

Bildungsevaluation und Bildungspolitik - Pädagogische Ambitionen zwischen Forschung und Politik

Evaluationen im Bildungsbereich werden oft im Auftragsverhältnis durchgeführt, wodurch die Freiheit der Forschung, nicht aber notwendigerweise die Unabhängigkeit, eingeschränkt werden kann. Diese Einschränkung, wenn es denn eine ist, kann aber auch den Realitätsbezug pädagogischer Forschung verstärken, weil der Auftraggeber brauchbare Informationen oder Erkenntnisse haben möchte. Dass Evaluationen jedoch konkrete Probleme lösen sollen, ist in der Regel eine unzutreffende Erwartung. Anhand eines Beispiels versuche ich diesen Zusammenhang zwischen Auftraggeber, Möglichkeiten von und Erwartungen an Evaluationen zu explizieren. – Am Anfang stand der Auftrag:

Der Regierungsrat des Kantons Zürich sah 1996 vor, dass die Gemeinden künftig zwischen Dreiteiliger Sekundarschule (Abteilungen A, B, C)¹ und Gegliederter Sekundarschule (Stammklassen und Niveaus mit erhöhter Durchlässigkeit, früher AVO-Modell genannt)² wählen können. Für den Meinungsbildungs- und Entscheidungsprozess im Kantonsrat, in den Gemeinden und in der Öffentlichkeit sollte eine wissenschaftlich gestützte Informations- und Entscheidungsgrundlage geschaffen werden. Die Erziehungsdirektion des Kantons Zürich hat dafür eine als Vergleichsstudie konzipierte Evaluation in Auftrag gegeben, die Aufschluss über die Leistungen der Schülerinnen und Schüler an der Gegliederten Sekundarschule, an der Dreiteiligen Oberstufe und am Langgymnasium geben sollte. Die Leistungsuntersuchung wurde als externe Evaluation vorgesehen.

Die vergleichende Leistungsmessung sollte in einen erweiterten evaluativen Kontext gestellt, insbesondere als Beitrag zur Schulentwicklung konzipiert werden und sollte Rückmeldungen für die je einzelnen, an der Evaluation beteiligten Schulen bereitstellen. Die Daten sollten somit auch als Selbststeuerung einzelner Schulen genutzt werden können und insofern einen 'top-down support' für eine 'bottom-up reform' ermöglichen. Weiter sollte die Untersuchung nicht als Sanktionsmittel angelegt sein, sondern als konstruktiver Beitrag zur Standortbestimmung der Schulen

1 In der *Dreiteiligen Sekundarschule* werden die Schülerinnen und Schüler in drei Abteilungen unterrichtet. Die Abteilungen A, B, C sind aufgrund der Leistungen der Schülerinnen zusammengesetzt und entsprechen den bisherigen Schultypen Sekundarschule, Realschule und Oberschule. Die Schülerinnen und Schüler werden in der Regel in allen Fächern in der gleichen Klasse unterrichtet. Gegenüber früher wird es bis zu drei mal pro Jahr möglich sein, zwischen den Abteilungen zu wechseln.

2 In der *Gegliederten Sekundarschule* werden die Schülerinnen und Schüler zwei unterschiedlich anspruchsvollen Stammklassen zugeteilt: Stammklassen mit grundlegenden Anforderungen und Stammklassen mit erweiterten Anforderungen. In zwei Fächern (Mathematik, Französisch oder Deutsch) wird der Unterricht in drei Leistungsniveaus erteilt: Niveau mit grundlegenden Anforderungen, Niveau mit mittleren Anforderungen und Niveau mit erweiterten Anforderungen.

durch die Lehrerinnen und Lehrer. Die Resultate der Studie hatten möglichst rasch in den laufenden politischen Meinungsbildungs- und Entscheidungsprozess einzufließen.

Die Vorgaben waren ziemlich komplex, und die zur Verfügung gestellte Zeit war äußerst knapp bemessen. Damit war umzugehen, aber es stellte sich eine zentrale Frage: Kann in eine bildungspolitisch kontrovers geführte Diskussion mit Hilfe wissenschaftlicher Evaluation in der Form eingegriffen werden, dass eine zusammenfassende, weithin akzeptierte Aussage über den Erfolg bzw. Misserfolg oder nur Teilerfolg eines neu erprobten Schulsystems möglich wird? Aufgrund von Vermutungen wird diese Frage eher mit nein beantwortet (AURIN & STOLZ 1990). Es gibt aber auch Ausnahmen, wie diese Evaluation für den Kanton Zürich. Die Frage ließ sich nicht a priori beantworten sondern erst a posteriori.

Ich werde weniger auf Planung, Methoden, Durchführung und Ergebnisse dieser Evaluation eingehen als vielmehr konzeptionelle Überlegungen und Kontexte darstellen, die sich zwar am erwähnt Beispiel orientieren, aber in einem unscharf begrenzten Rahmen auch verallgemeinerbar sein sollen.

Wenn zwei Schulsysteme miteinander verglichen werden, dann ist dies am Ende der neunziger Jahre kaum mehr möglich, ohne sich mit dem Thema Schulqualität auseinanderzusetzen, wobei der bildungspolitische Rahmen eine Rolle spielt. In einem ersten Teil gehe ich auf die Bestimmung von Kriterien von Schulqualität ein (1). Anschließend wird die Evaluationsforschung im Feld der Bildungs-, Schul- und Sozialforschung situiert (2). Abschließend versuche ich das soziale Feld von Evaluationsforschungen zu strukturieren und mit dem Beispiel aus dem Kanton Zürich in Verbindung zu bringen.

1. Kriterien zur Bestimmung von Schulqualität³

Die Frage nach der Qualität von einzelnen Schulen oder von Schulsystemen führt zu grundlegenden Fragen nach den gesellschaftlichen Zielsetzungen, der Art der Mitwirkung an Entscheidungsprozessen auf allen Ebenen des Bildungssystems und den tatsächlichen Zwecken und Zielen von Schule als Institution. Die OECD hat das Problem wie folgt umschrieben:

„Es gibt keinen einfachen, eindimensionalen Maßstab für Qualität. So wie die Definition für hohe Qualität von Bildung multidimensional ist, so gibt es auch kein einfaches Rezept dafür, welche Komponenten für das Erreichen qualitativ hochwertiger Bildung notwendig sind. Viele Faktoren stehen dabei in Wechselwirkung: Schülerinnen und Schüler mit ihrem jeweiligen Hintergrund und ihre Eigenschaften, Lehrerkollegien und ihre Fertigkeiten, Schulen mit ihrer Struktur und ihrem besonderen Charakter, Curricula und gesellschaftliche Erwartungen“ (KARMEL zit. nach OECD 1991, S. 36).

Qualität von Schule lässt sich über Ziele, Aufgaben Erwartungen und Produkte definieren, die je nach Personen- und Interessengruppen unterschiedlich sein können.

³ Teile diese Kapitels sind in MOSER & RHYN (1999) erschienen.

Schule als öffentliche Institution hat aber mehrere explizite und implizite Aufgaben und Funktionen, zum Beispiel die Qualifikation oder die Selektion der Schülerinnen und Schüler (vgl. RHYN 1999). Je nach Funktion, die herausgestrichen oder besonders gewichtet wird, kann die Qualität von Schule anders bestimmt werden.

