

Brügelmann, Hans

Funktionen und Formen der Evaluation

Brügelmann, Hans [Hrsg.]: Was leisten unsere Schulen? Zur Qualität und Evaluation von Unterricht. Seelze : Kallmeyer 1999, S. 142-152



Quellenangabe/ Reference:

Brügelmann, Hans: Funktionen und Formen der Evaluation - In: Brügelmann, Hans [Hrsg.]: Was leisten unsere Schulen? Zur Qualität und Evaluation von Unterricht. Seelze : Kallmeyer 1999, S. 142-152 - URN: urn:nbn:de:0111-pedocs-183107 - DOI: 10.25656/01:18310

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-183107>

<https://doi.org/10.25656/01:18310>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Funktionen und Formen der Evaluation

Rechenschaft kann im Bildungswesen auch mit Hilfe von Modellen bzw. Methoden organisiert und ausgestaltet werden, die anderen gesellschaftlichen Bereichen oder Subkulturen entlehnt sind (vgl. nebenstehende Übersicht):

- In der juristischen Tradition geht es um die streitige Auseinandersetzung zwischen konkurrierenden Sichtweisen über Sachverhalte und Normen, z. B. die Rangfolge von Zielen für Unterricht (Anwälte Pro und Contra): Etwas gilt als „gut“ oder „wahr“, weil es in einem institutionell geregelten Verfahren als zutreffend bzw. gültig festgestellt worden ist.
- In der Forschungstradition kommt es auf die Argumentation mit methodisch gewonnenen Befunden im Kreis kritischer KollegInnen an: Etwas gilt als „gut“ oder „wahr“, weil es einer systematischen Prüfung seiner theoretischen Stimmigkeit und empirischen Geltung standgehalten hat.
- In der Tradition der Kunstkritik zählt das Urteil fachlich ausgewiesener anerkannter Personen: Etwas gilt als „gut“ oder „wahr“, weil es von erfahrenen und glaubwürdigen Leitfiguren der Subkultur entsprechend gewürdigt wird.
- In der Tradition weltanschaulicher Auseinandersetzungen werden Positionen durch die autoritative Auslegung geteilter Normen bzw. Annahmen begründet: Etwas gilt als „gut“ oder „wahr“, weil es in einer Gruppe, die sich durch dieselben Grundwerte und Sichtweisen verbunden fühlt, als verträglich mit der geltenden Dogmatik anerkannt worden ist.
- In der ökonomischen Tradition wird die Effektivität (Wirksamkeit) und Effizienz (Sparsamkeit) von Maßnahmen mit technischer Präzision gemessen: Etwas gilt als „gut“ oder „wahr“, weil es vorgegebene Soll-Werte mit geringer Fehlertoleranz einhält.

Die Vorzüge und Schwächen der verschiedenen Ansätze werden in den folgenden Abschnitten diskutiert. Die stichwortartige Gegenüberstellung in diesem Kasten macht jedoch schon deutlich,

- dass alle fünf Modelle hilfreich sein können, um tatsächlich vorfindliche Stile der Evaluation und Rechenschaft zu beschreiben und zu analysieren;
- dass aber nicht alle gleich angemessene Modelle anbieten, Verfahren der Qualitätskontrolle im Bildungsbereich zu entwickeln;
- dass sich je nach Fragestellung auch im Schulwesen eine andere Tradition als Modell für die Evaluation anbietet, z. B. das juristische Modell eher für die Klärung normativer Kontroversen, das ökonomische Modell eher zur Überprüfung methodischer Wirksamkeit;
- dass verschiedene Ansätze nicht unmittelbar miteinander kompatibel oder ihre Befunde gar direkt miteinander verrechenbar sind.

Die fünf „Traditionen der Evaluation“ sind eingebettet in ein jeweils besonderes Verständnis pädagogischen Handelns. Diese Sichtweisen werden in der folgenden Übersicht mit einigen Stichworten charakterisiert, z. B. im Blick auf typische Produkte, Aktivitäten, den Status von Plänen usw. (zusammengefasst nach Brügelmann 1980, S. 284–288):

