Untersuchung besonders effektiver Schulen international gestützt (vgl. Scheerens & Bosker, 1997).

Entwicklung der Einzelschule kann nur indirekt beeinflusst werden.

Vertretern der Bildungsadministration wird die Arbeit nicht einfacher gemacht, wenn der Blick auf das Potenzial der Einzelschule fokussiert ist. Auf Schulformen, Zugangs-, Übergangs- und Abgangsregelungen, Prüfungs- und Zertifizierungsrichtlinien, Lehrämter und Stellenpläne, Stundentafeln und Lehrpläne hatte die Bildungsadministration unmittelbaren Zugriff. Das soziale Klima, die pädagogischen Normen und Werte, die kollegialen Arbeitsformen einer Schule und vor allem das entscheidende Glied in der Wirkungskette, die Qualität des Unterrichts selbst, entziehen sich jedoch dem Zugriff von aussen. Stattdessen erhalten die Schulleitung und das Kollegium als „professionelle Lerngemeinschaft“ eine zentrale Rolle. Ministerium, Schulaufsicht, Landesinstitut und andere können die Entwicklung der Einzelschule und die Professionalisierung ihres Personals nur indirekt beeinflussen. Ihre Aufgabe besteht nicht mehr darin, den Arbeitsalltag von Schulen durch Vorschriften und Eingriffe zu steuern, sondern vielmehr Rahmenbedingungen und Instrumente bereitzustellen, die der Schule selbst Entwicklungsmöglichkeiten eröffnen. Zu solchen Instrumenten gehören beispielsweise Fortbildungsmassnahmen für Lehrkräfte und Schulleitungen, externe Beratung und Supervision durch unabhängige Experten, und nicht zuletzt Verfahren und Instrumente zur (Selbst- und Fremd-)Evaluation. Eine Administration, die sich auf ein derartiges neues Steuerungskonzept einlässt, riskiert den Verlust unmittelbarer Kontrolle, gewinnt aber möglicherweise im Blick auf Partizipation, Transparenz, letztlich auch Effektivität und Effizienz des gesamten Systems. Zumindest ist dies die grundlegende „Philosophie“ des neuen Verwaltungshandelns,

die im Bildungsbereich als Verbindung von Schulautonomie einerseits, Standards und externer Evaluation andererseits seit Jahren diskutiert wird.

Diese Massnahmen für Qualitätsentwicklung in Einzelschulen, die dem Schulqualitätskonzept von Fend, Steffens und anderen verpflichtet waren, wurden spätestens „nach PISA“ herausgefordert durch einen neuen Trend im deutschen Bildungssystem: Die leistungsbezogene Qualitätsbewertung mittels *large-scale-assessments*, Vergleichsarbeiten und ähnliche Testerhebungen. Relativ unbefriedigende Ergebnisse internationaler Schulleistungsstudien hatten deutlich gemacht, dass nicht nur die Schulkultur und das pädagogische Klima, sondern auch (und vielleicht vor allem) das Niveau der Lernergebnisse, die starke Heterogenität von Lernresultaten und deren enge Koppelung an die soziale Herkunft der Schülerinnen und Schüler dringende Reaktionen erfordern. Als eine Reaktion unter anderen wurde nun für Deutschland adaptiert, was angelsächsische und skandinavische Länder, die Niederlande und viele andere Staaten schon seit Ende der 80er Jahre erprobt hatten: die regelmässige Überprüfung des „Output“ von Schulen durch Lernerfolgsmessung.

Neuer Trend: leistungsbezogene Qualitätsbewertung durch Lernerfolgsmessung

Noch lässt sich nicht sagen, wie Qualitätssicherung und -entwicklung zukünftig organisiert sein wird (vgl. dazu Steinert & Klieme, 2004). Aufgrund ausländischer Erfahrungen müsste aber Ziel sein, über ein mehrdimensionales Monitoringsystem negative Effekte zu vermeiden. In diesem Beitrag soll daher an Beispielen dargestellt werden, welche Kriterien ein Qualitätssicherungs- und Entwicklungssystem idealerweise zu integrieren hat (Abschnitt 4 und 5). Zuvor wird in Abschnitt 2 und 3 der theoretische Hintergrund erarbeitet, indem der Begriff der Evaluation in seiner Mehrdimensionalität näher beleuchtet wird und Aspekte eines integrierten Systems der Qualitätssicherung und Qualitätsentwicklung beschrieben werden.

Aufbau dieses Beitrags

2. Was ist Evaluation?

2.1 Evaluation als Prozess

Bestimmungen von Evaluation sind vielfältig (Scriven, 2005; Wesseler, 1999); ein gemeinsamer Kern scheint darin zu liegen, dass in Evaluationen Bewertungen vorgenommen werden, die auf systematischen Untersuchungen beruhen. Wir gehen im Folgenden von einer Vorstellung aus, die Evaluation als eine Abfolge von teilweise einander bedingenden Entscheidungen, d.h. als Prozess betrachtet. Diese Vorstellung lässt sich in Form eines Kreislaufs darstellen (Abb. 1, im Ansatz vergleichbare Konzeptionen finden sich bei Stufflebeam, 2003a, S. 786; Sanders & Davidson, 2005; S. 815; Büeler, 2004, S. 277; Rolff, 1998, S. 265):

Evaluation wird hier verstanden als der Prozess, innerhalb dessen eine zweckgerichtete Auswahl von Bewertungskriterien erfolgt, eine Institution oder Massnahme auf Basis dieser Kriterien systematisch untersucht und bewertet wird und eine Kommunikation über die Bewertung stattfindet mit dem Ziel, Konsequenzen abzuleiten.

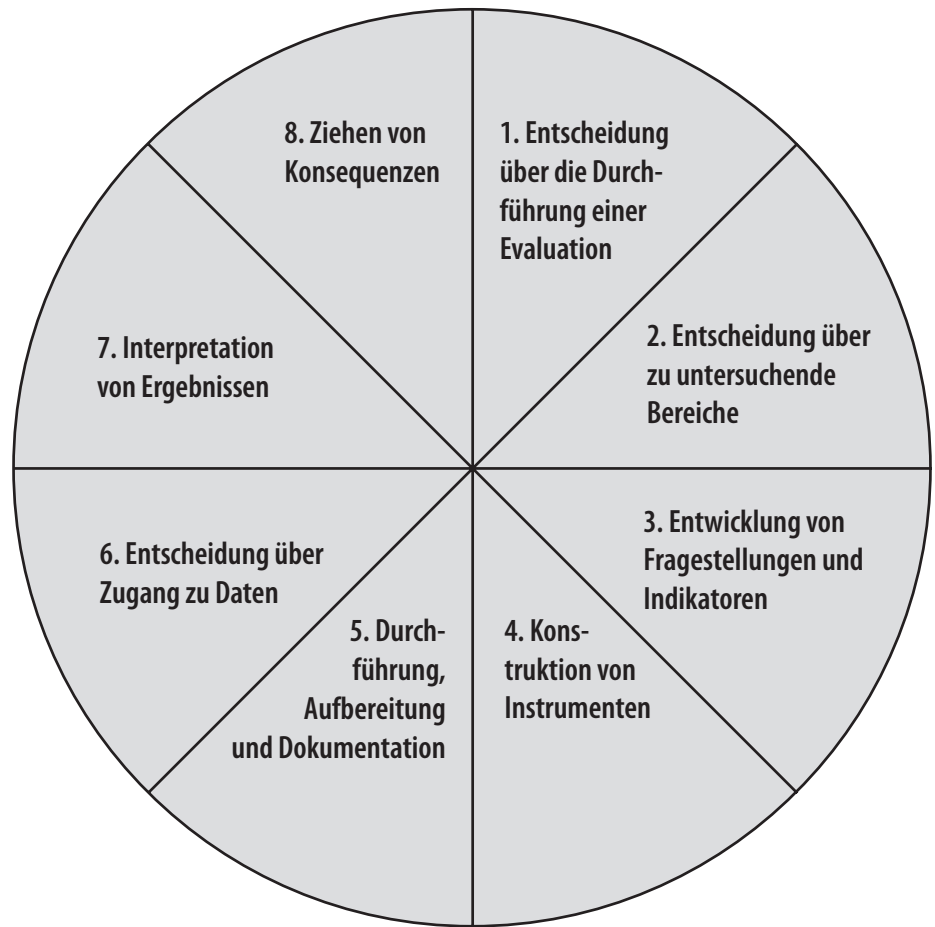


Abb. 1: Entscheidungsfelder in der Evaluation

Unterscheidung zwischen externer Evaluation und Selbstevaluation ist in der Regel unscharf.

Jede der Entscheidungen in diesem Kreislauf kann ausserhalb der Schule fallen oder in ihr. Deshalb ist die Unterscheidung zwischen externer Evaluation und Selbstevaluation in der Regel unscharf (Dubs, 1999; Abs & Döbrich, 2004). Eine Selbstevaluation im strikten Sinne, in der alle anstehenden Entscheidungen durch die Beteiligten selbst bewusst getroffen werden, wird sich kaum finden; vielmehr werden als Selbstevaluation solche Massnahmen bezeichnet, bei denen im Rahmen des Kontinuums zwischen externer und interner Evaluation einzelne Entscheidungen durch die Beteiligten selbst getroffen werden. Dabei sind in den jeweiligen Entscheidungsfeldern noch Abstufungen möglich. So fragt Altrichter (1999a) zum Beispiel danach, wann man sagen dürfe, eine Schule habe die Entscheidung zur Evaluation selbst getroffen: Welche Gruppen in der Schule müssen an der Entscheidung wie beteiligt gewesen sein? Reicht es zum Beispiel aus, dass die Schule den Zeitraum bestimmen durfte? Somit erscheint es auch wenig sinnvoll, voneinander unabhängige Standards für externe Evaluation und Selbstevaluation zu entwickeln.

Tiefer als die Abgrenzung von interner und externer Evaluation geht die Ausdifferenzierung epistemologischer Grundlagen und Zwecke, die mit einer Evaluation verbunden werden. Sie sollen im Folgenden idealtypisch als Paradigmen beschrieben werden. Weiter unten wird zu zeigen sein, dass diese Paradigmen in je spezifischer Weise die Phasen des Evaluationsprozesses und auch die unterschiedlichen Standards hervorheben.

2.2 Paradigmen der Evaluation

In einer funktionalen Betrachtung von Evaluation lassen sich drei Perspektiven unterscheiden (Chelimsky, 1997; Wesseler, 1999), die von Kromrey (2000) als Paradigmen benannt wurden. Wir greifen den Begriff des Paradigmas hier bewusst auf, weil sich mit jedem Paradigma nicht nur eine unterschiedliche Absicht verbindet, die durch eine Evaluation erreicht werden soll, sondern darüber hinaus auch jeweils unterschiedliche Aspekte der schulischen Wirklichkeit in den Blick kommen und unterschiedliche Methoden bevorzugt werden (Beywl, 1987; Donmoyer, 2001). Unterschieden werden

- das Forschungsparadigma
- das Entwicklungsparadigma
- das Legitimations- und Kontrollparadigma

Im Rahmen des *Forschungsparadigmas* orientieren sich die Ziele einer Evaluation an wissenschaftlichen Erkenntnisinteressen. Das bedeutet, dass Fragestellungen auf der Grundlage von Theorien und bisheriger Forschungsergebnisse generiert und begründet werden müssen. Evaluationsfragen werden von Wissenschaftlern entwickelt, die gegenüber dem Evaluationsgegenstand die Rolle eines externen Beobachters einnehmen. Von daher liegt es nahe, dass die Methoden innerhalb dieses Paradigmas durch das Ideal des Experiments gekennzeichnet sind. Als Ergebnis der Evaluation werden Erkenntnisse angestrebt, die in hohem Masse generalisierbar sind. Typische Fragen der Evaluation richten sich auf die Stärke von Zusammenhängen, auf die Wirkrichtung von Kausalität oder auf die Bedingungen von Entwicklung. Von Grundlagenforschung kann dieses Paradigma dadurch abgegrenzt werden, dass ihm ein Handlungsbezug inhärent ist. (Differenziert ausgearbeitet findet sich ein solches Evaluationsverständnis etwa bei Rost, 2000.)

Im Rahmen des *Entwicklungsparadigmas* werden die Ziele einer Evaluation an einem Verbesserungsinteresse in Bezug auf Institutionen orientiert. Deshalb müssen Fragestellungen auf der Grundlage von Zielen einer Organisation oder Massnahme entwickelt und begründet werden. Dies geschieht im Rahmen eines Beratungsprozesses, der entweder von den Mitgliedern einer Organisation bzw. den Akteuren einer Massnahme oder von aussen gesteuert werden kann. Jedoch stellen, auch bei einer Steuerung von aussen, die Zielsetzungen und Bedürfnisse der Beteiligten eine wichtige Grundlage dar. Methodisch bieten sich daher stärker interaktive Verfahren an. Eine hohe Interaktion zwischen Evaluatorinnen und Evaluatoren und Akteuren im „Feld“ erscheint hier als Voraussetzung für die Umsetzung der Evaluationsintention. Die Evaluation zielt unmittelbar auf Handlungsalternativen für den konkreten Fall, ohne Anspruch auf Verallgemeinerung. Typische Fragen der Evaluation richten sich auf die Ermittlung von Ressourcen, die Klärung von Zielvorstellungen oder die Identifikation von Hindernissen. Dieses Evaluationsverständnis findet sich besonders ausgeprägt in den Arbeiten von Rolff und Mitarbeitern (z.B. Dalin & Rolff, 1990) und bei Autoren, die der Tradition der Handlungsforschung verbunden sind, wie Posch (1999) und Altrichter (1999a; 1999b).