Die Frage nach der Qualität für wen und der Qualität in bezug auf was ist für ein Bildungssystem, aber auch für die einzelne Schule nicht beliebig und nicht nur von einer Seite zu beantworten. Im öffentlichen Bildungssystem ist dies nur über einen pluralen, möglichst öffentlichen Diskurs überhaupt möglich. Demokratisch verfasste Gesellschaften zeichnen sich auch dadurch aus, dass verschiedene Vorstellungen über die Gestaltung des Schulsystems und über die Qualität von Schulen vertreten und diskutiert werden. Aber auch pluralistische und sich wandelnde Gesellschaften können auf eine verbindliche Orientierung und auf einen – meist aus Kompromissen bestehenden (WINCH 1996) – Grundkonsens darüber, welchen Zwecken Schulen dienen und nach welchen Prinzipien sie gestaltet werden sollen, nicht verzichten. Ohne eine derartige Bezugs- und Orientierungsbasis, die auch gesetzlich verankert und daher verbindlich ist, kann kein Schulsystem funktionieren und kann keine Arbeit in und mit Schulen auf Dauer produktiv sein. Verbindliche und qualitätsrelevante Vorgaben werden in der Schweiz nebst anderem in den Zweckartikeln der kantonalen Verfassungen formuliert und sind vom Souverän, d.h. von der stimmberechtigten Bevölkerung beschlossen worden. Die Wissenschaft kann zwar die Qualität von Schule einigermaßen verlässlich überprüfen, sie kann und soll aber nicht alleine bestimmen, worin diese Qualität zu bestehen hat, ebensowenig wie ausschließlich die Lehrpersonen oder die Eltern.

Die Diskussion um Schulqualität ist im wesentlichen aus den USA importiert worden. Manche der zur Zeit diskutierten Qualitätsmerkmale lassen sich nicht einfach realisieren, vermutlich weil ein Missverständnis vorliegt. In den umfangreichen Forschungen zur Qualität von Schulen in den USA war das Kriterium zur Bestimmung guter Schulen fast ausnahmslos die Schulleistung. Diejenigen Schulen mit besonders guten Leistungen sind näher untersucht worden, wobei Merkmale dieser Schulen festgestellt worden sind, wie beispielsweise 'gemeinsame Planung und kollegiale Beziehungen' oder 'starke Schulleitung'.⁴ Keine der untersuchten Schulen hat aber alle diese Merkmale aufgewiesen und - das scheint mir von entscheidender Bedeutung zu sein - der Umkehrschluss, Schulen mit den gefundenen Merkmalen seien leistungsstark, ist nicht zwingend. Oft werden Merkmale guter Schulen mit Kriterien

⁴ PURKEY & SMITH (1983) nennen folgende Merkmale: *Organisationsbezogen*: Leitung der Schule, Unterrichtsbezogene Führung, Stabilität des Kollegiums, Artikulation und Organisation des Curriculums, Weiterbildung des Kollegiums, Unterstützung Mitarbeit der Eltern, schulweite Anerkennung fachunterrichtlicher Leistungserfolge, wirksam genutzte Unterrichtszeit, Unterstützung durch die regionale Schulbehörde. *Prozessbezogen*: Gemeinsame Planungen und kollegiale Beziehungen, Zusammengehörigkeitsgefühl, klare Ziele und von allen geteilte hohe Erwartungen, Ordnung und Disziplin.

zur Bestimmung von Qualität verwechselt.⁵ Wenn gute Teams in der Industrie, deren Qualität an der Effizienz und Produktivität gemessen werden, sich regelmäßig abends zum Kartenspiel treffen, dann wäre die Umkehrung von Ursache und Wirkung weder zulässig noch plausibel. Ein schlechtes Team wird nicht dadurch gut, dass die Personen sich abends oder am Wochenende zum Kegeln treffen.

In der Diskussion um Schulqualität wird zudem häufig vergessen, dass die Schule in den USA anders organisiert ist als in den meisten europäischen Ländern. Die Klasse spielt in US-amerikanischen Schulen infolge des hohen Grades an Differenzierungsmaßnahmen eine untergeordnete Rolle. Der Einfluss von Unterrichts- und Klassenmerkmalen wurde deshalb in amerikanischen Untersuchungen zum Thema "School Effectiveness" wenig berücksichtigt. Meist wurden nur zwei Ebenen – Schule und Schüler/Schülerinnen – untersucht. Ein großer Teil der Erklärung von Unterschieden in der Schulqualität liegt in der Schweiz, aber auch in anderen europäischen Staaten, bei Klassenmerkmalen.

Schulqualität ist vor allem dann ein sinnvolles Konzept, wenn es sich einigermaßen verlässlich erfassen lässt, wenn es weitgehend auf konsensualen und demokratisch legitimierten Normen basiert und wenn die Kriterien an Merkmalen ansetzen, die von den Schulen beeinflusst werden können.⁶

Wir sind deshalb von einem pragmatischen Konzept von Schulqualität ausgegangen, das wir so nennen, weil es sich nicht am Wünschbaren, sondern am Machbaren orientiert (vgl. dazu RHYN & MOSER 1998). Dabei ist der pragmatische Zugriff ein doppelter: Einerseits werden nur Kriterien aufgenommen, die mess- oder erfassbar sind, andererseits nur solche, die von den Schulen beeinflusst werden können. Damit schließt unser Modell mehr als nur die Schulleistungen und weniger als eine umfassend-ganzheitliche Wunschvorstellung ein, und es lässt sich den methodischen und schulpolitischen Entwicklungen entsprechend um- und ausbauen. Abbildung 1 zeigt, wie sich die in der Evaluation der Oberstufe im Kanton Zürich untersuchten Merkmale zu Kriterien einer Schul- oder besser Unterrichtsqualität zusammenfügen lassen.

5 Den Qualitätsforschungen verwandt sind die Erhebungen von Bildungsindikatoren. Bildungsindikatoren sind Statistiken, die es erlauben, Werturteile über zentrale Aspekte der Arbeitsweise von Bildungssystemen zu treffen. Diese Definition, wie sie von der OECD bei der Erhebung der Bildungsindikatoren verwendet wird, beinhaltet mindestens drei Besonderheiten: Erstens wird davon ausgegangen, dass wir es mit messbaren Eigenschaften von Bildungssystemen zu tun haben. Zweitens, dass keine tieferegreifende Beschreibung, sondern das Messen von als zentral erachteten Aspekten oder Bereichen angestrebt wird. Drittens beinhaltet die Definition die Forderung, dass Indikatoren etwas über die Qualität des Schulsystems aussagen und somit einen Bezugspunkt oder Standards haben, auf dessen Basis Werturteile gefällt werden können (vgl. SCHEERENS 1994, S. 62).

6 In ihren Bildungsindikatoren geht die OECD ebenfalls davon aus, dass diejenigen Indikatoren bedeutsam sind, die sich beeinflussen lassen (vgl. OECD 1994, insbesondere SCHEERENS 1994, S. 64).

Abbildung 1: Drei Kriterien von Schulqualität in einem pragmatischen Modell

Lernerfolg	
Variablen	Qualitätsbestimmung
- Fachleistungen	Die Qualität steigt - mit dem Lernerfolg in den Lernbereichen oder Fächern.