Fünf Traditionen der Evaluation					
SUBKULTUR	RECHT	FOR- SCHUNG	KUNST	RELIGION	INDUS- TRIE
Programm- Analogie	Richtlinien Lehrpläne	didakt. Kriterien method. Regeln	Meister- werke	Lehrstücke	Unterrichts- modelle
Modell der Entwick- lungsarbeit	Gesetz- gebung Verordnung	Systemati- sche Unter- suchung	Verdichtung von Erfahrung	Gebote Botschaften	Erfindung Konstruk- tion
Status des Programms	Vorschrift	Hypothese	Partitur	Katechis- mus	Prototype
Vermitt- lungsweg	Verwal- tungs-/ Erlassweg	Publikation Profession	Bewegung Mode	Verkündi- gung Mission	Produktion Vertrieb
Rolle/ Selbstver- ständnis	BeamtInnen	Forsche- rInnen	Regisseu- rInnen	Priester- Innen	Techniker- Innen
Bild von Unterricht	legale Anordnung	argumenta- tive Über- zeugung	Inszenie- rung Interpre- tation	Bekehrung	Vervielfälti- gung Anwen- dung
Methode der Evaluation	unabhängi- ge Gerichte	Experiment Diskussion	anerkannte Kritike- rInnen	Auslegung von Dogmatik	Messver- fahren Tests
Werte/ Kriterien	Gerechtig- keit Gehorsam	Stimmigkeit Aufklärung	Originalität Glaubwür- digkeit	Wahrheit Glaube	Präzision Sparsam- keit
Strategie der Schul- reform	Bürokratie- modell	Problem- lösungs- modell	Interakti- onsmodell	Ideologie- modell	Technolo- giemodell

Die vier als mögliche Modelle breiter diskutierten Analogien Recht, Forschung, Kunst und Industrie stellen wir in kurzen Skizzen mit ihren jeweils besonderen Chancen und Risiken vor.

Ein juristisches Modell der Evaluation

Das juristische Modell organisiert Rechenschaft als Streitige Auseinandersetzung zwischen Parteien. In einem formalisierten Verfahren werden umstrittene Sachverhalte und unterschiedliche Meinungen über die Auslegung von Normen geklärt. Dieses Modell hat aus verschiedenen Gründen Bedeutung für die Beurteilung von Unterricht.

Das Schulwesen ist im Laufe dieses Jahrhunderts, vor allem in den letzten 30 Jahren zunehmend „verrechtlicht“ worden. SchülerInnen und ihre Eltern klagen häufiger als früher gegen Entscheidungen. Dies mag auch damit zusammenhängen, dass Rechtsschutzversicherungen solche Verfahren risikolos erscheinen lassen. Hinzu kommt aber ein verändertes Verständnis von Schule. SchülerInnen werden seit den 70er Jahren nicht mehr in einem „besonderen Gewaltverhältnis“ zum Staat gesehen, d. h. Eingriffe in ihre Rechte aufgrund betriebsinterner Regelungen sind nicht zulässig („Gesetzesvorbehalt“).

SchülerInnen und ihre Eltern verstehen sich zudem im Gegensatz zu früheren Zeiten als Kunden mit bestimmten Ansprüchen, deren Erfüllung sie dann auch einklagen. Qualität von Unterricht, vor allem Leistungsbeurteilungen und disziplinarische Maßnahmen, werden damit zum Gegenstand gerichtlicher Würdigung.

Das juristische Modell ist aber noch in einer grundsätzlicheren Hinsicht von Bedeutung. Ziele und Prinzipien pädagogischen Handelns lassen sich (anders als technische Qualitätskriterien) nicht eindeutig definieren. So ist beispielsweise Ende der 70er Jahre der Versuch gescheitert, die Anforderungen des Abiturs in sogenannten „Normbüchern“ verbindlich und messbar festzuschreiben (Flitner/Lenzen 1977). Urteile über die Qualität von Unterricht setzen insofern Verfahren voraus, in denen Konflikte bei der Interpretation von Maßstäben geklärt werden können.

Schiedsverfahren und Gerichtsverfahren sind aus der Einsicht erwachsen, dass Normkonflikte nicht technisch zu lösen sind. Andererseits wird soziales Handeln unmöglich ohne Verlässlichkeit. So brauchen auch PädagogInnen Verfahren, zu einer zeitlich und räumlich begrenzten Verständigung und damit Berechenbarkeit wechselseitiger Erwartungen zu kommen, so dass die Alltagsgeschäfte trotz unterschiedlicher Meinungen weitergehen können. Das juristische Modell garantiert keine „richtige“ Anwendung von Normen, aber es verspricht eine höhere Akzeptanz von Lösungen, z. B. weil die unterschiedlichen Sichtweisen in einem unabhängig organisierten Verfahren prinzipiell gleiche Chancen haben, zur Geltung zu kommen.