Im Rahmen des *Legitimations- und Kontrollparadigmas* orientieren sich die Ziele der Evaluation am Konzept der Rechtfertigung gegenüber geltenden Vorgaben. Das bedeutet, dass Fragestellungen sich jeweils auf die Vergangenheit oder den aktuellen Stand einer Organisation oder Massnahme richten. Dies ist aus Sicht der Legitima-

Es lassen sich drei Paradigmen der Evaluation unterscheiden.

Das Forschungsparadigma: Ziele einer Evaluation orientieren sich an wissenschaftlichen Erkenntnisinteressen.

Das Entwicklungsparadigma: Ziele einer Evaluation orientieren sich an einem Verbesserungsinteresse von Institutionen.

Das Legitimations- und Kontrollparadigma: Ziel der Evaluation ist die Rechtfertigung gegenüber Vorgaben.

tions- bzw. Kontrollinstanz sinnvoll, weil im Rahmen dieses Paradigmas ein wesentlicher Effekt der Evaluation (nämlich die Ausrichtung an Vorgaben) bereits durch ihre Ankündigung erzielt werden soll. Die Fragen der Evaluation werden in der Regel von aussen vorgegeben. Gleichwohl kann die Kontrolle auch innerhalb einer Organisation institutionalisiert werden. Wie sich am Beispiel des betrieblichen Controllings zeigen liesse, wird dann die interne Kontrolle allerdings auch als „extern“ im Verhältnis zu den operativen Einheiten wahrgenommen. Die Evaluatorin oder der Evaluator nimmt eine beobachtende Rolle von aussen ein, jedoch nicht um das Geschehen im Feld möglichst unbeeinflusst zu lassen, sondern um auf alle Evaluierten in gleicher Weise einzuwirken. Die Methoden, die in diesem Paradigma zur Anwendung kommen, müssen in besonderer Weise der Herausforderung von Gefälligkeitsantworten und der Täuschung durch die Evaluierten begegnen. Deshalb ist es hier besonders wichtig, mehrere Perspektiven einzubeziehen und nach nicht-reaktiven Indikatoren (z.B. Wiederholerraten, Unfallziffern) oder objektiveren Daten wie z.B. landesweiten Vergleichsarbeiten Ausschau zu halten. Für die Schulevaluation kann dies bedeuten, dass schriftliche Befragungen kombiniert werden mit Leistungstests, die an Bildungsstandards orientiert sind, und mit Empfehlungen aus einer umfassenden Schulinspektion. Als Ergebnisse der Evaluation werden Aussagen über einen Einzelfall angestrebt, die eine Bewertung anhand der vorgegebenen Kriterien ermöglichen. Eine Skizze eines solchen Evaluationssystems findet sich in der Expertise zur Entwicklung nationaler Bildungsstandards (Klieme et al., 2003).

Sobald mehrere Zwecke gleichzeitig durch Evaluation erreicht werden sollen, kann die Komplexität des Feldes kaum angemessen berücksichtigt werden.

Innerhalb eines Paradigmas herrscht Einigkeit darüber, was als wichtig zu betrachten ist, welche Vorgehensweise legitim ist und wie argumentiert werden kann. Durch die Einschränkung auf eine Funktion leisten die Paradigmen eine Reduktion von Komplexität und geben der Evaluatorin oder dem Evaluator eine eindeutige Orientierung. Um diese nicht zu verlieren, wird das eigene Paradigma bisweilen durch die Abwertung oder Negierung anderer Herangehensweisen abgesichert. Somit ist jedem Paradigma eine Tendenz zur Verabsolutierung inhärent. Die Stärken der einzelnen Paradigmen werden in den Vordergrund gerückt und dabei wird übersehen, dass die Komplexitätsreduktion jeweils nur durch eine funktionale Einschränkung möglich war. Bei strenger Orientierung an den Paradigmen dient Evaluation entweder einem Zweck oder einem anderem, aber nicht mehreren gleichzeitig. Sobald mehrere Zwecke durch Evaluation gleichzeitig erreicht werden sollen (z.B. Kontrolle und Entwicklung), wird es schwierig, die Komplexität der im Feld vorhandenen Wirkkräfte und Ansprüche in einem Design angemessen zu berücksichtigen. Beispiele hierfür bilden die von Experten vielfach kritisierten *accountability*-Systeme in den USA, die nur begrenzt positive Impulse für pädagogische Schulentwicklung zu geben scheinen (Fuhrman & Elmore, 2004), aber auch die letztlich nicht voll befriedigenden Versuche, Ergebnisse aus stichprobenbasierten Schulstudien, wie z.B. PISA, so an die beteiligten Schulen in Deutschland zurückzumelden, dass neben dem Forschungs- ein Entwicklungszweck erfüllt wird (Watermann et al., 2003; Klieme et al., 2000). Widersprüchliche Zwecksetzungen können letztlich den Erfolg von Evaluationen gefährden, so dass es wichtig erscheint, die funktionale Berechtigung und Eigenständigkeit aller drei Paradigmen anzuerkennen.

2.3 Schwerpunkte des Evaluationsprozesses in den Paradigmen

Die Felder im Evaluationskreislauf erhalten je nach Paradigma unterschiedliche Gewichtung.

Die Entscheidungsfelder des Evaluationskreislaufs werden im Rahmen der Paradigmen in unterschiedlicher Weise repräsentiert. Dies zeigt sich besonders prägnant am ersten Schritt: Die Entscheidung über die Durchführung erfolgt im Entwicklungs-

paradigma idealerweise durch die Betroffenen selbst, im Kontrollparadigma durch übergeordnete Instanzen bzw. Regelungen, im Forschungsparadigma hingegen im Rahmen der Stichprobenziehung. Im Kontext des Forschungsparadigmas spielen sodann die Entscheidungsfelder zwei bis sieben eine wichtige Rolle; Entscheidungen werden hier weitgehend extern, d.h. ohne Beteiligung der Evaluierten, getroffen. Die Felder eins (Entscheidung über die Durchführung von Evaluation) und acht (Ziehen von Konsequenzen) werden im Rahmen des Forschungsparadigmas typischerweise durch Informationsveranstaltungen abgedeckt, in denen die Forscher den Beteiligten die geplanten Erhebungen vorstellen und Ergebnisse präsentieren. Beides erfolgt in erster Linie, um die Teilnahmemotivation zu erhöhen.

Im Rahmen des Entwicklungsparadigmas sind alle Entscheidungsfelder relevant; es handelt sich hier um das einzige Paradigma, in dem es eindeutig die Aufgabe der Evaluatorinnen und Evaluatoren ist, das achte Entscheidungsfeld (Ziehen von Konsequenzen) ins Zentrum der Evaluationsplanung zu rücken. Ein Design im Rahmen dieses Paradigmas, das sich auf Bereiche konzentriert, in denen die Evaluierten keine Konsequenzen ziehen können, verfehlt seinen Zweck. Es gehört zu den Aufgaben der Evaluatorin oder des Evaluators, diesen Aspekt schon bei der Entscheidung über die Evaluationsbereiche zu berücksichtigen und in das Gespräch mit den Evaluierten einzubringen. Im Unterschied zu den beiden anderen Paradigmen ist es charakteristisch für das Entwicklungsparadigma, dass Entscheidungen insbesondere in den Feldern zwei, sieben und acht unter Beteiligung der Evaluierten getroffen werden. Gleichwohl ist eine solche Beteiligung nicht zwingend. Wie sehr sie auf den Erfolg eines Evaluationsvorhabens einwirkt, ist eine empirisch zu klärende Frage.

Dagegen gehört es zu den Merkmalen des Legitimations- und Kontrollparadigmas, dass die Beteiligung der Evaluierten deutlich eingeschränkt ist. Oberflächlich betrachtet scheint dies eine Parallele zum Forschungsparadigma zu sein, jedoch erfolgt der Ausschluss aus entgegengesetzten Gründen und wird deshalb auch anders vermittelt. Während im Rahmen des Forschungsparadigmas das Vorgehen der Evaluation den Evaluationsgegenstand möglichst wenig beeinflussen und deshalb der Evaluation als solcher im Feld möglichst wenig Aufmerksamkeit zukommen soll, ist es im Rahmen des Legitimations- und Kontrollparadigmas genau umgekehrt. Hier soll schon die Entscheidung zur Durchführung einer Evaluation das Verhalten aller Beteiligten möglichst stark beeinflussen. Ein sozial erwünschtes Verhalten, das im Rahmen des Forschungsparadigmas möglichst ausgeschlossen werden soll, ist im Legitimations- und Kontrollparadigma das eigentliche Programm der Evaluation. Insofern verbindet sich mit dem Kontrollaspekt zumeist auch ein Anliegen des Qualitätsmanagements, allerdings im Sinne eines von aussen gesteuerten oder zumindest initiierten Prozesses. Inwieweit dies gelingen kann, ist ein zentraler Punkt der aktuellen Debatten um die Einführung von Bildungsstandards in den deutschsprachigen Ländern (Klieme, im Druck).

3. Schulevaluation als Teil eines systematischen Bildungsmonitorings

Pädagogische Bilanzen per Fragebogen, moderierte Selbstevaluation und vergleichende Leistungsstudien bilden zentrale Bestandteile eines Systems von „educational evaluation, assessment and monitoring“. So der Titel eines aktuellen Werkes, mit dem

Das Entwicklungsparadigma: Das Ziehen von Konsequenzen aus der Evaluation steht im Zentrum.

Das Forschungsparadigma: Das Vorgehen der Evaluation soll den Evaluationsgegenstand möglichst wenig beeinflussen.

Legitimations- und Kontrollparadigma: Es soll ein sozial erwünschtes Verhalten der Evaluierten bewirkt werden.

der Schulforscher Jaap Scheerens, der Psychometriker Cees Glas und die Evaluationspezialistin Sally M. Thomas verschiedene Ansätze für Evaluation im Schulbereich systematisieren (Scheerens et al., 2003). Nach einer mehr als fünfzehnjährigen Phase intensiver Erprobung von Qualitätssicherung in verschiedenen Staaten versuchen die Autoren, die Verfahren zu ordnen und mit einem „systemischen“ Ansatz theoretisch zu beschreiben. Die hier vorgelegte Systematik kann der Diskussion in Deutschland, wo man gerade erst beginnt, die internationale Entwicklung aufzuarbeiten und eigene Qualitätssicherungskonzepte zu formulieren, wichtige Orientierungen geben.

Orientierungspunkte für Qualitätssicherungskonzepte

Evaluationsverfahren unter einer systemischen Perspektive einzuführen und zu gestalten, bedeutet nach Scheerens et al. (2003, S. 15),

- dass die Verfahren nicht singulär, etwa für die Bewertung spezifischer Massnahmen und Programme, eingesetzt werden, sondern dass sie als Bestandteil des Steuerungssystems institutionell eingeführt und laufend genutzt werden,
- dass Evaluation Entwicklungs- und Lernprozesse in Gang setzt und am Leben hält, die dem Prinzip des feedbackgesteuerten Lernens genügen,
- dass verschiedene Arten der empirischen Erhebung (Testeinsätze, amtliche Daten u.a.m.), die je spezifischen Zwecken dienen, dennoch gemeinsam betrachtet und Synergien genutzt werden,
- und dass schliesslich zur Systematisierung und Interpretation von Befunden ein theoretisches Rahmenmodell verwendet wird, welches Input-, Prozess- und Output- Elemente zueinander in Beziehung setzt.

Schulevaluation und Bildungsmonitoring sollten in einem System zusammengeführt sein.

Die Autoren ordnen somit die Evaluation einzelner Institutionen (Schulen) und Programme in den grösseren Kontext einer laufenden Bestandsaufnahme des Schulbereichs (sog. Monitoring) ein. Schulevaluation und Bildungsmonitoring sollten in einem System zusammengeführt sein, was die Autoren mit dem durchweg verwendeten Kürzel „E & M“ kennzeichnen; wir sprechen im vorliegenden Text vom „System der Qualitätssicherung“. Ein solches umfassendes System speist sich den Autoren zufolge aus drei Datenquellen: Zum einen sind dies administrative Daten, die der laufenden Statistik (etwa im Rahmen eines Schulinformationssystems oder einer virtuellen Schülerkartei) entnommen sind. Zum zweiten spezielle Erhebungs- und Untersuchungsansätze, die zu bestimmten Zeitpunkten zusätzliche Informationen über einzelne Schulen liefern. (Hierzu gehören unterschiedlichste Verfahren der Selbstevaluation, Schulinspektion, Schulaufsicht, dienstrechtliche Bewertungen von Lehrerinnen und Lehrern und anders mehr.) Schliesslich kommen als dritter Bereich Testerhebungen hinzu – sei es im Rahmen von internationalen Programmen (TIMSS, PISA usw.), im Rahmen von nationalen und regionalen Leistungsstudien, von zentralisierten Prüfungen oder von Testeinsätzen im Rahmen der regelmässigen Schulevaluation.

Drei Funktionen von Qualitätssicherungssystemen

Auf der Basis dieser Daten erfüllen Monitoring- und Evaluationssysteme den Autoren zufolge drei Funktionen:

- Akkreditierung und Zertifizierung,
- Rechenschaftslegung („accountability“) sowie
- Diagnose und Organisationslernen.