Förderung	
Variablen	Qualitätsbestimmung
- Leistungsbereitschaft	Die Qualität steigt - mit der Steigerung der Leistungsbereitschaft.
- Überforderung	- mit der Verringerung von Über- und Unterforderung.
- Unterforderung	- mit der Steigerung der Individualisierung im Unterricht.
- Individualisierung	- mit der Förderung nach individuellem Bedarf.
- soziale Herkunft	- mit der Verringerung von Ungleichheiten in der Förderung
- Geschlecht	von Kindern unterschiedlicher Muttersprache, sozialer
- Muttersprache	Herkunft, Geschlecht usw.

Beurteilung	
Variablen	Qualitätsbestimmung
Notengebung und	Die Qualität steigt
- Leistung	- mit der Verlässlichkeit der Beurteilung.
- soziale Herkunft	- mit der Vergleichbarkeit der Beurteilung.
- Muttersprache	- mit der Verringerung der ungerechtfertigten Beurteilung
- Geschlecht	aufgrund der Herkunft, der Muttersprache und des Geschlechts.

Lernerfolg

Eine Aufgabe der Schule ist die Vermittlung von Wissen und Können, wie dies in den Lehrplänen festgehalten ist. Lernerfolg in den Fachleistungen ist ein geeignetes Kriterium zur Bestimmung von Schulqualität, wenn die Ausgangsbedingungen und Kontextvariablen berücksichtigt werden. Um die Schulqualität im Leistungsbereich zu bestimmen, ist es unzureichend, die Leistungen der Schülerinnen und Schüler einfach zu messen. Der Vergleich mit anderen Schulen wäre unfair, weil der schulische Kontext die Leistungen erheblich variieren lässt. Verschiedene, nicht-schulische Faktoren sind bekannt, welche die Schulleistungen beeinflussen. Dazu gehören Alter und Geschlecht, die soziale Herkunft, die Intelligenz, die Erstsprache und anderes. Schulen haben in aller Regel standortbedingte Vor- oder Nachteile. Die Schulen können sich ihre Schülerinnen und Schüler nicht aussuchen, deshalb sind diese Einflussfaktoren in aufwändigen statistischen Verfahren zu neutralisieren, damit ein Vergleich in aussagekräftiger Weise möglich ist und von der Qualität eines Unterrichts oder einer Schule gesprochen werden kann.

Der Lernerfolg kann in verschiedenen, aus praktischen Gründen aber kaum in allen Fächern erhoben werden. Einmal untersuchte Fächer sollten über längere Zeit erhoben werden, damit Vergleiche möglich und Qualitätssicherung sowie Qualitätssteigerungen messbar sind. Es ist üblicherweise aus finanziellen und arbeitsökonomischen Gründen kaum möglich, derartige Untersuchungen in allen Schulstufen durchzuführen.

Förderung

Die Förderung der Schülerinnen und Schüler ist eine zentrale pädagogische Aufgabe von Schulen, die kaum umstritten ist. Der allgemeine Begriff Förderung lässt sich spezifizieren in Interesse am Unterricht, Motivation, Selbstvertrauen und die Bereitschaft zur aktiven Teilnahme am Unterricht. Wenn diese feststellbaren Variablen steigen, nimmt die Qualität zu. Diese Komponenten von Förderung lassen sich auch als Leistungsbereitschaft bezeichnen.

Ein weiterer Indikator für eine gute Förderung sind das Maß an Über- und Unterforderung im Unterricht. Gelingt der Lehrperson ein Unterricht, der von den Schülerinnen und Schülern als individualisierend wahrgenommen wird, dann findet auch weniger Über- und Unterforderung statt, und die Unterrichtszeit kann effektiver genutzt werden. Die genannten Variablen für eine optimale Förderung und somit für eine hohe Unterrichtsqualität lassen sich relativ einfach und zuverlässig erfassen.

Damit von einer optimalen Förderung gesprochen werden kann, sollten Kinder unterschiedlicher sozialer Herkunft, Mädchen und Knaben sowie Schülerinnen und Schüler verschiedener Muttersprache gleich gut gefördert werden. Dieses anspruchsvolle Ziel ist schwierig zu erreichen und bleibt eine Herausforderung für das Schulsystem bzw. für jede einzelne Schule oder Klasse als Teil dieses Systems.

Beurteilung

Die Beurteilung von Lernfortschritten gehört zum schulischen Alltag und hat mindestens zwei Funktionen zu erfüllen: Erstens dient sie didaktischen Zwecken, nämlich dem gezielten Unterstützen von Lernprozessen. Zweitens hat die Beurteilung von Leistungen der Schülerinnen und Schüler eine promotionsrelevante Funktion, die sich in Zeugnissen oder ähnlichen Dokumenten niederschlägt. Innerhalb eines Schultyps sollte die Notenvergabe aufgrund eines ähnlichen Leistungsniveaus erfolgen, weil damit unfaire Selektionen, Einteilungen und Berechtigungen vermieden werden könnten. Der Zusammenhang von Note und Leistung ließe sich in verschiedenen Klassen ähnlich gestalten, wenn Vergleichswerte bekannt und zugänglich wären. Die Qualität der Beurteilung von Lernerfolgen ist um so höher je weniger Herkunft und Geschlecht diese Beurteilung beeinflussen.

Weitere Kriterien zur Bestimmung von Schulqualität, wie beispielsweise der Grad der Durchlässigkeit zwischen Schultypen, das Schulklima, die Arbeitszufriedenheit

von Lehrerinnen und Lehrern, der soziale Umgang zwischen den Schülerinnen und Schülern oder die Zufriedenheit der Eltern mit der Schule, könnten ebenfalls zur Bewertung der Schulqualität herangezogen werden. Ob sich diese Kriterien auch in ein pragmatisches Modell integrieren ließen, wäre im je einzelnen Fall zu prüfen. Es wäre aber darauf zu achten, dass die normativen Entscheidungen, welche jeder Kriterienbestimmung vorausgehen, nicht persönlichen Präferenzen entsprechen, sondern gesellschaftlich abgesichert sind, zumindest dann, wenn die Evaluation von externer Seite durchgeführt wird. Im weiteren müssten die Kriterien verlässlich erfassbar und von den Schulen beeinflussbar sein. Zudem ist ein allzu komplexes Modell von Schulqualität nicht anwendbar und wenig hilfreich. Die Veränderungen von politischen Rahmenbedingungen oder von schulhausinternen Verhältnissen usw. sind selbstverständlich möglich, aber auf anderen Ebenen zu bearbeiten, als Evaluationsforschung dies tun kann und soll.

Externe Evaluationsforschung kann nicht alle Aspekte der hochkomplexen Schul- und Unterrichtsrealität erfassen und nicht allen schulhauspezifischen Besonderheiten gerecht werden. Grenzen werden durch beschränkte Arbeitskapazität und beschränkte finanzielle Mittel, aber auch durch methodische Schwierigkeiten und die fehlende Vertrautheit mit lokalen Besonderheiten gesetzt. Sie kann aber Wissen und Informationen bereitstellen, die als Entscheidungsgrundlagen dienen und im besten Fall zur Verbesserung des öffentlichen Schulsystems beitragen können.