Das juristische Modell garantiert insofern weder Wahrheit noch Gerechtigkeit, aber es schränkt die Möglichkeit der Durchsetzung von mächtigen Deutungspositionen ein. Von seinen Grundsätzen und methodischen Verfahren kann ein Bildungssystem profitieren, das mit unterschiedlichen Wertvorstellungen und Erwartungen zu kämpfen hat (vgl. zum Verfahren im Einzelnen Brügelmann 1974; Sugarman 1974)

„Kunstkritik“ als Modell der Evaluation

Ein „Kunst“-Modell der Evaluation? Wer Evaluation mit Leistungstests gleichsetzt, muss dies als Widerspruch in sich selbst empfinden. Wenn man Qualität von Unterricht aber auf die Förderung der Persönlichkeitsentwicklung insgesamt bezieht und wenn man über die Wirkungen hinaus auch die Prozesse bewerten will, gewinnt diese Perspektive an Bedeutung und Attraktivität.

In der Tat gibt es ja auch in der Literatur, in der Ton- und der darstellenden Kunst eine etablierte Tradition der Beurteilung von Leistungen. Marcel Reich-Ranicki als Einzelperson, das „literarische Quartett“ als öffentlich diskutierende Jury signalisieren den Einfluss, aber auch die besondere Form einer Kunstkritik, die über personengebundene Urteile wirksam wird. Ihre Funktion lässt sich umreißen als „stellvertretende Erfahrung“: Personen mit einer speziellen und langjährigen Expertise oder mit einem in der Szene besonders anerkannten Urteil dokumentieren ihre Wahrnehmungen und Urteile für andere. Ob diese das Urteil übernehmen, hängt von ihrer Einschätzung der Person, insbesondere von deren Sachkenntnis, deren Position in der Subkultur und deren Glaubwürdigkeit ab. Die Subjektivität des Urteils ist also erträglich, weil der „Nutzer“ selbst entscheidet, auf wen er sich verlassen will. Eventuell wird sie ausbalanciert durch die Möglichkeit, Urteile verschiedener KritikerInnen zu vergleichen. Sie kann aber auch systematischer kontrolliert werden durch die Bildung einer Jury, in die bewusst Personen mit verschiedenen Positionen oder aus unterschiedlichen Traditionen berufen werden.

Diese Idee einer sozialen Kontrolle von Urteilen kommt dem juristischen Modell nahe, ist aber weniger formalisiert. Im pädagogischen Raum wird sie z. B. in Prüfungsausschüssen oder über Sachverständigenkommissionen realisiert. Dabei wird die Person als „Instrument“ der Evaluation genutzt, ihre Subjektivität bewusst nicht neutralisiert (z. B. durch technische Verfahren wie standardisierte Tests), sondern als besondere Qualität persönlicher Wahrnehmung geschätzt, andererseits aber möglichst transparent gemacht und durch die Einbindung in Gruppenprozesse in ihrer Wirkung begrenzt.

In der Rolle von InspektorInnen oder SchulrätInnen verbinden sich das Kunst- und das juristische Modell: Einerseits ist die Evaluation – trotz der rechtlichen Vorgaben – stark personengebunden, andererseits sind das Verfahren der Beurteilung und seine Konsequenzen bürokratisch verankert. Die Einbindung in eine hierarchische Aufsicht verschenkt bzw. überfordert jedoch das Potential des „Kunst“-Modells. Entweder müsste seine Funktion auf Beratung (wie beim Kritiker) beschränkt oder auch die Beurteilung selbst durch „checks and balances“ (wie im juristischen Modell) stärker formalisiert werden. Ein (heikler) Mittelweg wäre die Veröffentlichung von Berichten mit Kommentaren der Betroffenen wie in der Tradition von „Her Majesty’s Inspectorate“ im britischen Schulsystem: sozialer Druck statt rechtlicher Sanktionen (s. oben Niemann, S. 28).

Das Industrie-Modell der Evaluation

Das Industriemodell zielt auf eine technische Produktions- und Produktkontrolle. Seine Denkweise und konkrete Verfahren lassen sich deshalb nur nach sorg-

fältiger Prüfung der besonderen Umstände auf soziale Situationen oder Abläufe übertragen. Das Modell geht von fünf Annahmen aus:

1. Produktionsziele lassen sich vorweg genau und verbindlich spezifizieren.
2. Produktionsergebnisse lassen sich anhand dieser Vorgaben eindeutig bewerten.
3. Produktionsprozesse zur Herstellung der Produkte sind standardisierbar.
4. Zwischen Produktionsprozessen und -ergebnissen besteht eine berechenbare Beziehung.
5. Diese Beziehung ist auf der Grundlage vorgängiger Planung im Detail steuerbar.