Die Autoren entwerfen das Bild eines empiriegestützten Schulentwicklungsprozesses, der durch eine Verbindung von externer und interner Evaluation vorangebracht wird. Evaluation sehen sie (Scheerens et al., 2003, S. 94) als das Herzstück einer lernenden Organisation, d.h. einer Schule, die ihre interne Effektivität und ihre Verantwortlich-

keit gegenüber externen Ansprüchen verbessern will. Eine solche feedbackgestützte Strategie der Schulentwicklung sei vermutlich effektiver als Innovationsstrategien, die „vorwärts planend“ arbeiten. Entscheidend für das Modell ist, dass Ergebnis- und Prozessqualität gleichermaßen zum Gegenstand des Feedbacks gemacht werden. Insbesondere sollten möglichst alle wichtigen Prozessfaktoren abgedeckt werden, die – gemäss den Ergebnissen der Schulwirkungsforschung (Scheerens & Bosker, 1997) – für erfolgreiche Schulen typisch sind, also vor allem:

- Leistungsorientierung,
- Kooperation, Konsens und Kohärenz im Kollegium,
- pädagogische Führung,
- curriculare Qualität/Lerngelegenheiten,
- Schulklima und Schulkultur,
- Evaluation,
- Einbeziehung der Eltern.

Prozessfaktoren erfolgreicher Schulen

Mit dieser Breite von Evaluationskriterien geht das von Scheerens und Mitautoren vorgestellte Konzept deutlich über die so genannten „accountability systems“ hinaus, wie sie in den USA entwickelt worden sind (vgl. Fuhrman & Elmore, 2004). Dort dominiert immer noch eine rein leistungsbezogene Bewertung von Schulen, gestützt zumeist auf standardisierte Tests, die regelmässig – im Extremfall jährlich für jede Schülerin und jeden Schüler – eingesetzt werden. Die Erfahrungen der USA belegen, dass ein solches engmaschiges, ausschliesslich testbasiertes Monitoring-System zwar Effekte im Sinne von über die Jahre anwachsenden Leistungsniveaus zeigen kann – Effekte jedoch, deren Nachhaltigkeit nicht belegt ist, die möglicherweise auf elementare Leistungsbereiche begrenzt sind, welche schnell trainiert und leicht abgetestet werden können, und die vermutlich gerade jene Schulen, die besondere Probleme haben, gar nicht erreichen (vgl. Klieme, 2004). Gleichzeitig haben solche extensiven Testsysteme nachweislich negative Folgen, beispielsweise durch ein „teaching to the test“, das die Breite und die Tiefe des Curriculums beeinträchtigt. Experten der Schulentwicklung und Evaluation in den USA empfehlen daher auch dort den „Rückbau“ der accountability-Systeme in Richtung auf eine vorsichtigere, unterrichtsnahe, multiple Kriterien einbeziehende Schulevaluation.

Extensive Testsysteme haben negative Folgen.

Das entscheidende Argument für die Berücksichtigung von Ergebnis- und Prozess-Aspekten ergibt sich aus der Frage, wie denn *Qualitätsentwicklung* bzw. – in der Terminologie von Scheerens et al. – die Funktion des „Organisationslernens“ realisiert werden soll. Eine blosser Feststellung von Lernergebnissen mag für die Rechenschaftslegung nach aussen oder (in einem noch stärker deregulierten System) für eine Akkreditierung von Institutionen genügen. Eine Schule als „lernende Organisation“ muss aber auch wissen, an welchen Prozessfaktoren sie ansetzen kann, um ihre Ergebnisse zu sichern bzw. zu verbessern. Evaluation ohne Prozessanalyse bedient bestenfalls legitimatorische Zwecke und Kontrollwünsche, ermöglicht jedoch keine Weiterentwicklung. (Zu dieser Unterscheidung von Evaluationszielen vgl. Abs & Klieme, im Druck).

Evaluation ohne Prozessanalyse bedient bestenfalls legitimatorische Zwecke.

Auch die Bewertung von Ergebnissen muss Wahrnehmungen und Urteile aller Beteiligten einbeziehen, wenn Evaluation zum Ausgangspunkt für Schulentwicklung werden soll. Daneben bedarf es aber „objektiver“ Indikatoren wie z.B. Testwerte, die anhand schulübergreifender Standards bzw. Vergleichsmassstäbe bewertet werden. Ansonsten würde man auf eine Politik des „pädagogischen Goodwill“ zurückfallen, die sich – wie internationale Vergleiche nahe legen – als wenig erfolgreich erwiesen hat (Fend, 2002).

Um die Verbindung verschiedener Aspekte hervorzuheben, bezeichnen wir das Gesamtkonzept eines neuen Steuerungssystems für den Schulbereich in Anlehnung an Fend (2002) als „integriertes System der Qualitätssicherung und Qualitätsentwicklung“. (Der Begriff „Qualitätssicherung“ wird in der deutschsprachigen Fachdiskussion – entgegen seinem sprachlichen Sinn – zumeist auf Kontrolle eingeschränkt und blendet somit Lern- und Entwicklungsaspekte aus.)

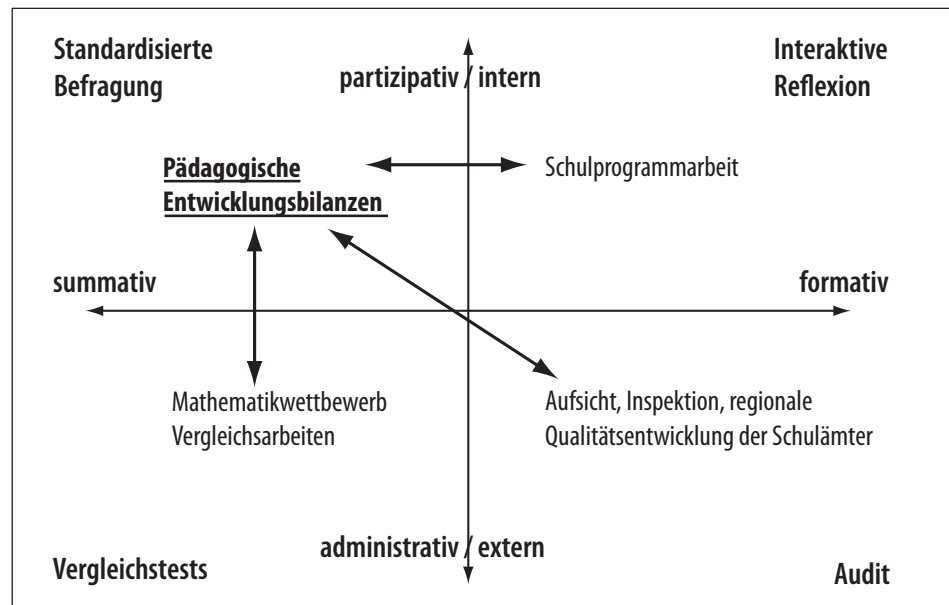


Abbildung 2: Typen der Schulevaluation, zugehörige Verfahren und Instrumente am Beispiel Hessens

Beispiele verschiedener Typen von Evaluation

In einem solchen integrierten System müssen unterschiedliche Formen von Evaluation kombiniert werden (vgl. Abb. 2). Wir veranschaulichen dies in einer zweidimensionalen Matrix, indem wir externe (zumeist administrativ gesteuerte) und interne (weitgehend partizipativ gesteuerte) Evaluationsformen unterscheiden und zugleich formative (entwicklungsbegleitende, prozessorientierte) von summativer (bilanzierender, ergebnisorientierter) Evaluation abheben. Typische Arbeitsformen, aber auch Instrumente von Schulevaluation können in diesem Schema verortet werden. So bilden standardisierte, von zentralen Institutionen vorgegebene Vergleichstests ein Beispiel für extern-summative Evaluation, während der Gegenpol, intern-formative Evaluation, zumeist als ein Reflexionsprozess mit hoher Interaktionsdichte abläuft, wie er beispielsweise in der Schulprogrammarbeit (wenn sie denn erfolgreich vonstatten geht) zu finden ist. Eine dritte, in den Niederlanden und Grossbritannien bereits flächendeckend eingeführte Form der Evaluation, die in Deutschland zunehmende Aufmerksamkeit erhält, sind „audits“, d.h. Inspektionen einer Schule durch spezielles Personal, durch „critical friends“ oder durch Arbeitsgruppen der Schulaufsicht, die alle Komponenten des Schullebens – von der Unterrichtspraxis bis zur Kooperation der Schule mit externen Partnern – beleuchten und Anstöße zur Schulentwicklung geben. Ein vierter Idealtypus schliesslich, der ebenso wie Vergleichstests summativ-bewertend arbeitet, aber doch intern gesteuert werden kann, sind Befragungen mittels (teil-)standardisierter Instrumente. Für diesen Typus von Evaluation wurden in den vergangenen Jahren eine Reihe von Instrumenten bereitgestellt, beispielsweise die von Ditton, Arnold und Bornemann (2002) eingesetzten Fragebogen, das Instrumentarium des österreichischen Modellversuchs „Qualität in Schulen (qis)“, das Fra-

gebogen-Instrumentarium des Modellversuchs „INIS“ der Bertelsmann-Stiftung und andere mehr. Zu diesem Typus von Evaluationsverfahren gehören auch die „Arbeitsplatzuntersuchungen“ bzw. „Pädagogischen Entwicklungsbilanzen (PEB)“ des DIPF (Döbrich, 2005). In diesen Befragungen geht es auch um Arbeits- und Lernprozesse, nicht zuletzt aber um die Einschätzung der „Ergebnisse“ im weitesten Sinne, z.B. im Sinne der Zufriedenheit und des Wohlbefindens der Beteiligten, und um ihre jeweilige Einschätzung der Zielerreichung.

Kombiniert man die unterschiedlichen Elemente, wie in Abb. 2 dargestellt, so entsteht ein Geflecht von Massnahmen, das dem „systemischen“ Ansatz von Scheerens et al. nahe kommt und anschlussfähig ist für Prozesse des Organisationslernens. Ein solches integriertes System wird im deutschsprachigen Raum seit einiger Zeit diskutiert (Böttcher, 2002; Dubs, 1999; Klieme et al., 2000; Rolf, 1998), ist aber wohl noch nie realisiert worden. Es scheint, als verfolgten viele Bildungspolitiker und -administratoren derzeit eine einseitig leistungsbezogene, testbasierte Art von Qualitätssicherung. Wie man dies mit Elementen einer prozessbezogenen, entwicklungsorientierten Schulevaluation verbinden könnte, soll, basierend auf den Erfahrungen im Bundesland Hessen, im Folgenden dargelegt werden.

Derzeit wird eine einseitig leistungsbezogene, testbasierte Art von Qualitätssicherung verfolgt.

4. Elemente des Qualitätssicherungssystems in Hessen

4.1 Pädagogische Entwicklungsbilanzen als Quelle für Prozess- und Ergebnismerkmale

Die Fragebogen-Erhebungen, die Döbrich in Hessen verantwortet, beziehen sich in ihrer jüngsten Version auf folgende Fragenkomplexe (vgl. Döbrich, 2005):

- Unterricht (Unterrichtsformen, Unterrichtsverlauf, Benotungspraxis, Vorbereitung, Hausaufgaben),
- Einschätzung der Motivation, Leistung und Zufriedenheit der Schüler,
- schulische Organisation,
- Kommunikationsverhalten und Zusammenarbeit im Kollegium,
- Probleme an der Schule und im Unterricht,
- Berufsansprüche und Ziele der Lehrkräfte sowie Einschätzung der Zielerreichung,
- persönliche Arbeitszeitgestaltung,
- Berufszufriedenheit,
- Belastungs- und Bewältigungserleben sowie Gesundheitszustand.

Fragenkomplexe des Qualitätssicherungssystems in Hessen

Schülerbefragungen und Befragungen von Eltern ergänzen die Sicht der Lehrkräfte. Das Prinzip dabei ist, die Beurteilungsdimensionen aus unterschiedlichen Perspektiven zu erhellen, um Debatten an den Schulen anzuregen. Die Fragebogenskalen decken im Wesentlichen die Schulqualitätsdimensionen ab, die Scheerens und Bosker (1997) identifiziert haben. Wo dies in den ersten Erhebungen noch nicht genügend realisiert war, wurde für Erhebungen ab 2005 das Instrumentarium auf der Basis einer Skalensammlung (Steinert et al., 2005) mit Rückgriff auf bewährte Instrumente anderer Studien erweitert.

Freiwillige Teilnahme der Schulen

Schulen in Hessen können freiwillig teilnehmen und erhalten eine ausführliche Dokumentation ihrer Ergebnisse, die zumeist im Rahmen eines pädagogischen Tages vorgestellt und mit dem Kollegium diskutiert wird. Voraussetzung für die Auswertung und Rückmeldung ist eine Mindestbeteiligungsquote von 75% im Lehrerkollegium, um sicherzustellen, dass die Aussagen repräsentativ für die Gesamtheit der Lehrkräfte und damit der Schule sind. Auch Vergleiche mit anderen Schulen derselben Schulform werden vorgenommen und helfen, das eigene Ergebnis zu interpretieren.

Über die Nutzung der Rückmeldungen für die schulinterne Reflexion, bei der häufig auch einzelne Fragen aus den Bögen angesprochen werden, hat Döbrich (2003) an anderer Stelle berichtet. Im vorliegenden Beitrag geht es schwerpunktmässig darum, die Aussagekraft der Erhebungen aus einer systematischen Perspektive darzustellen. Exemplarisch sollen hierzu aus den Lehrer/innenfragebögen sechs Skalen betrachtet werden; die psychometrischen Kennwerte entstammen Steinert et al. (2003).