2. Schulforschung und Evaluationen im Bildungsbereich

Die Bildungs- oder Schulforschung war und ist oft dem Vorwurf ausgesetzt, dass sie für die Praxis bedeutungslos sei. Allerdings hat nicht jede Bildungsforschung den Anspruch, direkt praxisrelevant zu sein und kann oder soll deshalb nicht mit dem Maßstab 'Praxis' gemessen werden. Schulforschung, soweit sie empirisch betrieben wird, ist sporadisch dem Vorwurf ausgesetzt, dass sie die relevanten Aspekte und Leistungen der Schule nicht erfassen könne oder nicht erfassen wolle. Grundsätzlich ist unter Forschern eine distanzierte, um Objektivität bemühte Haltung gegenüber dem Forschungsgegenstand üblich, während pädagogische Praktiker ihr Engagement, die Subjektivität der Wahrnehmung, die Schwierigkeiten in der Erfassung menschlichen Verhaltens und die Einzigartigkeit jedes Menschen betonen. Der differente Arbeitskontext und die deshalb unterschiedlichen Perspektiven zum gleichen Gegenstand, wie dies für Lehrkräfte und Bildungsforschende zutrifft (oder zutreffen sollte), sind strukturell konfliktbesetzt.

Vorbehalte gegenüber den Methoden empirischer Schulforschung sind daher verständlich. Als Gründe können genannt werden: a) Abwehr gegen Zahlen und Quantifizierungen, die gegenüber Fragen der Qualität oberflächlich und nichtssagend scheinen, b) Unübersichtlichkeit und Widersprüchlichkeit von Befunden, c) Schwierigkeiten mit der Nachvollziehbarkeit von Operationalisierungen und ihrer Aussagefähigkeit

keit, d) Misstrauen gegenüber der methodischen und analytischen Kraft, e) Trivialität der Befunde, f) Abwehr gegen unliebsame Befunde. Je nach Umständen können diese Gründe der Ablehnung von unterschiedlicher Bedeutung sein, lassen sich aber kaum je vollständig ausräumen (vgl. STEFFENS & BARGEL 1993).

Versuche, die „Kluft zwischen Schulehalten und Schulforschung“ (ebd., S. 8) zu überwinden, sind inzwischen verbreitet. Soweit mit dieser Kluft eine gewisse Berührungsangst zwischen Schulforschenden und Lehrenden angesprochen ist, mag diese Zielsetzung zu Recht bestehen. Soweit damit aber eine Annäherung der jeweiligen Aufgaben und eine Angleichung der verschiedenen Tätigkeiten anvisiert wird, sind Bedenken nicht leicht auszuräumen. Lehrerinnen und Lehrer sind oft nicht für die Forschung ausgebildet und Bildungsforscherinnen und Bildungsforscher sind nicht immer ausgebildete Lehrpersonen. Die Perspektiven von Bildungsforschung und Lehre sind zudem verschieden und deshalb werden gleiche Phänomene unterschiedlich eingeschätzt. Diese Differenz ist kein Defizit, weil sie in der Logik der Sache liegt. Andernfalls könnte sowohl an der Qualität der Lehre als auch an der Qualität der Forschung gezweifelt werden. Zudem können beide Parteien von der je anderen Sicht profitieren, und ein Ausgleich der Differenz wäre ein Verlust für beide Seiten. Schließlich geht die moderne Forschung nicht mehr von einem generellen und universalen Wahrheitskonzept aus, womit die Differenz auch forschungslogisch möglich ist.

Ein weiterer Grund für den schwierigen Stand von Schulforschung liegt in der nicht wenig verbreiteten reformistischen Schulkritik, die gerne Schule und Unterricht als genierliche Größe taxiert. Die Lehre in der Schule wird gegen das Leben außerhalb der Schule ausgespielt, wobei das Leben immer siegt und die Schule als defizitär ausgewiesen wird oder, wie Klaus PRANGE formuliert: „Schule und Unterricht sind akzeptabel besonders dann, wenn sie nicht als das erscheinen, was sie sind“ (PRANGE 1995, S. 41). Für die verbreitete Schulkritik gilt „Schule als besonders gut und tüchtig, wenn sie nicht Unterricht, nicht Vermittlung von Fertigkeiten und Kenntnissen ist, wenn sie offen und unscharf in ihren Grenzen ist, wenn sie anders ist“ (ebd.). Diese Haltung hat zur Folge, ob beabsichtigt oder nicht, dass empirische Schulforschung erschwert bis verunmöglicht wird. Historisch betrachtet, bestand eine entscheidende Komponente des sich formierenden öffentlichen Schulsystems in seiner rationalen und empirischen Konzeptionierung (vgl. OSTERWALDER 1995). Nur durch Abgrenzung von religiös-dogmatischer Kontrolle und Ausschluss esoterischer Einflüsse jedwelcher Couleur war es möglich, eine öffentlich-staatliche Schule zu realisieren, die ihre Funktion in einer Demokratie wahrnehmen kann.

Eine Reform kann erst dann empirisch untersucht und evaluiert werden, und also erst dann in nachweislicher Form erfolgreich sein, wenn sie operationalisierbare Ziele hat. Andernfalls kann immer Meinung gegen Meinung, Aussage gegen Aussage stehen, sofern die Reform weitergeführt wird.

Evaluationsforschung, die zu empirisch abgesicherten Erkenntnissen gelangen will, ist ähnlichen Problemen ausgesetzt wie die Schulforschung, und wenn die Evaluationsforschung zum Vermeiden unnötiger Missverständnisse zwischen den beiden

Bereichen Forschung und Lehre beitragen kann, ist dies sicherlich zu begrüßen, eine Differenz wird aber und soll auch aus den genannten Gründen aufrecht erhalten bleiben. Etwas anders als bei der Schulforschung sieht der Praxisbezug bei der Evaluationsforschung aus. Häufig hat sie die Aufgabe und den Anspruch, die Praxis zu beeinflussen, was zuweilen auch zu inadäquaten Erwartungen an eine Evaluation führt.⁷

Evaluationsforschung als Verfahren zur Überprüfung von Wirkungen von Interventionen oder Maßnahmen im schulischen Kontext hat in den letzten Jahren in Europa einen Aufschwung erlebt.⁸ Ähnlich wie in anderen Sektoren kann Evaluation im Bildungsbereich umschrieben werden als die systematische Sammlung von Daten, die als Erfolgs- oder Misserfolgsindikatoren für ein pädagogisches Programm interpretiert werden in der Absicht, die pädagogische Praxis zu verbessern. Folgt man dieser Beschreibung, wie sie etwa bei WOLF (1987) zu finden ist, dann unterscheidet sich die Evaluation von der üblichen Bildungsforschung und Sozialforschung durch ihre Anwendungsorientierung.