Diese Annahmen treffen für pädagogische Prozesse nicht oder nur sehr eingeschränkt zu (vgl. „Voraussetzungen für Verantwortlichkeit“, oben S. 138), denn für diese gilt:

1. Ziele sind umstritten, Einstellungen und Fähigkeiten lassen sich nicht eindeutig als beobachtbare Verhaltensweisen operationalisieren.
2. Menschliches Verhalten ist mehrdeutig, seine Deutung ist vom situativen Kontext und der Wahrnehmung der Beteiligten abhängig.
3. Unterricht lässt sich nicht nach Blaupause vervielfältigen, sondern muss in Abhängigkeit von den beteiligten Personen und konkreten Bedingungen jeweils neu „inszeniert“ werden.
4. Entwicklungs- und Lernprozesse lassen sich nicht auf berechenbare Bedingungen zurückführen.
5. Wesentliche Entscheidungen fallen erst im Unterricht, auszuhandeln zwischen den Beteiligten, nicht als bloße „Anwendung“ vorgängiger Planung.

Planung von Unterricht ist notwendig. Aber sie muss immer hypothetisch bleiben, weil LehrerInnen und SchülerInnen ihre Rollen nicht als Marionetten spielen – in einer demokratischen Schule auch nicht spielen sollen – und weil eine jede Unterrichtsplanung, aber auch Erkenntnisse pädagogischer Forschung praktisches Handeln nie vorschreiben können. Auf dieser Einsicht gründet das folgende Modell.

Ein Forschungs-Modell der Evaluation

In der gegenwärtigen Diskussion dominiert die Vorstellung, dass Evaluation einem Selbstverständnis verpflichtet sei, das dem von Forschung entspricht. Forschung wird aber oft missverstanden als Messung von Effekten in einem Input-Output-Modell, wie gerade am Beispiel des Industriemodells beschrieben. Lawrence Stenhouse hat mit seiner Arbeitsgruppe am „Centre for Applied Research in Education“ (CARE) in Norwich in den 70er Jahren ein anderes Konzept vom „teacher as researcher“ entwickelt, das zwischenzeitlich auch im deutschsprachigen Raum das Selbstverständnis von Evaluation geprägt hat (Brügelmann 1975, und vor allem Altrichter/Posch 1998).

Wenn pädagogisches Handeln so komplex und situationsspezifisch ist, dass Theorien und Programme immer nur hypothetischen Charakter, also allenfalls eine statistische Gültigkeit haben, dann ist Unterricht mehr als bloße Anwendung oder Durchführung solcher allgemeinen Regeln oder Programme. Gerade um im Sinne allgemeiner Ziele erfolgreich zu handeln, müssen LehrerInnen die beson-

deren Bedingungen ihres Unterrichts erkunden und die Wirkungen ihres Handelns ständig überprüfen. Nichts anderes aber meint Evaluation: eine erfahrungsorientierte und kritische Kontrolle pädagogischen Handelns.

Die Kompetenz von LehrerInnen hat in diesem Modell eine große Bedeutung, beschränkt sich jedoch nicht auf die mit dem täglichen Tun implizit wachsende Erfahrung. Gefordert sind auch ein explizites Nachdenken und eine kritische Auseinandersetzung mit den Sichtweisen und Deutungen anderer. Eine verlässliche Kooperation in und zwischen Kollegien ist dafür Voraussetzung, aber ebenso ein institutionell abgesicherter Austausch zwischen dieser Praxisforschung in den Schulen und der auf allgemeine Fragen spezialisierten Forschung an den Hochschulen. Insbesondere das Potential einer Zusammenarbeit mit PraktikantInnen und LehramtswärterInnen für eine kumulative Praxisforschung ist noch bei weitem nicht ausgeschöpft.

Den Fokus einer solchen Schulforschung bilden Prozesse im Unterricht, deren Qualität zu entwickeln ist. Daneben muss aber auch die Frage nach der Wirksamkeit von schulischen Maßnahmen und nach dem Aufwand, der für ihre Durchführung erforderlich ist, gestellt werden.