Sechs Skalen aus dem Lehrer/innenfragebogen

- **Unterrichtsbezogene Kooperation:** 9 Items (interne Konsistenz .80) erfassen spezifische Aspekte der Kooperation zwischen Lehrkräften, bei denen es um die Gestaltung von Unterricht geht. Die höchste Zustimmung hat die Aussage „Wir haben eine gute fachspezifische Zusammenarbeit“, die niedrigste findet sich für „Gegenseitige Unterrichtsbesuche sind ein selbstverständlicher Teil unserer Arbeit.“ Markieritem mit höchster Trennschärfe ist „Die Koordination der Unterrichtsarbeit innerhalb der Jahrgangsstufen ist gut organisiert.“
- **Pädagogische Führung:** Diese Skala umfasst 6 Items, mit denen die Lehrkräfte ihre Zufriedenheit mit der Arbeit der Schulleitung ausdrücken (interne Konsistenz .90). Beispielitems: „Ich bin mit dem Führungsstil der Schulleitung zufrieden.“ und „Die Schulleitung achtet auf kollegiale Mitbestimmung.“
- **Elternbeteiligung:** Die Lehrkräfte schätzen, wie hoch der Anteil der Eltern ist, die an schulischen Veranstaltungen, Elternabenden und Sprechtagen teilnehmen (3 Items, interne Konsistenz .89).
- **Belastung durch Schülerprobleme:** Auf die Frage „Wenn es in Ihrem Unterricht mal nicht so läuft, wie Sie es sich vorgestellt haben, was spielt Ihrer Meinung nach dabei eine Rolle?“ war die Bedeutung verschiedener Faktoren einzuschätzen. 4 Items, die auf unmotivierte, problembelastete, disziplinlose und unkonzentrierte Schüler verweisen, bilden eine Skala mit einer internen Konsistenz von .77.
- **Aggressionsniveau unter Schülern:** Mittels 6 Items (interne Konsistenz .84) wird das Ausmass der Gewalt zwischen Schülerinnen und Schülern bewertet. Es wird gefragt, wie häufig es vorkommt, dass Schüler schlagen, erpressen, einen Diebstahl begehen, zu wenig Rücksicht auf jüngere nehmen bzw. diese bedrohen oder im Unterricht stören.
- **Individuelle Förderung:** Mittels 5 Items (interne Konsistenz .67) beurteilten die Lehrkräfte, inwieweit es ihnen gelingt, sich für einzelne Schüler zu engagieren, Leistungsschwächere zu fördern und ausländische Schüler zu integrieren.

Unter den Komponenten der Schülerbefragung seien zwei erwähnt:

- **Aggressionsniveau unter Schülern:** Die 6 Items aus der Lehrerbefragung wurden hier gespiegelt. Die interne Konsistenz beträgt .82.
- **Wohlbefinden:** Diese Skala hat 6 Items (Konsistenz .75) mit dem Markieritem „Ich fühle mich in meiner Schule wohl.“

Schon diese Auswahl macht deutlich, dass die Befragungen sowohl Prozessmerkmale als auch Ergebnisvariablen erfassen. Hinsichtlich der Prozessqualität werden zentra-

le Komponenten des Scheerens/Bosker-Modells abgedeckt, hier etwa Kooperation im Kollegium, pädagogische Führung und Elternbeteiligung. Im Bereich der Ergebnisse geht es teils um erzieherische Kriterien, operationalisiert u.a. durch das Aggressionsniveau der Schüler/innen, teils um leistungsbezogene Kriterien, hier operationalisiert durch Fragen nach der Erreichung von Zielen im Bereich der individuellen Förderung. Merkmale wie das Aggressionsniveau oder das Wohlbefinden der Schüler lassen sich je nach Fragestellung als Prozess- oder Ergebnisvariablen betrachten.

Es werden sowohl Prozessmerkmale als auch Ergebnisvariablen erfasst.

4.2 Mathematik-Wettbewerb als Output-Indikator

Der Wettbewerb, der seit Jahren von engagierten hessischen Mathematiklehrer/innen unter Leitung von Klaus Allendörfer ehrenamtlich durchgeführt wird, sprach zunächst Schüler/innen mit Interesse an Mathematik in allen Schulformen auf freiwilliger Basis an. Seit dem Schuljahr 2002/03 hat ihn das Kultusministerium für sämtliche Schulen der Sekundarstufe I in Hessen verbindlich gemacht. Der Wettbewerb wird nun, eingeteilt in drei Schwierigkeitsstufen (A, B, C – entsprechend den Bildungsgängen der Sekundarstufe I), jährlich flächendeckend in Hessen durchgeführt. Für jede Stufe gibt es einige Pflichtaufgaben sowie anspruchsvollere Wahlaufgaben, unter denen jede Schülerin und jeder Schüler auswählen kann. Die Aufgabensätze werden jährlich nach Durchführung des Wettbewerbs veröffentlicht. Eine Inspektion dieser Quellen lässt vermuten, dass der Wettbewerb das Curriculum und die Alltagspraxis der Schulen (Klassenarbeiten) recht gut abbildet.

In Hessen wird jährlich ein flächendeckender Mathematik-Wettbewerb durchgeführt.

Die Ergebnisse der Schüler/innen werden – auf Schulebene aggregiert, d.h. als mittlere Zahl der an einer Schule erreichten Punkte – vom Kultusministerium ins Internet gestellt. Diese Daten – bezogen auf die Erhebung im Frühjahr 2004 – verwenden wir im Folgenden als Indikator für die Lernresultate der Schulen.

Bislang werden die Aufgaben des Wettbewerbs weder vorerprobt noch auf Individualebene statistisch ausgewertet. Die Korrektur der Schülerlösungen obliegt – auf der Basis zentral gesetzter Kriterien – den Mathematiklehrkräften vor Ort. Lediglich das Gesamtergebnis der Schule wird an die zentrale Arbeitsgruppe beim Kultusministerium zurückgemeldet, die landesweite Ergebnisse ermittelt und präsentiert. Aus pragmatischer Sicht hat dieses Vorgehen zwei Vorzüge: Zum einen ist es deutlich einfacher und kostengünstiger als eine Erhebung, die eine Datenerfassung und -verarbeitung auf zentraler Ebene erfordert. Zum anderen bleibt die Implementation des Leistungsvergleichs weitgehend in der Kontrolle der Schulen, was für die Akzeptanz des Verfahrens möglicherweise entscheidend ist. Allerdings lässt sich die psychometrische Qualität der Instrumente bei dieser Vorgehensweise nicht beurteilen, sodass letztlich auch die Aussagekraft des Wettbewerbs – so überzeugend seine Aufgaben auch sein mögen – ohne Beleg bleibt. Auf Dauer wird sich, so vermuten wir, eine Akzeptanz der Schulvergleiche nur herstellen lassen, wenn die Qualität des Verfahrens zumindest auf Stichprobenbasis empirisch kontrolliert und belegt werden kann.

Die Datenerfassung und -verarbeitung geschieht teilweise dezentral.

Die Ergebnisse eines einzelnen Tests wie des Mathematik-Wettbewerbs liefern an sich nicht mehr Information als die Stellung der jeweiligen Schule im Leistungsvergleich, ausgedrückt als *ranking* oder – wie in den offiziellen hessischen Tabellen – als Zugehörigkeit zu einer Ergebniskategorie (Quartil). Werden die Aufgaben – wie für die Zukunft geplant – auf Bildungsstandards bezogen, ergibt sich zusätzlich die Möglichkeit einer kriteriumsbezogenen Bewertung. Wenn unterschiedliche Dimensionen innerhalb der Mathematik unterschieden werden könnten oder wenn die Tests

Zukünftig sind kriteriumsbezogene Bewertungen und ipsative Vergleiche möglich.

mehrerer Jahre auf derselben Skala verankert wären (durch Übernahme einzelner Aufgaben und so genanntes „Testequating“), wären auch ipsative Vergleiche möglich, d.h. Aussagen zu relativen Stärken und Schwächen bzw. zu Veränderungen im *Output* einer Schule im Längsschnitt.

**Pädagogische Entwicklungs-
bilanzen ermöglichen zusätzliche
Analysen der Daten des
Mathematik-Wettbewerbs.**

Durch Hinzunahme der Pädagogischen Entwicklungsbilanzen ergeben sich aber schon jetzt zusätzliche Vergleichsmöglichkeiten. Wenn die Ergebnisse des Mathematik-Wettbewerbs aus dem Jahr 2004 im Folgenden mit Resultaten der Pädagogischen Entwicklungsbilanzen in Beziehung gesetzt werden, dann auch mit dem Zweck, nachzuweisen, dass schon bei der jetzigen Form des Wettbewerbs aussagefähige, theoriekonforme und innovative Ergebnisse zustande kommen. Wenn dies sogar mit Leistungsdaten gelingt, deren psychometrische Qualität noch nicht kontrolliert werden konnte, wäre ein starkes Argument für eine weitere Nutzung und Weiterentwicklung des Mathematik-Wettbewerbs gewonnen.

4.3 Derzeit noch offene Komponenten des Qualitätssicherungssystems: Input-Merkmale und darauf aufbauende differenzierte Rückmeldungen für die Schulen

**Input-Merkmale sind noch zu
wenig vertreten im Datengefüge.**

Relativ schwach sind in diesem Datengefüge bislang Input-Variablen vertreten. Die Frage nach der Belastung durch Schüler/innenprobleme könnte als Input-Indikator verstanden werden, allerdings gebrochen durch die Wahrnehmung der Lehrpersonen. Zukünftig sollten die Schüler/innen selbst im Rahmen der Pädagogischen Entwicklungsbilanzen zu Merkmalen ihres sozialen und familiären Hintergrunds wie z.B. dem Migrationsstatus befragt werden. Andere Aspekte wie die Grösse der Schule und die Lehrer - Schüler/innen-Relation sollten zukünftig der amtlichen Statistik entnommen werden. Zusammen mit Testuntersuchungen – derzeit noch begrenzt auf den Mathematik-Wettbewerb – entstünde so ein vollständiges Datentableau, das alle drei bei Scheerens et al. (2005) empfohlenen Quellen einbezieht.

**Bereinigung der Kennwerte um
den Effekt wichtiger Input-
Variablen ist unverzichtbar.**

Auf der Basis eines solchen Datentableaus könnte man schliesslich den Leistungsstand einer Schule um den Effekt wichtiger Input-Variablen (z.B. soziale Herkunft der Schülerschaft, Anteil von Kindern mit Migrationshintergrund) bereinigen. Im Kontext einer integrierten Strategie der Qualitätssicherung und -entwicklung erscheint eine solche Korrektur bei der Auswertung und Rückmeldung von Schulresultaten unverzichtbar, denn nur so kann die eigentliche Leistung der Schule, d.h. der von ihr erarbeitete „Mehrwert“ gegenüber den Eingangs- und Rahmenbedingungen fair evaluiert werden. Die Verwendung solcher statistisch bereinigter („value-added“) Kennwerte gehört heute – angesichts manch unerwünschter Vorkommnisse, die vor allem in Grossbritannien nach der Veröffentlichung unkorrigierter Testergebnisse in „league tables“ zu verzeichnen waren – zum Standardrepertoire der professionellen Schulevaluation.

Die Frage nach der Belastung durch Schüler/innenprobleme illustriert auch einen wichtigen methodischen Aspekt solcher Befragungen: Aggregiert auf Schulebene, als Durchschnitt der Urteile aller Lehrkräfte, mag die Frage tatsächlich die „Belastung“ der Schule durch ein schwieriges Klientel widerspiegeln. Auf der Ebene der einzelnen Lehrkraft jedoch geht es um die je individuelle Wahrnehmung bzw. Attribution von Schwierigkeiten im Unterrichtsablauf auf unterschiedliche mögliche Ursachen. Im Folgenden beschränken wir uns auf eine Betrachtung der (aggregierten) Daten auf Schulebene. Wissenschaftlich angemessener wäre eine so genannte Mehr-Ebe-

nen-Analyse, die zwischen individuellen und institutionellen Faktoren sowie zwischen Schul- und Klassenebene unterscheidet. Dies erlaubt jedoch die Datenlage für das hessische Projekt derzeit noch nicht.

In zweifacher Hinsicht wäre demnach bei der Datenerhebung eine Änderung des Untersuchungsplans sinnvoll: Zum einen sollten in den Entwicklungsbilanzen Inputdaten zur sozialen Herkunft der Schüler erfragt werden. Zum zweiten müssten die Ergebnisse des Mathematik-Wettbewerbs auch auf Individual- und Klassenebene mit den fragebogenbasierten Erhebungen verknüpft werden. Die *large-scale*-Studien wie PISA haben gezeigt, wie dies unter vollständiger Wahrung der Anonymität der Befragten möglich ist.

Bei der *Veröffentlichung* und *Nutzung* von Daten ist hingegen aus Sicht der Evaluationsforschung (Abs & Klieme, im Druck) grösste Zurückhaltung angezeigt – zumindest solange Betroffene und Öffentlichkeit noch lernen, mit solchen Befunden angemessen umzugehen. Dies betrifft nicht nur den vertraulichen Umgang mit Schulergebnissen, sondern auch einen vorsichtigen Umgang mit individuellen Testdaten, deren Messfehler von Lehrkräften häufig nicht bedacht werden. Zur Gestaltung und Nutzung des Feedbacks für Schulen geben neuere Forschungsarbeiten aus *large-scale-assessments* und Projekten der Schulevaluation in Deutschland wichtige Anstösse (Ditton & Arnold, 2004; Watermann & Stanat, 2004).

Einbezug von Inputdaten zur sozialen Herkunft und Verknüpfung der Mathematikergebnisse mit Fragebogenerhebungen

Im Umgang mit Veröffentlichung und Nutzung von Daten ist Zurückhaltung angezeigt.