Es gibt aber keine plausiblen Argumente, die dafür sprechen, dass die Evaluationsforschung eine eigene Disziplin mit spezifischem und genuinem Methodenrepertoire und eigenem Forschungsgegenstand ist. Evaluationsforschung, die nicht unterhalb wissenschaftlichen Niveaus stehen bleiben soll, wird mit Vorteil konzipiert als Anwendungsvariante wissenschaftlicher Forschungsmethoden auf eine spezielle Gruppe von Fragestellungen (vgl. ROSSI & FREEMAN 1993). Anders formuliert: „Mit dem Begriff *Evaluationsforschung* soll zum Ausdruck gebracht werden, dass Evaluationen wissenschaftlichen Kriterien genügen müssen, die auch sonst für empirische Forschungsarbeiten gelten - eine Position, die keineswegs durchgängig in der Evaluationsliteratur geteilt wird“ (BORTZ & DÖRING 1995, S. 96). Der Kanon empirischer Forschungsmethoden ist der gleiche wie bei der Grundlagenforschung, welche die Überprüfung von wissenschaftlichen Hypothesen und Theorien zur Aufgabe hat, während die Evaluationsforschung eher technologische Theorien überprüft. Es gibt deshalb auch keine festgefügten Forschungsdesigns. Der Begriff 'Evaluation' steht für ein ganzes Bündel von Ansätzen, Methoden und Instrumenten, aus denen je nach

7 Diese Beeinflussung kann sich auch auf die Praxis der Bildungspolitik beziehen, nämlich dann, wenn Evaluationsforschung die Grundlage für bildungspolitische Entscheidungen liefern soll. Dies gilt, obschon bekannt ist, dass sich die Bildungspolitik kaum oder nur schwer von der Bildungsforschung beeinflussen lässt (KÖTTER 1988).

8 Historisch wird die Evaluationsforschung im 20. Jahrhundert in vier bis sechs Phasen unterschieden (vgl. GRÜNDER 1993, S. 31ff.). In den USA wird seit den achtziger Jahren eine dominant kundenorientierte Form der Evaluation, die 'Fourth Generation Evaluation' von GUBA & LINCOLN (1989), propagiert (vgl. dazu RUSSELL & WILLINSKY 1997). So neu wie zuweilen angenommen wird ist das Phänomen der Evaluationsforschung auch im deutschsprachigen Raum nicht: Beispielsweise fand 1978 in Montreux eine vom Europarat, der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren und vom Bundesamt für Bildung und Wissenschaft getragene Arbeitstagung zum Thema „Innovation und Evaluation von Schulversuchen“ statt (vgl. MEYLAN 1981). 1981 wurde ein OECD/CERI-Seminar in Muttenz (Schweiz) durchgeführt, das dem Thema „Evaluation von Innovationen im Bereich der Primarschule“ gewidmet war (vgl. MEYLAN 1983).

spezifischer Aufgabenstellung ein geeignetes Design „maßzuschneidern“ ist (KROMREY 1995, S. 315).

Evaluation im Rahmen empirischer Bildungsforschung lässt sich durch mindestens drei Besonderheiten charakterisieren, die durch den Forschungsgegenstand bedingt sind: Erstens gelten Bildung und Erziehung als integraler und instrumentaler Teil einer strategischen Modernisierung der Gesellschaft. Die Gefahr einer ideologischen oder normativen Belastung von bildungs- und erziehungswissenschaftlichen Studien erreicht deshalb ein hohes Maß. Evaluationsstudien, die sich kritisch mit Wirkungen bestimmter Programme, Interventionen oder Reformen befassen, sind rasch einer politischen Codierung, entweder mit dem Label 'progressiv' oder - bei kritischen Studien - meist 'konservativ' ausgesetzt. Unabhängigkeit und Unvoreingenommenheit der Forschenden, die sich aber der Problematik bewusst sind, sind deshalb wichtige Voraussetzungen, um verlässliche Daten zu erhalten.

Soweit, zweitens, Evaluationsforschung im Bildungsbereich international vergleichend arbeitet, sind die Studien nach relativ gleichförmigen Modellen aufgebaut; ein hoher Grad an Standardisierung der Untersuchungsanlagen ist dabei auszumachen. Das Gegenteil gilt für Evaluationsforschung im Schulbereich, die nicht international vergleichend angelegt ist. Dabei ist anzustreben, Forschungsprojekte zu entwerfen, die auf besondere Verhältnisse und Programme oder Interventionsprojekte adaptiert sind und zugleich internationalen Qualitätsstandards entsprechen.

Und drittens: Während etwa in politikwissenschaftlichen Evaluationsstudien eher eine makroanalytische Perspektive, quantitative Daten und 'objektivistische' Forschungsmodelle favorisiert wurden, orientierte sich die erziehungswissenschaftliche Evaluation vornehmlich an lokalen Projekten und Reformen, ergänzte die quantitativen durch qualitative Methoden, setzte Interviews, Beobachtungen und offene Fragebogen ein, um Einstellungen von Betroffenen zu ermitteln. Eine Kombination beider Forschungsansätze drängt sich im Rahmen von Schulqualitätsforschung auf, ebenso wie bei der Evaluation von externen Projekten, die in schulische Kontexte implementiert werden (vgl. KLÖTI 1995, S. 83 f.).

In der Regel finden Evaluationen als Auftragsforschung statt. Der Auftraggeber ist meistens zugleich der Geldgeber, und die Versuchung, das vom Auftraggeber gewünschte Ergebnis zu suchen und zu finden, ist durchaus vorhanden (so etwa WOTTAWA & THIERAU 1990, S. 27). Die Verflechtung von Erkenntnis und Interesse sind kaum irgendwo so ausgeprägt wie in der Evaluationsforschung (vgl. BRANDTSTÄDTER 1990, S. 224). Deshalb ist es im Sinne einer Qualitätssicherung der Evaluationsforschung bedeutsam und letztlich eine Notwendigkeit, die Untersuchungsberichte nachvollziehbar und vollständig einer Fachöffentlichkeit zugänglich zu machen. Den öffentlichen Fachdiskurs zu suchen, ist eine Möglichkeit, Unabhängigkeit und Qualität der Evaluationsforschung zu sichern. Eine zusätzliche Möglichkeit ist, Evaluationsforschungen durch externe Evaluatoren durchführen zu lassen und von Selbstevaluationen eher abzusehen (BORTZ & DÖRING 1995, S. 98) oder zumindest keinen Wissenschaftsstatus zu reklamieren. Die Form der externen Evaluation ist vom Kanton

Zürich gewählt worden, indem die Evaluation der Sekundarstufe I an die Universität des Kantons Bern vergeben wurde. Die Qualität kann im weiteren durch Meta-Evaluationen festgestellt werden, die mit einem spezifischen Kriterienraster arbeiten (WIDMER 1996). Schließlich bestehen internationale Standards bei der Evaluation von Projekten im Erziehungsbereich, deren Einhaltung für eine Qualitätssicherung grundlegend sind (vgl. SANDERS 1994, dt. 1999).

3. Zum sozialen Kontext von Bildungsevaluationen

Eine Besonderheit der Evaluationsforschung ist, dass sich die an den Programmen, Interventionen, Reformen und Versuchen beteiligten Personen für die Forschung, sowohl für den Prozess als auch für das Ergebnis, interessieren. Forschungslaien, die in einem Reformgeschehen unterschiedliche Funktionen und Interessen haben, interessieren sich plötzlich für Forschung. So entstehen interessierte Interessengruppen. Formal können die verschiedenen Akteure oder Personengruppen eingeteilt werden in: politische Entscheidungsträger, Programmträger, Auftraggeber von Evaluationen, Programmteilnehmer, Programmleitung, Programmmitarbeiter, Programmkonkurrenten, Evaluationsforscherinnen, 'Zaungäste' (z.B. Behörden, Politiker, Presse usw.) und die 'Evaluationsgemeinde' (andere Evaluationsforscher) (vgl. ROSSI u.a. 1988, LÖSEL & NOWACK 1987). Diese verschiedenen Rollenträger und Interessenvertretungen ergeben ein soziales Umfeld, das die Evaluationsforschung vor erhebliche Probleme stellen kann, weil Lobbyismus, interessegeleitete Kritik und Intrigen nicht selten auftreten.