Kosten und Nutzen von Kosten-Nutzen-Analysen

In den OECD-Ländern werden im Durchschnitt 6 % des Bruttosozialprodukts für Bildung ausgegeben. Der Anteil an den öffentlichen Haushalten beträgt rund 13 % – in Deutschland jährlich 149 Mrd. DM. Das sind, wie in Italien und den Niederlanden, etwas weniger als 10 %, während in Ländern wie der Schweiz, Neuseeland und Norwegen mehr als 15 % der Staatsausgaben auf Bildung entfallen (Block/Klemm 1997, S. 24 ff.; s. zur laufenden Aktualisierung der OECD-Daten im Internet: <http://www.oecd.org/els/stats/eag97/chapterb.htm>).

Bei solchen Summen liegt es auf der Hand, dass die Öffentlichkeit Rechenschaft verlangt über den Ertrag: Leisten Schulen das, was sie leisten sollen? Und werden die dafür aufgewandten Mittel optimal genutzt, d. h. sparsam eingesetzt? Effektivität und Effizienz sind Kriterien der Bildungsökonomie. Sie legen eine betriebswirtschaftliche Prüfung der Arbeit von Schulen nahe. Die Anwendung auf Bildungsprozesse ist aber schwierig.

Schon der *Nutzen* ist schwer zu fassen und unterschiedlich zu bewerten, je nachdem, welche Perspektive man einnimmt:

- Bewerten wir den Ertrag aus der Sicht der SchülerInnen oder aus der bestimmter gesellschaftlicher Instanzen? Dass alle BürgerInnen gemeinsame Erfahrungen mit kulturellen Traditionen teilen, mag für den Bestand einer Gesellschaft wichtig, für die einzelne Person dagegen unwichtig sein. Ein individuelles Talent zu entwickeln kann den einzelnen sehr befriedigen, auch wenn die Gesellschaft weder kulturell noch ökonomisch davon profitiert.
- Erheben wir nur kurzfristige Effekte, z. B. Leistungen in einzelnen Fächern am Ende einer Einheit bzw. eines Schuljahres, oder auch langfristige Wirkungen, z. B. den Berufserfolg von AbsolventInnen verschiedener Programme oder Einrichtungen? Studien aus den USA zeigen für den akademischen Bereich, dass dort kaum eine Korrelation zwischen beiden Merkmalen besteht.

„Qualitätsmanagement“ von Bildung?

In der Wirtschaft haben sich mit der ISO 9000-Serie seit zehn Jahren international bestimmte Standards für die Kontrolle von Qualität in Unternehmen durchgesetzt. Diese Normen geben keine inhaltlichen Merkmale für Produkte vor. Sie beschreiben vielmehr bestimmte Anforderungen an den Prozess der Herstellung und seine Überwachung. Definiert wird also ein System der Qualitätskontrolle:

„Nach der klassischen Endkontrolle, bei der man meinte, Qualität durch Aus-sortieren erzeugen zu können, folgte die Strategie, Prozesse zu kontrollieren und dabei die Qualität in die Produkte hinein zu produzieren. Später begriff man immer mehr die Notwendigkeit zum übergreifenden Managementansatz, wenn Spitzenqualität erreicht werden soll.“ (Rühl 1998, S. 22)

Übertragen auf den Bildungsbereich bedeutet das: Effekte, z. B. Leistungen von SchülerInnen, sind nur ein Aspekt des Systems, an dem Qualität erkennbar wird, und deshalb unzureichend als Indikator für den Wert und das Niveau pädagogischer Arbeit. Aber darf man die Idee des Qualitätsmanagements (QM) überhaupt unbesehen auf das Bildungswesen übertragen?

Schon länger wird die Diskussion im Bereich der Weiterbildung geführt (z. B. Meifort/Sauter 1991), seit kurzem auch im Hochschulbereich. Dabei werden zum Beispiel Studierende als Kunden eines Fachbereichs betrachtet und aus ihrer Perspektive Anforderungen an die Qualität der Ausbildung (und ihrer Kontrolle) bestimmt:

„Allerdings: der Kunde entscheidet selbstständig und eigenverantwortlich über das von ihm gewünschte Produkt und die von ihm gewünschte Qualität; dies können (und sollen) Studierende nicht in allen Fällen. Weiterhin ist die Grundidee der Kunden-Lieferanten-Beziehung, dass der Kunde möglichst genau das bekommt, was er wünscht. Dies könnte im Extremfall dazu führen, dass der Studentenwunsch ‚Examen‘ bzw. ‚Schein‘ dominiert.- Und ein Kunde der ‚durchfällt‘? Also Studierende als ‚Produkt‘? Und als ‚Kunde‘ die Gesellschaft? Oder der Staat? Oder die einstellenden Institutionen? Auch dies bereitet Probleme: Wie passt der maßgebliche Anteil, den Studierende am Studienerfolg haben, zu einem ‚Produkt‘? Soziale Interaktion mit einem ‚Produkt‘? Motivation bei einem ‚Produkt‘? Entsprechende Fragen lassen sich genauso stellen beim ‚Kunden‘ Staat, Gesellschaft oder Absolventen einstellende Institution.“ (Stawicki [Beauftragter für die Evaluation an der FH Wiesbaden]1998, S. 73 f.)

Deshalb der Versuch einer anderen Modellierung:

„Produkt für eine Bildungsstätte ist der Lehrstoff oder das Lehrmaterial. Einfacher noch als im üblichen Produktionsbetrieb ist der Erfolg am Ergebnis, zum Beispiel an Prüfungs- oder Examensergebnissen abzulesen. [...] Hauptzweck der Produktion ist im übertragenen Sinne die Wissensvermittlung, das heißt das Unterrichten ... Das setzt die Qualität der Prozesse voraus, das heißt in dem vorliegenden Fall [,] die Qualität des Ablaufs in Bildungseinrichtungen [...]

Ähnlich verhält es sich bei der Lenkung fehlerhafter Produkte und bei Korrekturmaßnahmen. Man denke dabei an Fälle, wo der Student ein gestecktes Ziel nicht erreicht hat.“ (Fuhr [Deutsche Gesellschaft für Qualität, e. V.] 1998, S. 54 ff.)

Die Einwände liegen auf der Hand:

Je mehr „... wir an den Hochschulen versuchen, von den Bedingungen des etablierten QM in den Unternehmen auszugehen, um so mehr verstricken wir uns in den schier unauflösbaren definitiven Problemen von ‚Kunde‘, ‚Lieferant‘, ‚Qualitätsverantwortung‘ oder ‚fehlerhaftem Produkt‘. [...] Je mehr auch in den Hochschulen selbst Lehren und Lernen als quasi technischer Prozess der Produktion von Qualifikationen verstanden wird, in dem es Produzierende und Produzierte gibt, umso mehr verlieren wir eine Grundwahrheit unserer Arbeit aus den Augen, die einen entscheidenden Unterschied zwischen einer Hochschule und einem Unternehmen ausmacht: Nämlich dass Lehren und Lernen nur in einer gemeinschaftlichen Anstrengung von Lehrenden und Studierenden fruchtbar werden und dass diese Gemeinschaftlichkeit angesichts der bestehenden Autonomierechte der Beteiligten ein hohes Maß an Partizipation und Interaktion erforderlich macht.“ (Rieck [Professor für Personalwesen] 1998, S. 81 f.)

Wirtschaft und Bildung – zwei wesensverschiedene Welten?

„Eine Möglichkeit, den oben geschilderten Schwierigkeiten auszuweichen, besteht darin, die Norm so zu adaptieren, dass sie für den Bereich der Hochschule ‚passt‘ (so wie man vor einiger Zeit eine eigene Norm für den Dienstleistungsbereich geschaffen hat) – also eine ‚DIN EN ISO 9005 – Qualitätsmanagement an Hochschulen‘. Dort finden sich dann keine Lieferanten, keine Kunden, keine Produkte, sondern Fachbereiche, Studierende, Lehrende, Vorlesungen, Prüfungen etc. – die Dinge heißen so, wie sie eben heißen. Die Adaptation der Norm muss dann dafür sorgen, dass Studierende da, wo sie als Kunden auftreten, auch entsprechend behandelt werden; aber in anderen Situationen eben auch anders.“ (Stawicki 1998, S. 75)

Was können Schulen lernen aus dieser Diskussion über Qualitätsmanagement im Bildungswesen? Das Wichtigste: Simple Analogien verbieten sich. Schulen sind keine Betriebe. Aber auch Schulen sollten sich Gedanken darüber machen, wie sie ihre Arbeit durchsichtiger machen können, so dass die Selbstkontrolle wirksamer und eine Rechenschaft nach außen möglich wird. Fünf Forderungen sind besonders wichtig:

1. Ziele klarer bestimmen: Was wollen wir? (vgl. „Schulprogramm“, „Schulprofil“)
2. Verantwortungen definieren: Wer ist für was zuständig?
3. Außenorientierung verdeutlichen: Was wollen wir für wen leisten?
4. Prozesse und Bedingungen klären: Wie und womit können wir das erreichen?
5. Systeme der Kontrolle und Verbesserung installieren: In welchem institutionellen Rahmen und mit welchen Verfahren können wir sichern, dass wir Probleme rasch erkennen und wirksam lösen?