5. Grundlagen und Ziele der empirischen Analysen

5.1 Fragestellungen und Erwartungen

Die folgenden Analysen der hessischen Daten sollen exemplarisch zeigen, welcher Erkenntnisgewinn sowohl für einzelne Schulen als auch schulübergreifend, für das gesamte System, aus dem Instrumentarium der Qualitätssicherung gezogen werden kann. Fragestellungen und Hypothesen hierzu lassen sich aus den Theorien und Befunden der Schulforschung, aber auch aus PISA ableiten.

a) Multiperspektivität

Die Untersuchung von Prozess- und Ergebnismerkmalen aus unterschiedlichen Perspektiven (Schüler, Lehrer, Eltern) gibt einen facettenreichen Blick auf die Realität des Schullebens. Eine Übereinstimmung der Urteile „im Grossen und Ganzen“, über alle Schulen hinweg, würde für die Validität der Urteile sprechen. Allerdings ist nicht damit zu rechnen, dass die Beteiligten immer in ihren Wahrnehmungen bzw. Urteilen übereinstimmen, und gerade aus den Diskrepanzen ergeben sich an der einzelnen Schule wichtige Anlässe für Debatten über den Entwicklungsbedarf. Schulevaluation kann so den verschiedenen Gruppierungen – vor allem den Schülerinnen und Schülern, aber auch den Eltern – ein wichtiges Sprachrohr liefern und zur demokratischen Schulentwicklung im Sinne von Transparenz und Partizipation beitragen.

Gerade aus den Diskrepanzen der Wahrnehmungen ergeben sich wichtige Anlässe für Debatten.

b) „Harte“ und „weiche“ Ergebniskriterien

Analog sollten sich „harte“ Leistungsdaten und „weiche“ Einschätzungen der Ergebnisqualität insgesamt stützen. Diskrepanzen werden auch hier Fragen auslösen nach der Stimmigkeit von objektiven Leistungsmessungen, deren subjektiver Einschätzung durch die Beteiligten und anderen Ergebniskriterien (z.B. dem Wohlbefinden an der Schule). In der internationalen Schulforschung findet sich eine Reihe von Hinweisen darauf,

Positiver Zusammenhang zwischen Zufriedenheit / Wohlbefinden und Lernergebnissen.

dass Zufriedenheit und Wohlbefinden der Beteiligten mit positiven Lernergebnissen einhergehen, wobei Wirkungen in beide Richtungen anzunehmen sind. Allerdings liegen für Deutschland hierzu bislang kaum aussagefähige Befunde vor. Beispielsweise gelang es Fend nicht, für seine überwiegend in den 70er-Jahren erhobenen Daten zu Schulklima, Schulkultur und Schulzufriedenheit Relationen mit Leistungskriterien abzusichern (Fend, 1998). In der PISA 2000 E-Studie korrelieren auf der Schulebene (n= 1472) die mittlere Zufriedenheit der Schüler/innen und das mittlere Ergebnis im Lesetest zu .42, bei Kontrolle von Schulform, sozialem Hintergrund und mittleren kognitiven Grundfähigkeiten immerhin noch zu .11, in der Gruppe der Gymnasien (n = 408) sogar zu .44 bzw. zu .18. Diese Werte sind statistisch hoch signifikant, beruhen allerdings nur auf der Befragung und Testung von ca. 45 Schülerinnen und Schülern pro Schule. Immerhin deuten sie an, dass auch in Deutschland „harte“ und „weiche“ Kriterien der Ergebnisqualität von Schulen sich tendenziell stützen, aber keinesfalls deckungsgleich oder redundant sind.

c) Beziehung zwischen Prozessen und Ergebnissen

Schliesslich lassen sich Ergebnisqualitäten – seien es Leistungsdaten oder „weiche“ Kriterien – mit Prozessqualitäten in Beziehung setzen. Dass bei Untersuchungen einer grösseren Zahl von Schulen solche Beziehungen nachweisbar sind, gehört zum gesicherten und für die Praxis der Schulentwicklung wesentlichen Kenntnisstand der internationalen Schulforschung (Fend, 1998; Scheerens & Bosker, 1997; Steffens & Bargel, 1993). Auch hier ist allerdings die Datenlage in Deutschland noch unbefriedigend. Da Schulstruktur, organisatorische und professionelle Voraussetzungen in Deutschland spezifisch geartet sind, können die internationalen Befunde nicht einfach übertragen werden. Leider konnte PISA 2000 hierzu wenig beitragen, weil dort nur wenige schulbezogene Merkmale erfasst wurden, zu denen meist auch nur das Urteil des Schulleiters vorlag. Klieme und Rakoczy (2005) konnten immerhin zeigen, dass Disziplinprobleme (ein Indikator für geringe Prozessqualität einer Schule) in vielen Staaten, so auch in Deutschland, und zumal in Gymnasien, mit geringeren Leistungen einhergehen. Die PISA-Studie hat zudem nochmals deutlich gemacht, dass „Effekte“ von Schulen in Deutschland massiv durch die Schulform moderiert sind (Baumert et al., 2005).

Disziplinprobleme gehen mit geringen Leistungen einher.

Für die Einzelschule könnte eine auffällige Diskrepanz zwischen Prozess- und Ergebnisqualität beispielsweise darauf hinweisen, dass die Schule ihre Potenziale nicht ausreichend nutzt.

5.2 Datenbasis

Nachdem die Ergebnisse der zweiten Staffel des hessischen Mathematik-Wettbewerbs im Internet publiziert waren, wurde im Frühjahr 2004 am DIPF geprüft, welche der Schulen des Sekundarbereichs, deren Wettbewerbsergebnisse verfügbar waren, in einem der vergangenen Jahre an den Arbeitsplatzuntersuchungen bzw. Entwicklungsbilanzen teilgenommen hatten. Insgesamt gelang bei 139 Schulen eine solche Zuordnung. Für 26 davon lagen neben den Lehrer/innendaten auch Ergebnisse von Schüler/innenbefragungen vor. Alle nachfolgenden Aussagen stützen sich auf die Kombination der im Internet veröffentlichten Kennwerte des Mathematik-Wettbewerbs 2004 mit Befragungsergebnissen, die seit Herbst 1998 durch das DIPF gewonnen worden waren. Die Interpretation unserer Befunde muss daher berücksichtigen, dass zwischen Befragung und Leistungsmessung unterschiedliche Zeitintervalle lagen, und dass die Stichprobe von 139 Schulen in keiner Weise repräsentativ ist für die hessischen Se-

Trotz methodischer Einschränkungen kann das Potenzial einer integrierten Qualitätssicherung ausgelotet werden.

kundarschulen. Dennoch scheint uns eine Zusammenführung der Daten interessant und aussagefähig, um das Potenzial einer integrierten Qualitätssicherung auszuloten.

Für jede Schule wurden die Ergebnisse der dort vertretenen Bildungsgänge getrennt im Internet berichtet. Die Bildungsgänge sind in der Tat nicht vergleichbar, können also prinzipiell nicht gemeinsam verrechnet werden. Im Folgenden verwenden wir, sofern Leistungsergebnisse einbezogen sind, ausschliesslich die Ergebnisse zum Aufgabenset A, d.h. zum gymnasialen Bildungsgang, wobei neben 34 Gymnasien auch 21 integrierte und 37 kooperative Gesamtschulen einbezogen sind.

Einschränkend ist schliesslich festzuhalten, dass die Berechnungen allein auf der Schulebene durchgeführt werden können, da a) Leistungsdaten nur aggregiert vorliegen und b) bei Tests und Fragebogenerhebungen unterschiedliche Schüler/innengruppen untersucht wurden. Auch eine Kontrolle der Input-Bedingungen ist hier noch nicht möglich.

5.3 Ergebnisse

Vergleich von Lehrer/innen- und Schüler/innenperspektive

In den Pädagogischen Entwicklungsbilanzen werden Schüler/innen und Lehrer/innen anhand derselben Fragen um ihre Einschätzung zum Thema „Aggression unter den Schülerinnen und Schülern“ gebeten. Wie Abbildung 3 anhand jener 26 hessischen Sekundarschulen belegt, aus denen bislang vollständige Ergebnisse (Lehrer/innen- und Schüler/innendaten sowie Mathematikleistungen) bekannt sind, stimmen die beiden Perspektiven im Vergleich zwischen den Schulen überraschend genau überein. (Die Korrelation ist mit $r = .75$ hoch signifikant, $p < .001$.)

Mehrperspektivische Einschätzungen zum Thema „Aggression unter den Schülerinnen und Schülern“

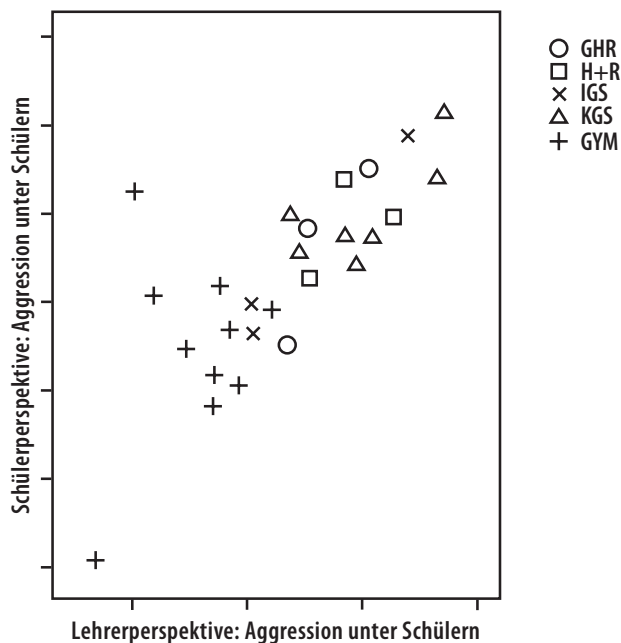


Abbildung 3: Vergleich von Lehrerpersonen und Schüler/innenwahrnehmungen in Bezug auf Aggressionen unter Schülerinnen und Schülern. Basis: Mittlere Schüler/innen- und Lehrer/innenurteile an 26 hessischen Sekundarschulen unterschiedlichen Typs (GHR: Grund- Haupt- und Realschulen; HRS: Haupt- und Realschulen; IGS: Integrierte Gesamtschulen; KGS: kooperative Gesamtschulen; GYM: Gymnasium)

¹ Hier und im Folgenden sind die statistischen Signifikanzniveaus folgendermassen bezeichnet: * $p < .05$, ** $p < .01$, *** $p < .001$.

Positiver Zusammenhang zwischen Wohlbefinden in der Schule und Lernergebnissen

Neben der insgesamt hohen Übereinstimmung, mit der sich die Perspektiven wechselseitig validieren, sind für die Praxis die „Ausreisser“ besonders wichtig. Beispielsweise wird die Rückmeldung an jenes Gymnasium, das in Abbildung 3 „links oben“ zu finden ist, besonderes Gewicht legen auf die Frage, warum denn hier die Schüler/innen von einem recht hohen Niveau der Aggressionen berichten, das die Lehrkräfte gar nicht wahrnehmen. Solche Kontrastierungen eignen sich besonders gut als Einstieg in die schulinterne Reflexion. In den Schulen „rechts oben“ wird man – auch unter Berücksichtigung von Schulform-Unterschieden – ein offensichtlich extrem hohes Aggressionsniveau konstatieren und zum Anlass von pädagogischen Interventionen machen.

Mathematikleistungen und andere Ergebniskriterien

Abbildung 4 belegt, dass erwartungsgemäss auch in hessischen Sekundarschulen ein Zusammenhang besteht zwischen subjektiven Qualitätskriterien (hier: Angaben der Schüler/innen, inwieweit sie sich in der Schule wohl fühlen) und Lernergebnissen (hier: Resultate des Mathematikwettbewerbs für den gymnasialen Bildungsgang). Die Korrelation ist mit $r = .44$ statistisch signifikant, und dies liegt nicht nur an Schulform-Effekten, denn auch innerhalb der Gymnasien und der Kooperativen Gesamtschulen besteht jeweils ein solcher Zusammenhang, der nur aufgrund der geringeren Fallzahl nicht statistisch auffällt.

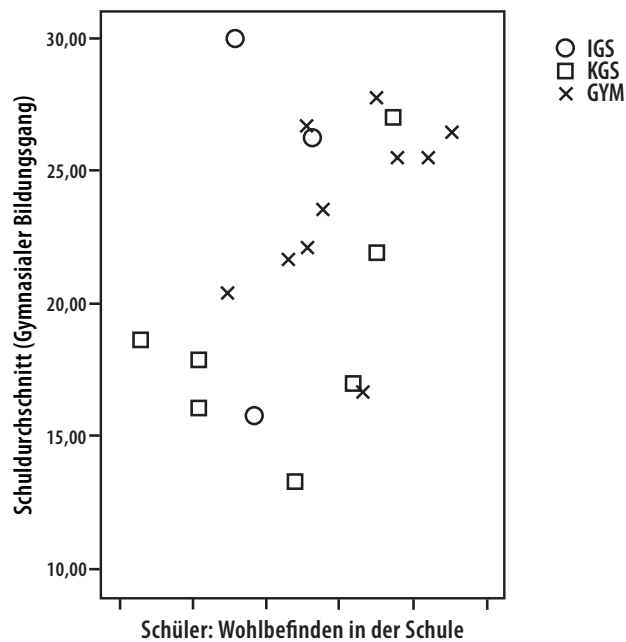


Abbildung 4: Beziehung zwischen subjektiven Qualitätsurteilen (hier: durchschnittliche Zustimmung der Schülerinnen und Schüler in der Skala „Wohlbefinden in der Schule“) und Leistungskriterien (hier: Ergebnis des Mathematik-Wettbewerbs im gymnasialen Bildungsgang); (IGS: Integrierte Gesamtschulen; KGS: kooperative Gesamtschulen; GYM: Gymnasium)

Gleichwohl fallen beide Kriterien nicht zusammen. „Ausreisser“ müssten in den Schulen kritisch diskutiert werden. Zu fragen ist, ob eine solche Schule unterschiedliche Ziele ausgewogen berücksichtigt, aber auch die Gültigkeit der Ergebnisse muss im Einzelfall hinterfragt werden.