Handelt es sich beim zu evaluierenden Programm um einen Schulversuch oder eine Schulreform an einer staatlichen Bildungseinrichtung, so sind mindestens vier Interessengruppen beteiligt: die Schule, in der die Verwirklichung von pädagogischen Anliegen im Vordergrund steht, die Bildungsverwaltung mit bildungspolitischen Absichten, die an Schulfragen interessierte Öffentlichkeit und schließlich die Wissenschaft, die mit kritischem Blick etwa die Qualität des Unterrichts überprüfen will. WEBER (1992, S. 111) benutzt den Begriff „Arena“, um dieses Zusammen- oder Gegenspiel von Akteuren mit unterschiedlichen Interessen, finanziellen und rechtlichen Mitteln, Machtbefugnissen, sozialen Beziehungen sowie Informationen und Wissen zu bezeichnen.

Bei der Evaluation von Schulversuchen lassen sich die unterschiedlichen Erwartungen der verschiedenen Akteure einigermaßen allgemein festhalten: die Schulbehörden erwarten Auskunft über die Tauglichkeit des Schulversuchs. Sie wollen wissen, ob das Modell sich für eine Generalisierung eignet. Die Schulversuchsleitung erwartet von der Evaluation konkrete Hinweise über Stärken und Vorteile des Versuchs und über Schwächen, Nachteile, Konfliktherde sowie Anregungen, wie diese zu beseitigen sind. Die Lehrerschaft erwartet weitgehend Unterstützung, Bestätigung und Rechtfertigung ihrer Tätigkeiten. Die Eltern erwarten eine Kontrolle des Schul-

betriebs, eine Garantie, dass im Laufe des Versuchs (oder der Reform) nichts geschieht, das die schulische Laufbahn der Kinder gefährdet. Befürworter einer Innovation erhoffen sich in der Regel eine Entwicklung, eine Absicherung, eine Qualifizierung und einen Werbeeffect für die Reform. Gegner hoffen genau auf das Gegenteil: Entwicklungsverhinderung, Verunsicherung, Disqualifizierung, Diskreditierung (vgl. JENZER 1992, S. 29). In dieser Konstellation wird es immer Akteure geben, deren Erwartungen durch eine Evaluation nicht erfüllt werden und die dann dazu neigen, eine Evaluation als mangelhaft, unbrauchbar oder überflüssig zu bezeichnen.

Gerade für im wissenschaftlichen Denken nicht ausgebildete oder nicht geübte Personen ist zuweilen die Unterscheidung von Ursachenzuschreibungen schwierig. Ein Mangel in einer durchgeführten Reform kann auf systematische, konzeptionelle, funktionale, systemische oder personale Ursachen zurückgeführt werden. Häufig ist es auch der Evaluation nicht möglich, klare Kausalitäten aufzuzeigen. In der Rezeption der Ergebnisse wird die Kausalattribution dann unkontrolliert und vorschnell vollzogen.

Die Überprüfung von Schulentwicklungen, Schulversuchen und Schulreformen wird von den Reformbefürwortern teils als Bedrohung, teils als Chance wahrgenommen. Evaluation als Bedrohung ist deshalb ernstzunehmen, weil die Verlässlichkeit der Ergebnisse nicht immer gesichert ist, weil viele Möglichkeiten der Beeinflussung und Manipulation, aber auch von Missverständnissen bestehen. Selbst beim Verfassen von Evaluationsberichten, die sich in der Regel an ein nicht wissenschaftliches Publikum wenden, sind Verzerrungen möglich und kaum vermeidbar (vgl. WOTTAWA 1981, FEND 1981). Evaluation wird dann als Chance erlebt, wenn die Bestätigung und Legitimation der Reform erwartet wird. Ansatzweise sind Steuerungen der Ergebnisse und damit auch der Bedrohungscharakter von Evaluationen bei der Mandatsvergabe möglich.

Die Auftraggeber von Evaluationsstudien verbinden mit ihrem Auftrag mehr oder weniger konkrete Zielsetzungen. Unter Umständen werden gezielt solche Evaluatoren beauftragt, die einer gewünschten Position oder Haltung nahestehen (WOTTAWA 1993, S. 719), oder die Evaluatoreninnen haben bei unerwünschten Ergebnissen mit negativen Reaktionen (z. B. keine weiteren Aufträge, keine Weiterempfehlung) der Auftraggeber zu rechnen (ROSSI u.a. 1988, S. 194). Der Auftraggeber kann auch Einfluss auf die Durchführung, die Personalauswahl und die Berichtlegung ausüben (WOTTAWA & THIERAU 1990, S. 25). Handelt es sich beim Auftraggeber um eine Behörde, können die Eingriffe besonders schwerwiegend sein. In Deutschland wurde der Zugang zum Forschungsfeld Schule reglementiert und beschränkt sowie teilweise Einfluss auf Inhalt, Umfang, Durchführung und Berichtlegung der Forschungsprojekte genommen (FABICH 1993, S. 194 ff.).

Keine Wissenschaft ist wertfrei, was für die Evaluation in besonders hohem Masse zutrifft, weil sie meistens Bewertungen und Empfehlungen für weiteres Handeln abgeben soll. Für die Rolle des Evaluators, der Evaluatorin, sind deshalb vier Rollenbilder entworfen worden, die die Erwartungen steuern können. Die Evaluation als

Faktensammlung legt das Schwergewicht auf das Messen und Erfassen bestimmter Phänomene. Die Evaluation als *Gutachten* bewertet die empirisch festgestellten Fakten. Bei der Evaluation als *Entwicklung* werden die erfassten und bewerteten Maßnahmen als Grundlage für heuristische Anregungen herangezogen. Die Evaluation als *Moderation* versucht Reformabläufe durch Beratung und sozialtechnische Hilfen zu optimieren (vgl. WOTTAWA 1991).

Schulversuche und Schulreformen sollen der Verbesserung des Regelschulsystems dienen. Die Evaluation kann daher nicht nur belegen, dass eine Änderung möglich ist, sie muss auch mit Schulen des Regelsystems vergleichen. Schule gehört zu den größten gesellschaftlichen Institutionen - wie das Gesundheitswesen oder das Militär - mit entsprechend großem Finanzaufwand. Ein derartiges Unternehmen kann sich auf Dauer nicht von Überprüfungen dispensieren, ohne unter Legitimationsdruck zu geraten. Hier hat Evaluation (und Schulforschung) eine Funktion, die der Optimierung des Systems dient, ohne normativen Druck aufzusetzen.

Die Frage, ob die Ergebnisse von Evaluationsforschungen von den Betroffenen, der Öffentlichkeit und den Reformträgern eher wahrgenommen und akzeptiert werden als solche der Bildungs- bzw. Schulforschung, lässt sich kaum beantworten, weil insbesondere bei der Evaluation von Einzelschulen oder kleineren Schulversuchen die personelle und bildungspolitische Situationsabhängigkeit sehr groß ist. Wenn die Evaluationsforschung ähnlich eingeschätzt würde wie die Bildungsforschung, nämlich dass sie allenfalls dazu taugt, „ex post empirisch zu belegen, dass Reformen nicht stattgefunden haben“ (ROEDER 1983, S. 82), dann wäre mit Peter Martin ROEDER zu antworten, dass sie immerhin dies kann.