- Stützen wir uns auf abstrakte Vergleichsdaten (z. B. Leistung in einem Test) oder fragen wir nach dem subjektiven Kompetenzgefühl (z. B. bezogen auf bestimmte berufliche Anforderungen). Beispielsweise zeigt die International Adult Literacy Study (OECD 1995), dass es Personen im unteren Leistungsviertel der Schriftkundigkeit gibt, die mit ihren Fähigkeiten durchaus zufrieden sind, während andererseits Personen im oberen Leistungsviertel ihre Fähigkeiten als unzureichend erleben. Analog ist zu fragen: Reicht der leicht messbare berufliche Status als Kriterium für Erfolg im Leben oder kommt es auf das subjektiv empfundene Lebensglück an? Und bis zu welchem Grad lässt sich dieses auf spezifische Erfahrungen in der Schule zurückführen?
- Nehmen wir nur die in Lehrplänen ausgewiesenen Fachziele in den Blick oder auch die in den allgemeinen Richtlinien propagierten Grundqualifikationen und Leitideen der Persönlichkeitsentwicklung? Untersuchungen zeigen, dass viele LehrerInnen diese Perspektiven als Zielkonflikt erleben und unterschiedliche Prioritäten setzen (s. u. a. Weinert/Helmke 1996, Kap. 11).

Was andererseits sind relevante *Kosten*? Die Gehälter der LehrerInnen und anderen Personals, Lehr- und Lernmaterialien, der Bau und die Ausstattung von Gebäuden, deren Instandhaltung und allfällige Verbrauchskosten wie Strom, Telefon usw. – alle diese materiellen Aufwendungen lassen sich zahlenmäßig verrechnen. Aber wie steht es mit psychischen und sozialen Kosten wie

- individueller Lebenszeit, die SchülerInnen als persönlich sinnlos verbracht erleben;
- Unzufriedenheit von LehrerInnen, die mit Lehrwerken oder nach einer Methode unterrichten sollen, mit denen sie inhaltlich nicht übereinstimmen;
- Spannungen bzw. offene Konflikte zwischen LehrerInnen und Eltern (oder innerhalb dieser Gruppen), weil sie unterschiedlichen Auffassungen von „gutem Unterricht“ haben;
- der wegen Fortbildungen zur Einführung eines neuen Lehrplans ausgefallener Unterrichtszeit?

Schon diese kurze Skizze macht deutlich: Kosten-Nutzen-Analysen im pädagogischen Bereich stehen selbst unter dem Vorbehalt, dass ihr Nutzen die Kosten rechtfertigt. Wenn sie die oben genannten Randbedingungen und Nebenwirkungen erfassen wollen, sind sie sehr aufwendig und verlieren notwendigerweise auch an Präzision. Je weniger sie andererseits von diesen Effekten und ihren Voraussetzungen erfassen, umso weniger aussagekräftig sind die Befunde. Umfassende Ansätze, die nicht nur unmittelbare finanzielle Kosten und Erträge berücksichtigen, sind selten (vgl. zu einem differenzierten Ansatz und seinen Schwierigkeiten: Block/Klemm 1997 und Klemm 1998b). Auf viele Fragen kann eine unaufwendige Sekundärauswertung von Statistiken, die in der Verwaltung sowie so anfallen, Auskunft geben (vgl. z. B. zum Nutzen von Schulkindergärten: Mader u. a. 1991; Rossbach/Tietze 1991).