Weitere Analysen ergänzen das Bild um die Einschätzung der Ergebnisqualität durch Lehrkräfte. Ausgewählt sind hierfür Aussagen zur Erreichung von Förderzielen. Es

geht also darum, inwieweit es dem Kollegium nach Auskunft seiner Mitglieder gelingt, Schüler/innen individuell zu fördern und zu integrieren. Hier finden sich nun deutliche Schulformeffekte. Die Lehrkräfte der Integrierten Gesamtschulen sehen diese Ziele eher erreicht als ihre Kolleginnen und Kollegen an Kooperativen Gesamtschulen und noch stärker als Gymnasiallehrkräfte. Interessanter noch ist der Befund, dass die Einschätzungen zwischen den Integrierten Gesamtschulen sehr stark streuen und dort ganz klar positiv mit der Leistung im Mathematikwettbewerb zusammenhängen ($r = .55^{**}$), während in den Gymnasien ($r = -.40^{**}$) und den Kooperativen Gesamtschulen ($r = -.17$) der Zusammenhang umgekehrt gerichtet ist. Eine hohe Förderorientierung des Kollegiums scheint demnach in Gymnasien mit relativ niedrigem Leistungsniveau einherzugehen, in Integrierten Gesamtschulen jedoch mit hohem Leistungsniveau (jedenfalls für die Schüler/innen, die dort den gymnasialen Bildungsgang besuchen). Über Ursachen lässt sich hier nur spekulieren: Hängt auch dieser Effekt mit den unterschiedlichen Schüler/innenpopulationen der Schularten zusammen (Auswahl nach Leistungsfähigkeit und sozialer Herkunft)? Geben Gymnasien mit geringerer Förderorientierung ihre schlechten Schüler/innen ab und erzielen so ein höheres Leistungsergebnis? Ist Förderung im Gymnasium eher remedialer Art, greift also nur in Problemlagen? Oder gelingt es Integrierten Gesamtschulen eher, Fördermassnahmen in tatsächliche Leistungsgewinne ihrer „gymnasialen“ Schüler/innen umzusetzen? Alle diese Fragen müssen hier offen bleiben, da uns eine Reihe von Hintergrundvariablen fehlt und da bloße Querschnittserhebungen ohnehin keine Kausalität anzeigen können. Jedenfalls deutet sich hier ein weiterer Beleg für die von Baumert, Trautwein und Artelt (2003) bzw. Baumert und Schümer (2001) vertretene These an, dass die Schulformen in Deutschland sehr unterschiedliche pädagogische Milieus bilden.

Aussagen zur Erreichung von individuellen Förderzielen aus Sicht der Lehrpersonen

Schulformspezifischer Zusammenhang zwischen individueller Förderorientierung und Leistungsergebnissen

Schulformen in Deutschland bilden sehr unterschiedliche pädagogische Milieus

Zusammenhänge zwischen Schulkultur und Ergebnisvariablen

Zunächst soll geprüft werden, ob sich schon innerhalb einer Befragungsperspektive, nämlich nur auf der Basis der Lehrer/innenurteile, aussagefähige Verknüpfungen zwischen Prozess- und Ergebnisvariablen herstellen lassen. Tabelle 1 beschreibt auf der Basis von 139 hessischen Sekundarschulen den Zusammenhang zwischen verschiedenen Aspekten der Schulkultur einerseits und der Einlösung pädagogischer Ziele andererseits, wie sie in den Kollegien wahrgenommen wurden.

Tabelle 1: Zusammenhänge zwischen Aspekten des Schulklimas und der Schulkultur einerseits, erzieherischen und fachlichen Wirkungen andererseits im Lehrer/innenurteil. Basis: Gemittelte Urteile der Lehrkräfte an 139 hessischen Sekundarschulen

Aspekt der Schulkultur aus Sicht der Lehrkräfte	Korrelation (auf Schulebene, n = 34 Gymnasien und 105 andere Sekundarschulen) mit Lehrer/inneneinschätzungen zu ...	
	... Aggressionsniveau unter Schülerinnen und Schülern	... Individuelle Förderung
Unterrichtsbezogene Kooperation	Gymnasium: -.02, n.s. andere: -.10, n.s.	Gymnasium: .45** Andere: .50***
Partizipative Führung	Gymnasium: .03, n.s. Andere: -.30***	Gymnasium: .31* Andere: .22*
Elternbeteiligung	Gymnasium: .02, n.s. Andere: -.68***	Gymnasium: -.06, n.s. Andere: .13, n.s.
Belastung durch problematische Schüler/innen	Gymnasium: .32* Andere: .69***	Gymnasium: .12, n.s. Andere: -.08, n.s.

Geringes Aggressionsniveau geht einher mit einer niedrigeren Problembelastung der Schüler/innen, höherer Elternbeteiligung und positiver Beurteilung der Schulleitung im Sinne einer partizipativen Führung.

Zusammenhang Elternbeteiligung, Problembelastung der Schüler/innen und Leistungen

Die Kombination von Lernerfolgsmessungen mit Befragungen kann eine aussagefähige Basis für die Schulevaluation liefern.

Die Einschätzungen der Lehrkräfte wurden für jede der 139 Schulen gemittelt. Auf der Schulebene ergeben sich sodann, wie Tab. 1 belegt, zum Teil hoch signifikante und bedeutsame Zusammenhänge. Hinsichtlich des Ergebniskriteriums „Aggressionsniveau unter den Schülern“ gibt es kaum Varianz zwischen den Gymnasien, und hier lassen sich auch kaum Zusammenhänge mit Prozessvariablen finden. Bei den übrigen Schulen zeigt sich aber ein klares Bild: Geringes Aggressionsniveau geht einher mit einer niedrigeren Problembelastung der Schüler/innen, höherer Elternbeteiligung und positiver Beurteilung der Schulleitung im Sinne einer partizipativen Führung. Da alle Merkmale als Urteile der Kollegien erhoben wurden, darf man keine Kausalität unterstellen; vielmehr können die Zusammenhänge durch eine mehr oder weniger positive Grundfärbung des Urteils zustande kommen; auch der soziale Hintergrund der Schülerschaft kann diese Zusammenhänge vermitteln. Interessanterweise bleiben die „Effekte“ der Führungsqualität und der Elternbeteiligung aber bestehen, wenn man sie gleichzeitig berücksichtigt und dabei die Problembelastung (die ja auch als grober Input-Indikator angesehen werden kann) kontrolliert.²

Die wahrgenommene Förderung hängt sowohl in Gymnasien als auch in anderen Schulen deutlich mit dem Ausmass der unterrichtsbezogenen Kooperation zusammen, auch wenn die übrigen Merkmale kontrolliert werden².

Als weitere Faktoren können „Elternbeteiligung“ und „Problembelastung der Schüler/innen“ in Relation zu den „objektiven“ Leistungen der Schüler/innen, gemessen am Mathematik-Wettbewerb, gesetzt werden. In allen drei Schulformen korreliert die Elternbeteiligung positiv mit den Leistungen ($r = .61^{**}$ bei IGS, $.39^{**}$ bei KGS, $.21$ bei Gymnasien), die Problembelastung hingegen negativ ($r = -.47^*$ bei IGS, $-.56^*$ bei KGS, $-.25$ bei Gymnasien). Bei Kooperativen wie auch bei Integrierten Gesamtschulen werden die Beziehungen statistisch signifikant. Berücksichtigt man beide Merkmale gleichzeitig², so bleibt für die Integrierten Gesamtschulen, zwischen denen die Werte besonders breit streuen, ein signifikanter „Effekt“ der Elternbeteiligung sichtbar.

Die Befunde zum Zusammenhang von Schulkultur und Ergebnisvariablen stützen die Aussage der Schulforschung, dass Führungsqualitäten der Schulleitung, Kooperation der Lehrkräfte und Elternbeteiligung wichtige Faktoren der Schulqualität sind.

6. Konsequenz: Leistungsmessung und Befragungen als notwendige Komponenten der Qualitätssicherung

Welche Konsequenzen ergeben sich nun für eine produktive Gestaltung von Evaluation für die Schulentwicklung? Trotz der methodischen Mängel, die dem Charakter des vorliegenden Datenmaterials geschuldet sind, konnte aufgezeigt werden, dass die Kombination von Lernerfolgsmessungen (hier: Mathematik-Wettbewerb des Landes Hessen) und Befragungen unterschiedlicher Personengruppen (hier: Pädagogische Entwicklungsbilanzen) eine aussagefähige Basis für Forschung und Evaluation an Schulen liefern kann. Es lohnt sich, die beiden Untersuchungskomponenten weiter zu entwickeln und – möglicherweise ergänzt um stärker formative Verfahren wie

² Die ergänzenden Regressionsanalysen, bei denen mehrere Merkmale kontrolliert wurden, können an dieser Stelle nicht im Einzelnen dokumentiert werden.

Audits und interaktive, schulinterne Reflexion – in einem integrierten System der Qualitätssicherung und Qualitätsentwicklung zusammen zu führen.

Erst wenn Ausgangs- und Rahmenbedingungen (Input), Prozessmerkmale auf Schul- und Unterrichtsebene und Ergebnisse (vor allem schulische Leistungen) zusammengebracht werden, erhält die einzelne Schule das Feedback, das sie braucht, um sich als „lernende Organisation“ weiter zu entwickeln. Dabei ist die Verbindung von internen und externen Evaluationen besonders fruchtbar, wobei die Qualitätskriterien breit gefasst werden sollten (neben kognitiven auch nicht-kognitive „Ergebnisse“ sowie Prozessmerkmale). Die Bewertung der erfassten Daten ist dabei an sozialen, ipsativen und vor allem kriteriumsorientierten Maßstäben unter Berücksichtigung von Rahmenbedingungen und Eingangsvoraussetzungen der untersuchten Schüler/innen vorzunehmen. Wichtige Standards für die Durchführung von Schulevaluation finden sich dazu bei den *Standards for Educational Accountability Systems* (Baker et al., 2002; Baker & Linn, 2004). Fasst man die Daten solcher Erhebungen schulübergreifend in anonymisierter Form zusammen, so lässt sich zudem „Systemwissen“ generieren, das sowohl den Praktikern als auch der Bildungspolitik und -administration wichtige Hinweise gibt und in ein systembezogenes Bildungsmonitoring einfließen kann.

Die Konturen eines integrierten Systems der Qualitätssicherung und -entwicklung sind in den Ländern der Bundesrepublik Deutschland noch sehr unscharf. Reflektiert man aber die eigenen Erfahrungen und bezieht die Erfahrungen anderer Staaten ein (vgl. Scheerens et al., 2003), so zeichnen sich Strukturen und Prozesse ab, die es ermöglichen könnten, Irrwege anderer Staaten – z.B. ein überbordendes Testprogramm oder ein blindes Vertrauen auf Marktmechanismen – zu vermeiden. Deutlich ist auch, dass eine Vielzahl von Personen und Institutionen gefordert sind – von den Lehrkräften und Leitungen in den Schulen über Berater/innen und wissenschaftlich geschulte Evaluatorinnen und Evaluatoren in den Unterstützungsagenturen bis zu Entscheidungsträgerinnen und Entscheidungsträgern in Ministerien –, die mit hoher Professionalität zusammenwirken.

Das Zusammenführen von Input, Prozess und Output hat grosses Potenzial für die Qualitätsentwicklung.

Breit gefasste Evaluationskriterien sowie die Bewertung an sozialen, ipsativen und kriteriumsorientierten Maßstäben sind notwendig.

Mit einem integrierten System der Qualitätssicherung könnten Irrwege vermieden werden.



Prof. Dr. Eckhard Klieme
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Schloßstraße 29
60486 Frankfurt am Main