Evaluationen, die empirisch nachweisen, dass reformierte Schulen die Zielsetzungen nicht in erwartetem Masse erreichen, können Anlass sein, die Argumente und Schlagworte der Reformdiskussion mit noch größerer Leidenschaft und Vehemenz aufzunehmen, oder sie können die Reform und ihre Fortführung gefährden. Oft geschieht beides zugleich, wobei politische, wirtschaftliche und soziale Konstellationen weitgehend bestimmen, welche Position sich durchsetzt. Über den Erfolg einer Evaluation entscheidet meistens das Umfeld. Eine zwar erfolgreiche Evaluation ist letztlich dennoch als gescheitert einzustufen, wenn sie und ihre Ergebnisse nicht wahrgenommen werden. Der Berichtlegung und der Art der Verbreitung der Ergebnisse kommt deshalb eine wichtige Bedeutung zu, die auch für den Erfolg oder den Misserfolg maßgebend sein kann (vgl. SANDERS 1999).

Mit der häufig an eine Evaluation herangetragenen Erwartung, direkte Handlungsanweisungen zu geben oder gar Entscheidungen zu treffen oder wenigstens vorwegzunehmen, ist vorsichtig umzugehen. Einerseits könnten zumindest externe Evaluatoren für die Folgen ihrer Entscheidungen kaum zur Verantwortung gezogen werden und andererseits soll über das öffentliche Bildungswesen öffentlich und politisch, im Minimum demokratisch kontrolliert, entschieden werden. Evaluation kann im Normalfall Probleme nicht lösen, sie kann höchstens Grundlagen für Entscheidungen bereitstellen und emotional oder ideologisch geführte Debatten auskühlen.

Aus gesellschaftlichem und politischem Interesse ist für ein sozial gerechtes und Chancengleichheit anstrebendes Volksschulsystem zu sorgen, das verhältnismäßig ähnlich guten Unterricht in allen Schulen anbietet. Je stärker der Trend zur Profilierung und Unterscheidung der einzelnen Schulen voneinander wird, desto wichtiger werden vergleichende Evaluationen der Schulqualität. Wird der Unterschied zwischen guten und schlechten Schulen zu groß, werden die guten Schulen von sozial starken und bildungsnahen Gesellschaftsschichten besetzt oder sie weichen in Privatschulen aus. Für die anderen bleiben die schlechten Schulen. Es ist deshalb für die Zukunft der Volksschule, aber auch in anderen Bildungsbereichen, nicht nur wichtig gute Schulen zu haben, sondern auch schlechte Schulen zu vermeiden.

Ich komme zum Schluss, indem ich auf wenige Ergebnisse aus der Evaluation der Sekundarstufe I im Kanton Zürich hinweise:

Der Vergleich der beiden Schulmodelle wurde im 8. Schuljahr durchgeführt. Von den insgesamt 11.895 Schülerinnen und Schülern des 8. Schuljahrs wurden im Juni 1996 über 3.000 aus 168 Klassen an zwei Halbtagen befragt und getestet. Zwischen den Schulmodellen ergaben sich keine bedeutenden Unterschiede der Leistungen in den Fächern Deutsch und Mathematik, auch dann nicht, wenn nur diejenigen Testaufgaben berücksichtigt wurden, deren Inhalt in den Klassen unterrichtet worden war. Auch die Zufriedenheit wird von den Schülerinnen und Schülern in beiden Schulmodellen gleich eingeschätzt. Keine Unterschiede gibt es zudem in der Leistungsbereitschaft im Fach Deutsch: Motivation, Interesse und Beteiligung am Unterricht, aber auch das Selbstvertrauen in die eigene Leistungsfähigkeit sind bei den Schülerinnen und Schülern beider Schulmodelle gleich ausgeprägt. Allerdings sind im Fach Mathematik Motivation, Interesse und Selbstvertrauen in die eigene Leistungsfähigkeit in der Gegliederten Sekundarschule leicht höher als in der Dreiteiligen Oberstufe. Entscheidend sind unter anderem auch die Muttersprache, das Geschlecht und die soziale Herkunft der Schülerinnen und Schüler; also Merkmale, die nicht zum direkten Einflussbereich der Schule gehören, die aber im Unterricht eine pädagogische Herausforderung darstellen.

Bildungspolitisch war zu entscheiden, ob eines der beiden Schulsysteme, nämlich das bessere, flächendeckend und definitiv für den ganzen Kanton eingeführt werden sollte, oder ob die Wahl des Schulsystems auf Gemeindeebene delegiert werden könnte. Die Evaluation hat nur geringfügige Unterschiede zwischen den Schulsystemen eruiert. Am 28. September 1997 wurde vom Souverän eine Revision des Volksschulgesetzes gutgeheißen, die vorsieht, dass sich die Gemeinden zwischen Dreiteiliger Sekundarschule und Gegliederter Sekundarschule (vormals AVO) entscheiden können und müssen. Die Abstimmungsvorlage und das Abstimmungsergebnis wären vermutlich anders ausgefallen, wenn die Evaluation zu anderen Ergebnissen gekommen wäre. Die Wahl der Schulmodelle wäre vom Regierungsrat nicht zur Abstimmung vorgelegt worden, wenn die Evaluation gezeigt hätte, dass ein Schulmodell deutlich bessere Resultate erzielt.

Zugleich ist aber auch anzunehmen, dass die Ergebnisse der Evaluation nicht ohne weiteres akzeptiert worden wären, wenn sich gezeigt hätte, dass eines der beiden Schulsysteme eindeutige Vorteile hat. Die Befürworterinnen und Befürworter des als schlechter beurteilten Schulsystems hätten am Auftraggeber, an den Evaluatoreninnen und Evaluatoren, an den Methoden, den Auswertungen und den Interpretationen der Daten gezweifelt und entsprechende Bedenken und Kritiken vorgebracht. Insgesamt hat das Ergebnis der Evaluation sicher auch zur Akzeptanz der Evaluation insgesamt beigetragen.

Ein Preis, der bei dieser Form von Evaluationen bezahlt werden muss besteht darin, dass der Zeitpunkt der Veröffentlichung der Resultate nicht von den Forschenden bestimmt werden kann. Je nach bildungspolitischer Konstellation haben die Ergebnisse für die Evaluierenden zu früh vorzuliegen oder sie dürfen erst zu spät veröffentlicht werden. In unserem Fall war aber garantiert, dass die Ergebnisse selber in keiner Weise vom Auftraggeber beeinflusst wurden. Vielleicht war es ein Glücksfall, dass sich die Bildungspolitik ernsthaft für Evaluationsforschung interessierte. Erziehungswissenschaftliche Forschung in Form von Evaluationen im Bildungs- und Schulbereich kann aber von solcherlei Aufträgen profitieren, sich weiterentwickeln und sich sogar als nützlich erweisen.