Kritische Rückfragen an „harte Daten“

Die Aufgeregtheit der gegenwärtigen Diskussion über „Schulqualität“ produziert viele Daten und noch mehr Interpretationen dieser Daten. Die folgenden Fragen

verweisen auf häufige Schwächen empirischer Untersuchungen und der Darstellung bzw. Deutung ihrer Befunde. Sie sollten bei einer kritischen Lektüre von Aussagen zum Leistungsstand bzw. zur Leistungsentwicklung von SchülerInnen im Blick behalten werden:

- Decken die gewählten Dimensionen ein zureichend breites Leistungsspektrum ab (z. B. nicht nur Schriftsprache und Arithmetik als „Leistung der Grundschule“)?
- Werden alle bedeutsamen Dimensionen einer Leistung berücksichtigt (z. B. nicht nur die Rechtschreibung als Indikator für „Sprachkompetenz“)?
- Erfassen die Instrumente tatsächlich, was beurteilt werden soll (z. B. nicht nur Diktate als Ausweis von „Rechtschreibkönnen“)?
- Werden die Aufgaben dem Inhalt gerecht (z. B. keine Rechtschreibfallen oder exotischen Wörter, wenn es um die „grundlegende Rechtschreibfähigkeit“ geht)?
- Sind die Niveaus „ausreichender“ oder „guter“ Leistungen sinnvoll bestimmt (z. B. mit Fehlertoleranz bei Alltagsanforderungen an Rechtschreibung statt einer fiktiven „allgemeinen Rechtschreibsicherheit“)?

Besondere Anforderungen gelten für Untersuchungen in historischer Perspektive und für Systemvergleiche:

- Wurden die Eingangsvoraussetzungen kontrolliert (z. B. Zahl der Kinder anderer Muttersprache in verschiedenen Schulen)?
- Werden die Rahmenbedingungen des Unterrichts bei der Interpretation von unterschiedlichen Effekten berücksichtigt (z. B. Ausfall von LehrerInnen in einer Klasse; Grad der Aussonderung von leistungsschwachen SchülerInnen)?
- Sind die untersuchten Gruppen vergleichbar (z. B. „HauptschülerInnen“ 1955 und 1995; „10. Schuljahr“ in Deutschland und USA)?
- Sind die Aufgabenformen gleich vertraut (z. B. Tests unter Zeitdruck in Japan und Deutschland; Diktate 1910 und 1980)?
- Sind die Inhalte gleich bedeutsam – für den beruflichen/privaten Alltag, aber auch in der Schule (z. B. „1 x 1“ in den 20er und in den 70er Jahren)?
- Sind die statistischen Grundregeln für empirische Studien eingehalten worden (vor allem: Repräsentativität von Stichproben und Prüfung von Unterschieden auf Zufälligkeit)?
- Werden die gewonnenen Daten sinnvoll interpretiert (z. B. eine Fehlerdifferenz von 2 in 100 auf 3 in 100 Wörtern nicht als „Zunahme um 50 %“)?
- Vermeidet die Form der Darstellung suggestive Stilmittel (z. B. das „Abschneiden“ der unteren 50 % in Schaubildern, um kleine Differenzen optisch zu verstärken)?

Memento: „Harte Daten“ sind unverzichtbar, aber:

1. Nicht alles, was wichtig ist, lässt sich in Zahlen fassen.
2. Zahlen sprechen nicht für sich, ohne Kontext sind sie missverständlich.
3. Statistiken müssen mit Sachkunde interpretiert werden.

Auch die scheinbar harten Daten von Tests, auch die vordergründige Objektivität ökonomischer Analysen haben also ihre Schwächen. Welches Fazit lässt sich aus dem Vergleich der verschiedenen Modelle der Evaluation ziehen?

Sowohl-als-Auch statt Entweder-Oder:

Die Balance von Verfahren und Zuständigkeiten

Es gibt nicht *die* Methode der Evaluation oder *das* Verfahren der Rechenschaft. Alle haben ihre Stärken und ihre spezifischen Schwächen. Vieles spricht für eine Kombination verschiedener Elemente bzw. für eine aufgaben- und situationsspezifische Auswahl. In der gegenwärtigen Situation allerdings erfordert das Kriterium der Balance eher eine Stärkung

- lokaler gegenüber zentraler Verantwortung;
- demokratischer gegenüber bürokratischer Normen;
- professioneller und klientenorientierter gegenüber politischer und administrativer Rechenschaftsflegung;
- forschungs- und kunstorientierten gegenüber rechtlichen und industriellen Formen der Evaluation;
- der Kriterien Legitimität und Stimmigkeit gegenüber Wirksamkeit und Sparsamkeit;
- der Funktionen Diagnose und Beratung gegenüber Außenrechtfertigung und Akkreditierung;
- von Beschreibung und Analyse gegenüber Bewertung.

Der folgende Vorschlag skizziert einen pragmatischen Weg zur Einrichtung eines differenzierten Evaluationssystems.