Klieme@DIPF.de

Literaturverzeichnis

- Abs, H. J. & Döbrich, P. (2004). Evaluation and Quality of the Individual School. In: Döbert, H., Klieme, E. & Sroka, W. (Hrsg.), *Conditions of School Performance in Seven Countries. A Quest for Understanding the International Variation in PISA Results* (S. 366 – 372). Münster: Waxmann.
- Abs, H.J., & Klieme, E. (im Druck). Standards für schulbezogene Evaluation. Erscheint in *Beiheft 2005 der Zeitschrift für Erziehungswissenschaft*.
- Altrichter, H. (1999a). Selbstevaluation von Schulen als Strategie der Qualitätsentwicklung. Begriffe, Erwartungen und Erfahrungen. In: *Evaluation und Qualität im Bildungswesen. Problemanalysen und Lösungsansätze am Schnittpunkt von Wissenschaft und Bildungspolitik. Dokumentation eines internationalen Workshops in Blumau/Steiermark*, 18.-21. Februar 1999. Graz.
- Altrichter, H. (1999b). Anmerkungen zur Diskussion mit Detlev Leutner. In: J. Thonhauser & J.-L. Patry (Hrsg.), *Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog* (S. 135 – 136). Innsbruck, Wien: Studien Verlag.
- Baker, E. L./Linn, R. L. (2004). Validity Issues for Accountability systems. In: S.H. Furman & R.F. Elmore (Hrsg.), *Redesigning Accountability systems for Education* (S. 47 – 72). New York, London: Teachers College Press.
- Baker, E. L., Linn, R. L., Herman, J. L. & Koretz, D. (2002). Standards for Educational Accountability Systems. In: *CRESSST - National Center for Research on Evaluation, Standards and Student Testing. Policy Brief 5*.
- Baumert, J. & Schümer, G. (2001). Schulformen als selektionsbedingte Lernmilieus. In: J. Baumert et al. (Deutsches Pisa-Konsortium) (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 454 – 467). Opladen: Leske und Budrich.
- Baumert, J., Trautwein, U., Artelt, C. (2005). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In: J. Baumert et al. (Deutsches Pisa-Konsortium) (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261 – 352). Opladen: Leske und Budrich.
- Beywl, W. (1987). *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Frankfurt a.M.
- Böttcher, W. (2002). *Kann eine ökonomische Schule auch eine pädagogische sein? Schulentwicklung zwischen Neuer Steuerung, Organisation, Leistungsevaluation und Bildung*. Weinheim und München: Juventa.
- Büeler, X. (2004). Qualitätsevaluation und Schulentwicklung. In: R. Stockmann, R. (Hrsg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, 2.Aufl.* (S. 259 – 286). Opladen: VS Verlag.
- Chelimsky, E. (1997). Thoughts for a New Evaluation Society. *Evaluation*, 5, S. 97-118.
- Dalin, P. & Rolf, H.G. (1990). *Institutionelles Schulentwicklungsprogramm*. Soest: Landesinstitut für Schule und Weiterbildung.
- Ditton, H., Arnold, B. & Bornemann, E. (2002). Entwicklung und Implementation eines extern unterstützten Systems der Qualitätssicherung an Schulen – QuaSSU. In: *Bildungsqualität von Schule. 45. Beiheft der Zeitschrift für Pädagogik* (S. 374 – 389). Weinheim und Basel.
- Ditton, H. & Arnold, B. (2004). Wirksamkeit von Schülerfeedback zum Fachunterricht. In: J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule* (S. 152 – 170). Münster: Waxmann.
- Döbrich, P. (2005). Pädagogische Entwicklungsbilanzen (PEB). *Pädagogische Führung*, 14 (1), S. 27 – 30.
- Donmoyer, R. (2001). Paradigm Talk Reconsidered. In: V. Richardson (Hrsg.), *Handbook of research on teaching* (S. 174 – 197). Washington: American Educational Research Association.
- Dubs, R. (1999). Im Dialog: Durchführung und praktische Folgen: Externe Evaluation. In: J. Thonhauser & J.-L. Patry (Hrsg.), *Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog* (S. 153 – 171). Innsbruck, Wien: Studien Verlag.
- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim und München: Juventa.
- Fend, H. (2002). Mikro- und Makrofaktoren eines Angebot-Nutzungsmodells von Schulleistungen. Zum Stellenwert der Pädagogischen Psychologie bei der Erklärung von Schulleistungsunterschieden verschiedener Länder. *Zeitschrift für Pädagogische Psychologie*, 16, S. 141-149.
- Fuhrman, S.H. & Elmore, R.F. (2004) (Hrsg.). *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press.
- Klieme, E. (2004). Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde. *Zeitschrift für Pädagogik*, Heft 5/2004, S. 625-634.

- Klieme, E., Baumert, J. & Schwippert, K. (2000). Schulbezogene Evaluation und Schulleistungsvergleiche – Eine Studie im Anschluss an TIMSS. In: H.G. Rolff, W. Bos, K. Klemm, K. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung, Band 11*. München: Juventa.
- Klieme, E. & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiss (Hrsg.), *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 353-359). Opladen: Leske + Budrich.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Volmer, H.J. (2005): *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: BMBF.
- Kromrey, H. (2000). Qualität und Evaluation im System Hochschule. In: R. Stockmann (Hrsg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder* (S. 253 – 258). Opladen: VS Verlag.
- Posch, P. (1999). Im Dialog: Durchführung und praktische Folgen: Interne Evaluation. In: J. Thonhauser & J.-L. Patry (Hrsg.), *Evaluation im Bildungsbereich. Wissenschaft und Praxis im Dialog* (S. 139 – 152). Innsbruck, Wien: Studien Verlag.
- Rolff, H.-G. (1998). Schulprogramm und externe Evaluation. In E. Risse (Hrsg.), *Schulprogramm – Entwicklung und Evaluation* (S. 254 – 266). Neuwied: Luchterhand.
- Rost, J. (2000). Allgemeine Standards für die Evaluationsforschung. In: W. Hager, W., Patry, J.-L. & Brenzing, H. (Hrsg.), *Evaluation psychologischer Interventionsmassnahmen. Standards und Kriterien: Ein Handbuch* (S. 129-140). Bern, Göttingen, Toronto, Seattle: Huber Verlag.
- Sanders, J.R. & Davidson, E.J. (2005). A Model for School Evaluation. In: T. Kellaghan & D.L. Stufflebeam (Hrsg.), *International Handbook of Educational Evaluation. Part Two: Practice* (S. 806-826). Dordrecht: Kluwer.
- Scheerens, J., Bosker, R. (1997). *The foundations of educational effectiveness*. New York: Pergamon.
- Scheerens, J., Glas, C. & Thomas, S. M. (2005). *Educational evaluation, assessment, and monitoring. A systematic approach*. Lisse: Swets & Zeitlinger.
- Scriven, M. (2005). Evaluation Theory and Metatheory. In: T. Kellaghan & D.L. Stufflebeam (Hrsg.), *International Handbook of Educational Evaluation. Part One: Perspectives* (S. 15 – 50). Dordrecht: Kluwer.
- Steffens, U. & Bargel, T. (1995). *Erkundungen zur Qualität von Schule*. Neuwied: Luchterhand.
- Steinert, B., Gerecht, M., Klieme & E., Döbrich, P. (2003). *Skalen zur Schulqualität: Dokumentation der Erhebungsinstrumente . ArbeitsplatzUntersuchung (APU), Pädagogische Entwicklungsbilanzen (PEB)*. Frankfurt a. M.: GFPP.
- Steinert, B. & Klieme E. (2004). Was kommt mit der Einführung der KMK-Bildungsstandards auf die Schulen zu? *Schulverwaltung*, 8 (2), S. 56 – 59.
- Stufflebeam, D. L. (2005). Institutionalizing Evaluation in Schools. In: T. Kellaghan & D.L. Stufflebeam (Hrsg.), *International Handbook of Educational Evaluation. Part Two: Practice* (S. 775-806). Dordrecht: Kluwer.
- Watermann, R. & Stanat, P. (2004). Schülerrückmeldungen in PISA 2000: Sozialnorm- und kriteriumsorientierte Rückmeldeverfahren. *Empirische Pädagogik*, 18 (1), S. 40-61.
- Watermann, R., Stanat, P., Kunter, M., Klieme, E. & Baumert, J. (2005). Möglichkeiten und Grenzen der Nutzung von Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen: Das Disseminationskonzept von PISA 2000. In: Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 393-409). Opladen: VS Verlag.
- Wesseler, M. (1999). Evaluation und Evaluationsforschung. In: R. Tippelt (Hrsg.), *Handbuch Erwachsenenbildung/Weiterbildung* (S. 671-687). Opladen: VS Verlag.

Versuch einer Zusammenschau: Bei der Übergabe des Steuers in voller Fahrt das Ziel nicht aus den Augen verlieren!

Silvia Grossenbacher

Das Motto der Tagung „Schule wohin?“ stellt die Frage nach der Zielorientierung von Schule. Zusammen mit dem Untertitel „Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert“ lässt sich interpretieren, dass es darum geht, zu umreissen, wie das Steuer im Bildungswesen gestaltet werden soll und in welche Richtung es auszurichten ist. Im folgenden Beitrag wird versucht, einige der im Laufe des Symposiums vorgetragenen Hinweise resümierend hervorzuheben und in einen Zusammenhang zu bringen.

Veränderung der
Steuerungskonzepte

Auf den ersten Teil der Frage nach den Gestaltungsmöglichkeiten der Steuerung im Bildungswesen gibt das Symposium viele Antworten. Helmut Fend schildert in seinem Beitrag eindrücklich und reich illustriert, wie sich die Steuerungskonzepte verändert haben: von einem politico-administrativen Muster, in dem über rechtliche Vorgaben und entsprechend zugeordnete Ressourcen gesteuert wird, hin zu einem Muster der Orientierung an Ergebnissen. Damit geht es nicht mehr (nur) darum, die rechten Dinge zu tun, sondern auch darum, die Wirksamkeit des Tuns zu überprüfen, um den rechtlich vorgegebenen Auftrag bestmöglich zu erfüllen. Als Ansatzpunkt zur Überprüfung bieten sich Messungen der Leistungen von Schülerinnen und Schülern an, die objektiver sind als die von Lehrpersonen in ihren Schulklassen vorgenommenen Benotungen. Von der Bildungsforschung seit langem eher im Stillen verfolgt, hat diese Strategie in der Bildungspolitik heute einen zentralen Stellenwert. Auch die Stärkung der einzelnen Schulen durch den Zyklus von Schulentwicklung und Evaluation oder die Professionalisierung der Tätigkeit von Lehrpersonen und deren regelmässige Bewertung sind Teil des neuen Systems der Qualitätssicherung.

Neue Steuerungsformen forschungsgestützt einsetzen

Bei der Umsetzung neuer
Programme ist auf die
Anschlussfähigkeit zu achten.

Was hier geprüft und optimiert wird, ist die Effektivität der eingesetzten Mittel. Dagegen ist nichts einzuwenden. Wie Helmut Fend betont, sind mehr Transparenz über die Leistungen des Bildungswesens, die Stärkung der Professionalität von Schule und Lehrpersonen wie auch die Erneuerung des schulischen Programms und dessen Anpassung an gesellschaftliche Erfordernisse (Sprachkompetenzen, naturwissenschaftliche Kenntnisse) und Erkenntnisse der Forschung (Lernmöglichkeiten in der frühen Kindheit, Frühförderung) durchaus plausible Massnahmen zur Qualitätsverbesserung. Sollten die angestrebten Reformen allerdings vom System angenommen und produktiv umgesetzt (statt absorbiert und unschädlich gemacht) werden, so müssten sie *anschlussfähig* sein. Im Vergleich zwischen alten Regelwerken oder Grammatiken von Schule und neuen Formen der Regulierung macht Helmut Fend die Probleme sichtbar, die auftreten können beim Versuch, eine neue „Grammatik“ durch Reformen

zu installieren. Er warnt insbesondere vor voreiligen Teilimporten aus anderen (z.B. bei PISA besser abscheidenden) Systemen, da eine solche Übernahme einzelner Elemente das Bedingungsgefüge von Regelwerken missachtet. Und er warnt vor dem Spardruck, unter dem die Politik geneigt ist, mit Kosten senkenden Regularien und Kontrollen einzugreifen und die Steigerung der Effizienz durch Standarderhöhung bei gleichzeitiger Ressourcenverknappung zu verlangen.

Auf den Verwertungsprozess achten

Helmut Fend schreibt der Bildungsforschung bei der Installation neuer Steuerungsformen eine wichtige Rolle zu. Sie sollte die Auswirkungen von Qualitätssicherung über Bildungsstandards, Vergleichsarbeiten, zentrale Tests usw. überprüfen, insbesondere auch im Hinblick auf die optimale Rückmeldekultur von solchen Ergebnisprüfungen. Je nach Verwertung der Ergebnisprüfungen (Abstrafung vermeintlich Verantwortlicher *oder* tatkräftige Unterstützung jener, die Ziele nicht erreicht haben) werden die Wirkungen sehr unterschiedlich sein. In diesem Sinne plädiert Helmut Fend dafür, die Nutzungsprozesse genau zu beobachten und die diesbezügliche Forschung zu intensivieren.

Die Verwertung der Ergebnisprüfung kann sehr unterschiedliche Wirkungen erzielen.

Qualitätskontrolle ist gut, Qualitätsentwicklung ist besser

Hier wäre allerdings zu fragen, ob die nachfolgende oder begleitende Forschung tatsächlich genügt oder ob es nicht nötig wäre, schon vorher (also bei der Planung und Vorbereitung von Entscheidungen über Steuerungsmodelle und Verbesserungsmaßnahmen) bereits vorliegende Forschungsergebnisse stärker miteinzubeziehen. Denn gerade in Zusammenhang mit der Entwicklung von Leistungsmessungen und Standardsetzung beschäftigen sich viele in der Schulentwicklung tätige Frauen und Männer mit der Frage, die Helmut Fend in seinen Ausführungen ebenfalls gestreift hat: Zielt die derzeit beobachtbare Konzentration auf Leistungsmessung nicht am Kern der Sache, am Kern des Gebotes der *Qualitätsentwicklung*, vorbei?

Wird Konzentration auf Leistungsmessung dem Gebot der Qualitätsentwicklung gerecht?

Die eigentliche Arbeit beginnt erst

Die Befürchtung könnte im Beitrag von Jürgen Baumert Bestätigung finden. An anderer Stelle hat er klipp und klar gesagt, dass die eigentliche Arbeit in der Schule, in den sie unterstützenden Einrichtungen und in der Politik erst nach den Untersuchungen (Leistungsmessungen, Evaluationen etc.) beginne (Baumert, 2001, S. 33). In seinem Referat zeichnete er, datenreich belegt, den von Inglehart prognostizierten „cultural shift“ nach. Er konzentriert sich dabei auf einen Aspekt: den Anstieg des Bildungsniveaus in der Bevölkerung. Im Zuge der Bildungsexpansion reduzierte sich der noch in den 60er Jahren weit überwiegende Anteil der deutschen Hauptschule auf einen „Rest“ zugunsten von Realschule und Gymnasium. Damit einhergehend veränderte sich die Schulbevölkerung in allen Schulformen, entstanden dramatische intergenerationelle Unterschiede im Bildungsniveau und ergab sich allgemein ein signifikanter Anstieg des Bildungsniveaus der Bevölkerung. Es stellt sich die Frage, ob die Menschen auch mehr lernen, wenn sie mehr Zeit in Schulen verbringen. Anhand

Das gestiegene Bildungsniveau kann nicht adäquat für Fachleistungen genutzt werden.

verschiedener Untersuchungen zeigt Jürgen Baumert auf, dass Ingleharts Rechnung (kognitive Mobilisierung) mit Blick auf die Fachleistungen nicht aufging. Offensichtlich – um es auf einen ganz kurzen Nenner zu bringen – gelingt es der Schule nicht, die über die Generationen hinweg beobachtbare und vom Schulsystem mitverursachte Erhöhung kognitiver Potentiale adäquat auch für Fachleistungen nutzbar zu machen. Ursachen sieht der Autor in der verminderten gesellschaftlichen Wertschätzung von Leistung und Schule, im Generationenwechsel beim Lehrpersonal, im Wechsel pädagogischer Überzeugungen, bei der Rekrutierung von Lehrpersonen und der mangelnden Modernisierung ihrer Ausbildung sowie im Ansehensverlust des Lehrberufes durch mangelnde Professionalität. Eine weitere Erklärung sieht Jürgen Baumert darin, dass die verschiedenen Schulformen mit der durch die Bildungsexpansion veränderten Schulbevölkerung, die sich durch verstärkte Heterogenität auszeichnet, nicht klarkommen.