Literatur

- AURIN, K. & STOLZ, G. E. (1990): Erfahrungen aus der Aufarbeitung von Evaluationsvorhaben am Beispiel der Projektgruppe 'Gesamtschule' der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. In: *Zeitschrift für Pädagogische Psychologie*, 4, Heft 4, S. 269-282.
- BORTZ, J. & DÖRING, N. (1995): *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. 2., vollständig bearbeitete und aktualisierte Auflage. Berlin.
- BRANDTSTÄDTER, J. (1990): Evaluationsforschung: Probleme der wissenschaftlichen Bewertung von Interventions- und Reformprojekten. In: *Zeitschrift für Pädagogische Psychologie*, 4, Heft 4, S. 215-227.
- DRERUP, H. (1987): *Wissenschaftliche Erkenntnis und gesellschaftliche Praxis. Anwendungsprobleme der Erziehungswissenschaft in unterschiedlichen Praxisfeldern*. Weinheim.
- FABICH, F. (1993): *Forschungsfeld Schule. Wissenschaftsfreiheit, Individualisierung und Persönlichkeitsrechte. Ein Beitrag zur Geschichte sozialwissenschaftlicher Forschung*. (Studien und Berichte des Max-Planck-Instituts für Bildungsforschung, Bd. 54). Berlin.
- FEND, H. (1981): Stellungnahme zu Wottawa „Die Kunst der manipulativen Berichtlegung in der Evaluationsforschung“. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 13, Heft 1, S. 61-70.
- GRÜNDER, H. (1993): Evaluation und Evaluationsforschung im Bildungswesen. In: *Pädagogische Rundschau*, 47, Heft 1, S. 29-52.
- GUBA, E. & LINCOLN, Y. (1989): *Fourth Generation Evaluation*. California: Sage.
- JENZER, C. (1992): Wem nützt Evaluation - Wozu? Mit Beispielen aus dem Erfahrungshintergrund der Lehrplan- und Gesamtschulevaluation. In: GONON, P. (Hrsg.): *Evaluation in der Berufsbildung*. Aarau: Sauerländer, S. 13-30.
- KLÖTI, U. (1995): Politische Evaluation im Bildungsföderalismus. In: TRIER, U. P. (Hrsg.). *Wirksamkeitsanalyse von Bildungssystemen*. (NFP 33, Symposium Bern 1995). Aarau, S. 80-92.
- KÖTTER, L. (1988): Empirische Erkenntnis und bildungspolitische Entscheidung. In: *Zeitschrift für Pädagogik*, 34, Heft 5, S. 657-671.

- KROMREY, H. (1995): Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeiten ihrer Realisierung. In: *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 15, Heft 4, S. 313-336.
- LÖSEL, F. & NOWACK, W. (1987): Evaluationsforschung. In: SCHULTZ-GAMBARD, J. (Hrsg.): *Angewandte Sozialpsychologie*. München/Weinheim, S. 57-87.
- MEYLAN, J.-P. (Red.) (1981): *Innovation und Evaluation von Schulversuchen*. Bern: Haupt.
- MEYLAN, J.-P. (Red.) (1983): *Evaluation von Innovationen im Bereich der Primarschule*. Bern: Haupt.
- MITTER, W. & WEISHAUPT, H. (Hrsg.) (1977): *Ansätze zur vergleichenden Analyse der wissenschaftlichen Begleitung bildungspolitischer Innovationen*. Weinheim.
- MOSER, U. & RHYN, H. (1999): *Schulmodelle im Vergleich. Eine Evaluation der Leistungen in zwei Schulmodellen der Sekundarstufe*. Aarau: Sauerländer.
- OECD (Organisation for Economic Co-operation and Development) (1991): *Schulen und Qualität: ein internationaler OECD-Bericht*. Frankfurt a. M u.a.
- OECD (Organisation for Economic Co-operation and Development) (1994): *Die internationalen Bildungsindikatoren der OECD: ein Analyserahmen*. Frankfurt a. M u.a.
- OSTERWALDER, F. (1995): *Rationale Wirkungsannahme von Schule – Die Herausforderung eines empiristischen Schulkonzepts und sein Misserfolg in der Schweiz*. In: TRIER, U. P. (Hrsg.): *Wirkungsanalyse von Bildungssystemen*. Schweizerische Koordinationsstelle für Bildungsforschung: Aarau, S. 218-228.
- PRANGE, K. (1995): *Die Zeit der Schule*. Bad Heilbrunn.
- PURKEY, St. C. & SMITH, M. S. (1983): *Effective Schools: A Review*. In: *The Elementary School Journal*, Vol. 4, S. 21-31.
- RHYN, H. & MOSER, U. (1998): *Schulqualität und externe Evaluation*. In: *schweizer schule*, Heft 3, S. 21-31.
- RHYN, H. (1999): *Beurteilung und Selektion im demokratischen Schulsystem. Ein argumentatives Potpourri*. In: *Beiträge zur Lehrerbildung*, 17, Heft 1, S. 41-51.
- ROEDER, P. M. (1983): *Bildungsreform und Bildungsforschung*. In: BENNER, D.; HEID, H. & THIERSCH, H. (Hrsg.): *Beiträge zum 8. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft*. (Zeitschrift für Pädagogik, 18. Beiheft) Weinheim & Basel, S. 81-96.
- ROSSI, P. H. & FREEMAN, H. E. (1993): *Evaluation*. Beverly Hills: Sage Publications.
- ROSSI, P. H.; FREEMAN, H. E. & HOFMAN, G. (1988): *Programm-Evaluation: Einführung in die Methoden angewandter Sozialforschung*. Stuttgart.
- RUSSELL, N. & WILLINSKY, J. (1997): *Fourth Generation Educational Evaluation. The Impact of a Post-Modern Paradigm on School Based Evaluation*. In: *Studies in Educational Evaluation*, Vol. 23, No. 3, S. 187-199.
- SANDERS, J. R. (Ed.) (1994): *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. 2nd edition. Thousand Oaks: Sage Publications (dt. Opladen 1999).
- SCHEERENS, J. (1994): *Prozessindikatoren der Arbeitsweise von Schulen*. In: *OECD/CERI: Die internationalen Bildungsindikatoren der OECD: ein Analyserahmen*. Frankfurt a. M. u.a.: Peter Lang, S. 62-92.
- SCHNUER, G. (1986): *Die deutsche Bildungskatastrophe. 20 Jahre nach Picht, Lehren und Lernen in Deutschland*. Herford.
- SCRIVEN, M. (1991): *Evaluation Thesaurus*. London: Sage.
- STEFFENS, U. & BARGEL, T.: *Erkundungen zur Qualität von Schule*. Neuwied.
- WEBER, K. (1992): *Engagement und Distanz in der Evaluation*. In: GONON, P. (Hrsg.): *Evaluation in der Berufsbildung*. Aarau: Sauerländer, S. 109-120.
- WIDMER, T. (1996): *Meta-Evaluation: Kriterien zur Bewertung von Evaluationen*. Bern: Paul Haupt.
- WOLF, R. M. (1987): *The nature of educational evaluation*. In: *International Journal of Educational Research*, 11, S. 7-19.
- WOTTAWA, H. & THIERAU, H. (1990): *Lehrbuch Evaluation*. Bern, Stuttgart: Huber.
- WOTTAWA, H. (1991): *Zum Rollenverständnis in der Evaluation und der Evaluationsforschung*. In: *Empirische Pädagogik*, 5, Heft 2, S. 151-168.
- WOTTAWA, H. (1981): *Die Kunst der manipulativen Berichtlegung in der Evaluationsforschung*. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 13, Heft 1, S. 45-60.