Zentrales Problem: Umgang mit Heterogenität

Damit ist ein Problem angesprochen, das alle im Bildungswesen Tätigen täglich umtreibt: der Umgang mit Heterogenität. Diese Heterogenität widerspricht der auf Homogenisierung und Gleichschritt-Didaktik ausgerichteten „Grammatik“ der Schule in hohem Masse, doch gerade diese Grammatik zu ändern erweist sich als besonders schwierig, wie Tyack und Tobin (1994) aufgezeigt haben und worauf auch Helmut Fend hinweist.

Selbstreflexion stärken statt Differenzen dramatisieren

Hannelore Faulstich-Wieland präsentiert dazu in ihrem Beitrag einen Ansatz, der weit über die Geschlechterfrage in der Schule hinaus relevant ist. Sie liefert einen Schlüssel, der für die Aus- und Weiterbildung von Lehrpersonen und die Schulentwicklung von zentraler Bedeutung ist. Sie plädiert dafür, auf der Basis von Genderkompetenz (und – allgemeiner gesprochen – eines vertieften Wissens um strukturelle Ungleichheiten) die individuelle wie auch die institutionelle Selbstreflexion zu erhöhen.

Entdramatisierung von Geschlecht führt zur Erweiterung des Verhaltensspektrums.

Anhand der Geschlechterfrage in der Schule zeigt Hannelore Faulstich-Wieland, dass der auf Theorien zur Geschlechterdifferenz basierende Ansatz die Gefahr der Festschreibung von Geschlechterstereotypen birgt. Der Ansatz, dass Geschlecht in Interaktionen „konstruiert“ wird (doing gender), erlaubt das Verhalten von Individuen unter dem Aspekt der Geschlechterkonstruktion und ihrer interaktionellen Herstellung zu betrachten. In bisherigen Schulprogrammen zur Mädchen- und Jungenförderung in Deutschland erfolgte eine Dramatisierung von Geschlecht durch geschlechtergetrennte Angebote. Möglicherweise hat diese Strategie das Gegenteil der intendierten Wirkung erzielt, nämlich eine Verstärkung von Geschlechterstereotypen. Dagegen plädiert Hannelore Faulstich-Wieland für eine Entdramatisierung, um statt der Festschreibung von Geschlechterstereotypen eine Erweiterung des individuellen Verhaltensspektrums zu erhalten. Um eine solche Horizonterweiterung in den Geschlechterbildern und Verhaltensweisen zu ermöglichen, müssen Lehrpersonen sich mit Ungleichheiten bzw. Benachteiligungen aufgrund des Geschlechts und mit Geschlechterstereotypen auseinandersetzen und einen subtilen Umgang mit Differenzen finden, ohne diese selbst zu verschärfen oder „herzustellen“. Während Lehrpersonen strukturelle Ungleichheiten, Benachteiligung und die stereotype Festschreibung von Unterschieden also durchaus als dramatisch aufzufassen hätten, wäre die Dramatisierung von Geschlecht im täglichen Umgang mit Kindern und Jugendlichen zu vermeiden.

Wie diese Balance zwischen Dramatisierung und Entdramatisierung im schulischen Alltag umzusetzen ist, dazu bleiben zahlreiche Fragen vorläufig offen. Ganz allgemein kann man aber festhalten, dass ein nicht diskriminierender Umgang mit Heterogenität verlangt, dass Lehrpersonen um Ungleichheiten wissend, sich mit der eigenen Position im sozialen Gefüge auseinander setzen und differenziert auf unterschiedliche Lernvoraussetzungen eingehen können. Dies wiederum ist unter der herrschenden „Grammatik“ der Schule ein uneingelöstes Postulat.

Evaluation für Schulen fruchtbar machen

Auf die verschiedenen Aspekte, die bei Schulevaluationen berücksichtigt werden müssen, wenn sie Schulen nützen und diese stärken sollen, macht Eckhard Klieme aufmerksam. Evaluation ist ein wichtiger Teil der Schulentwicklung und ein Merkmal, das erfolgreiche Schulen auszeichnet. Dabei lassen sich verschiedene Paradigmen feststellen: Evaluation als *Wirkungsforschung*, als *entwicklungsbegleitender* Bewertungsprozess und als *Kontrolle* der Zielerreichung zum Zweck der Beurteilung von Effektivität und Effizienz. Evaluationen dienen als Grundlage für praktische Entscheidungen; die Kombination verschiedener der oben genannten Paradigmen muss ausgehandelt werden. Für die Durchführung von Evaluationen gelten Standards, die kurz umrissen Transparenz, Multiperspektivität durch unterschiedliche Datenquellen, Berücksichtigung der Mehr-Ebenen-Struktur von Schulen, verständliche Publikation der Resultate usw. beinhalten. Als wichtige Anforderung an Schulevaluation nennt Eckhard Klieme deren Bezug zu den Zielen der Einzelschule. Evaluationen nützen den Schulen dann, wenn diese aus dem Feedback lernen können, wenn die Rückmeldung konkret ist, sich auf die Zielformulierung bezieht und ebenso handlungsrelevante wie wertschätzende Empfehlungen macht. Aus Feedback zu lernen scheint allerdings keine einfache Sache zu sein, wie Visscher & Coe (2002) aufzeigen. Laut Klieme gibt es gewisse Faktoren, welche die Offenheit für Qualitätsentwicklung begünstigen; dazu gehören Partizipationserfahrungen im Kollegium und ein einfühlsamer Führungsstil der Schulleitung. Doch Evaluation ist und bleibt eine anspruchsvolle Aufgabe, insbesondere wenn sie als relevant erlebt werden will und zu Veränderungen führen soll. Das setzt, so vermutet Eckhard Klieme, Änderungen im professionellen Selbstverständnis der Lehrpersonen und Schulleitungen voraus und macht eine Veränderung der „Grammatik“ der Schule nötig.

Evaluation ist ein Merkmal erfolgreicher Schulen.

Schulevaluation muss Bezug nehmen auf die Ziele der Einzelschule.

Fazit: Qualität verbessern heisst Benachteiligungen abbauen

Das Symposium hat viel beigetragen zur Klärung von Fragen der Systemsteuerung, Schulentwicklung und Evaluation. Zumindest implizit wurde dabei auch immer wieder auf Unvollendetes, noch Ausstehendes verwiesen, wurden offene Baustellen aufgezeigt.

Auf der Systemebene geht die Verschiebung von Verantwortung hin zu den Schulen (mit Profil, mit erweitertem Gestaltungsspielraum, mit Teilautonomie etc.) einher mit der Einführung von Bildungsstandards, die der Systemevaluation dienen (z.B. Harmons¹). Soll dieses Instrument nicht allein Kontrollfunktion haben, muss es verzahnt

Evaluationselemente müssen miteinander verzahnt werden.

¹ Harmonisierung der obligatorischen Schule; Programm der EDK

Auftragserfüllung von
Schulen muss umfassend
überprüft werden.

werden mit anderen Evaluationselementen, die stärker auf Entwicklungsbegleitung und auf Wirkungsforschung ausgelegt sind. In Zeiten knapper Finanzen steht der Ruf nach Kontrolle jedoch im Zentrum. Viel Geld wird derzeit investiert in den Aufbau eines entsprechenden Instrumentes, und es stellt sich die Frage, ob noch genügend Ressourcen für den ebenso dringenden Ausbau der anderen Elemente übrig bleiben. Auf der Schulebene sind Entwicklung und Evaluation bisher sehr stark als Organisationsentwicklung und -evaluation gesehen und konzipiert worden. Hier muss vermehrt auch der Unterricht ins Spiel kommen (siehe dazu auch EDK, 2004). Dabei genügt wiederum die Teilnahme an internationalen und nationalen Leistungsmessungen nicht. Denn zum einen kann der Auftrag der Schule nicht reduziert werden auf die Vermittlung von Kompetenzen in den Kernfächern. Die darüber hinaus gehenden Aspekte der Auftragserfüllung sollten ebenfalls regelmässig überprüft werden. Zum andern gibt die Erfassung von Leistungen bzw. Leistungsunterschieden noch keine Auskunft darüber, wie die Leistungen zustande kommen und allenfalls verbessert werden können bzw. woher die Leistungsunterschiede rühren und wie sie verringert werden können. Nur vertiefte Kenntnisse über Wirkzusammenhänge im Unterricht ermöglichen eine fundierte Weiterentwicklung dieses Unterrichts.

Sensibilität und
Kompetenz als Basis der
Unterrichtsentwicklung

Unterrichtsentwicklung kann zudem – ähnlich wie Schulorganisationsentwicklung – nur bedingt von aussen angestossen werden. Es bedarf der Bereitschaft der Lehrpersonen und der Schulen als pädagogische Handlungseinheiten zur ständigen Reflexion ihres Tuns, mit dem Ziel, dieses Tun den sich verändernden Verhältnissen, Bedürfnissen, Möglichkeiten anzupassen. Dazu aber brauchen Lehrpersonen und Schulen Sensibilität (beispielsweise für strukturelle Ungleichheiten) und Kompetenz (beispielsweise in Genderfragen). Die Grundlage solcher Sensibilität und Kompetenz ist Wissen. Dieses bereitzustellen ist Sache der Forschung und die muss – darüber lässt das Symposium keinerlei Zweifel zurück – auf allen angesprochenen Ebenen ausgebaut werden. Systemsteuerung und insbesondere auch ein Umbau in dieser Steuerung ist ohne Forschung nicht denkbar. Der FS&S – Forschungsbereich Schulqualität & Schulentwicklung leistet seit 10 Jahren in diesem Sinne unverzichtbare und erfolgreiche Arbeit und es ist ihm für die Zukunft weiterhin viel Erfolg zu wünschen!

Ungleichheiten müssen
abgebaut werden.

Wo die Prioritäten bezüglich der Probleme im Bildungswesen zu setzen sind, darüber entscheidet die Forschung nicht allein. Das ist primär ein politischer Aushandlungsprozess. Gleichwohl hat das Symposium auch dazu einige wichtige Impulse gegeben. Die Qualität von Schule und Unterricht zu steigern, um beispielsweise aus dem PISA-Mittelfeld an die Spitze vorzurücken, ist sicher ein in Politikerkreisen unbestrittenes Ziel. Dass dabei auch die Unterschiede in der Qualität der Angebote verkleinert und die breite Streuung in den Leistungen der Schülerinnen und Schüler verengt werden sollten, ist von Seiten der Forschung klar (siehe auch Terhart, 2000) und am Symposium mehrfach betont worden. Die Forschungsbereiche, die eng mit Bildungsplanung und -verwaltung zusammenarbeiten, haben nicht zuletzt die Aufgabe, aufzuzeigen, weshalb es sich auch politisch durchaus lohnt, Ungleichheiten abzubauen. Und sie sollten auf der Basis ihrer Forschungsergebnisse Wege markieren, auf denen wir dem Ziel „equity“ näher kommen können. Auch darin ist dem FS&S viel Erfolg zu wünschen!

Die Referierenden am Symposium haben aus verschiedenen Blickwinkeln dargelegt, wie schwierig es ist, die „Grammatik“ der Schule zu verändern. Die Erhöhung des äusseren Drucks durch Ressourcenverknappung und die öffentliche Abstrafung vermeintlich Verantwortlicher sind untaugliche Mittel. Eine nachhaltige Verbesserung

der Schulqualität im oben genannten Sinne kann nur das gemeinsame Werk aller Beteiligten sein (Schüler/innen und Eltern, Lehrpersonen, Schulen, Verwaltung und Politik). Um sie auf einen gemeinsamen Weg zu bringen, braucht es zum einen die diskursive Aushandlung von Zielen und zum anderen Anreize und Motivation, wie Helmut Fend zum Schluss seiner Ausführungen deutlich gemacht hat. Möge es FS&S in seiner interdisziplinären Zusammensetzung weiterhin gelingen, diese Überzeugung wirkungsvoll nach aussen zu tragen!



Dr. Silvia Grossenbacher
Schweizerische Koordinationsstelle für Bildungsforschung (SKBF)
Entfelderstrasse 61
5000 Aarau

silvia.grossenbacher@swissonline.ch

Literaturnachweis:

- Baumert, J. (2001). Vergleichende Leistungsmessung im Bildungsbereich. In: *Zeitschrift für Pädagogik*, 45. Beiheft, S. 15-36.
- EDK (Schweizerische Konferenz der kantonalen Erziehungsdirektoren) (2004). *Unterrichtsentwicklung – zum Stand der Diskussion*. Bern: EDK (Studien und Berichte 21).
- Terhart, E. (2000). Qualität und Qualitätssicherung im Schulsystem. In: *Zeitschrift für Pädagogik*, 6, 2000, S. 809-829.
- Tyack, D. & Tobin, W. (1994). The „Grammar“ of Schooling: Why Has it Been so Hard to Change? In: *American Educational Research Journal*, 3, 1994, S. 453-479.
- Visscher, A. J. & Coe, R. (2002). *School Improvement Through Performance Feedback*. Lisse: Swets & Zeitlinger.