

Schwippert, Knut; Frey, Hanno
Schulevaluation: ein Beispiel aus England

Tertium comparationis 10 (2004) 1, S. 64-81

urn:nbn:de:0111-opus-29514

in Kooperation mit:



<http://www.waxmann.com>

Nutzungsbedingungen

pedocs gewährt 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 dem Gebrauch von pedocs und der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Kontakt:

peDOCS

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Informationszentrum (IZ) Bildung

Schloßstr. 29, D-60486 Frankfurt am Main

eMail: pedocs@dipf.de

Internet: www.pedocs.de



Schulevaluation: ein Beispiel aus England

Knut Schwippert und Hanno Frey
Universität Hamburg

*„Man begnügt sich nicht damit, daß man dem andern einfach seine Meinung sagt,
man bemüht sich zugleich um ein Maß, damit sie den andern nicht umwirft,
sondern ihm hilft; wohl hält man ihm die Wahrheit hin,
aber so, daß er hineinschlüpfen kann“*
(Max Frisch, 1965, S. 48)

Abstract

During the last years, effective evaluation has become one of the key-processes for school development. Not surprisingly, the two most important methods within this area, internal and external evaluation, are being discussed – mostly in a controversial way. This is especially true for Germany, where school evaluation has only just started to gain public attention. In the article, however, the authors argue that internal and external evaluation are not necessarily to be understood as opposing strategies but that parts of both can, instead, be combined to a positive end. In order to sustain their thesis they relate to a topical example which illustrates that a combined method is highly effective; namely evaluation in private schools in England by the *Independent School Inspectorate*. The article, furthermore, discusses the benefits that can be drawn from an adoption of this method as well as certain risks it may contain.

1 Einleitung

In der aktuellen pädagogischen Fachliteratur wird ebenso wie in Berichten der (neuen) Medien gegenwärtig eins deutlich: Evaluationen im Bildungsbereich haben Konjunktur. Dies gilt insbesondere nach dem schlechten Abschneiden deutscher Schülerinnen und Schüler in internationalen Schulvergleichsstudien wie TIMSS oder PISA, in dessen Folge es zu einem regelrechten Evaluationsboom in Bildungseinrichtungen gekommen ist.

In diesem Beitrag soll kurz vorgestellt werden, wie sich Evaluationen, die von Schulen selber bzw. die von Außenstehenden durchgeführt werden, unterscheiden. Im

Anschluss daran wird die Evaluation einer englischen Schule durch eine private Gesellschaft vorgestellt, die seit Jahren Erfahrungen mit der Evaluation von Bildungseinrichtungen und der Rückmeldung von Evaluationsergebnissen hat. Abschließend wird aufgezeigt, welche Perspektiven sich aus dieser Evaluationspraxis für die in der BRD durchgeführten Evaluationen ergeben, um schließlich den Beitrag mit einem Ausblick auf zukünftige Entwicklungen der Evaluationsmethoden zu schließen.

2 Der Evaluationsbegriff

Schulevaluation kann als eine kriteriumsorientierte Untersuchung von Prozessen oder Ergebnissen beschrieben werden, bei der neben dem Untersuchungsgegenstand auch Rahmenbedingungen mit in den Blick genommen werden. Allerdings: Nur durch Beobachtung und Rückmeldung der Ergebnisse wird der Unterricht nicht besser. Wichtige Voraussetzung für einen erfolgreichen Evaluationsprozess ist deshalb die Bereitschaft, Ziele zu formulieren und darauf aufbauend Strategien zur Qualitätsverbesserung einzuleiten (Becker, Ilsemann & Schratz, 2001), denn bei einer Evaluation geht es in erster Linie nicht um ein Bewerten des Evaluationsgegenstandes sondern darum, Informationen zu ermitteln, die in einen Entscheidungsprozess zur Auswahl von Handlungsalternativen genutzt werden können. Ein zentrales Merkmal der Evaluation ist somit die Nutzenüberlegung (Wottawa & Thierau, 1998).

Im Folgenden werden zwei Formen der Schul- bzw. Unterrichtsevaluation vorgestellt, die sich bezüglich der Evaluatoren unterscheiden. Neben diesem Merkmal können Evaluationen auch bezüglich des Auftraggebers, des Zeitpunkts, den eingesetzten Instrumenten und verschiedener Nutzenaspekte differenziert werden. Hieran interessierte Leser seien die weiterführenden Beschreibungen von van Ackeren (2003), Helmke (2003) und Wottawa & Thierau (1998) und die Homepage des Landesinstituts für Schule in Soest empfohlen: www.learn-line.nrw.de/angebote/standardsicherung/.

3 Interne und externe Evaluation

Nach einer von Wottawa und Thierau (1998) gegebenen Definition lassen sich zwei Evaluationsarten unterscheiden, wobei sich das Hauptkriterium dieser Differenzierung auf die beteiligten Evaluatoren bezieht: Bei internen Evaluationen stammen diese aus dem betrachteten Evaluationssystem, während sie bei der externen aus anderen Institutionen kommen. Als einfaches Beispiel einer internen Evaluation kann die Beobachtung des Unterrichts durch eine aus dem eigenen Kollegenkreis stammende Person genannt werden. Demgegenüber kommt der Evaluator bei externen Evaluationen z.B. aus einem Landesinstitut oder einem Ministerium. Auch groß angelegte Schulvergleichsuntersuchungen wie TIMSS, PISA oder IGLU zählen zu den externen Evaluationen.

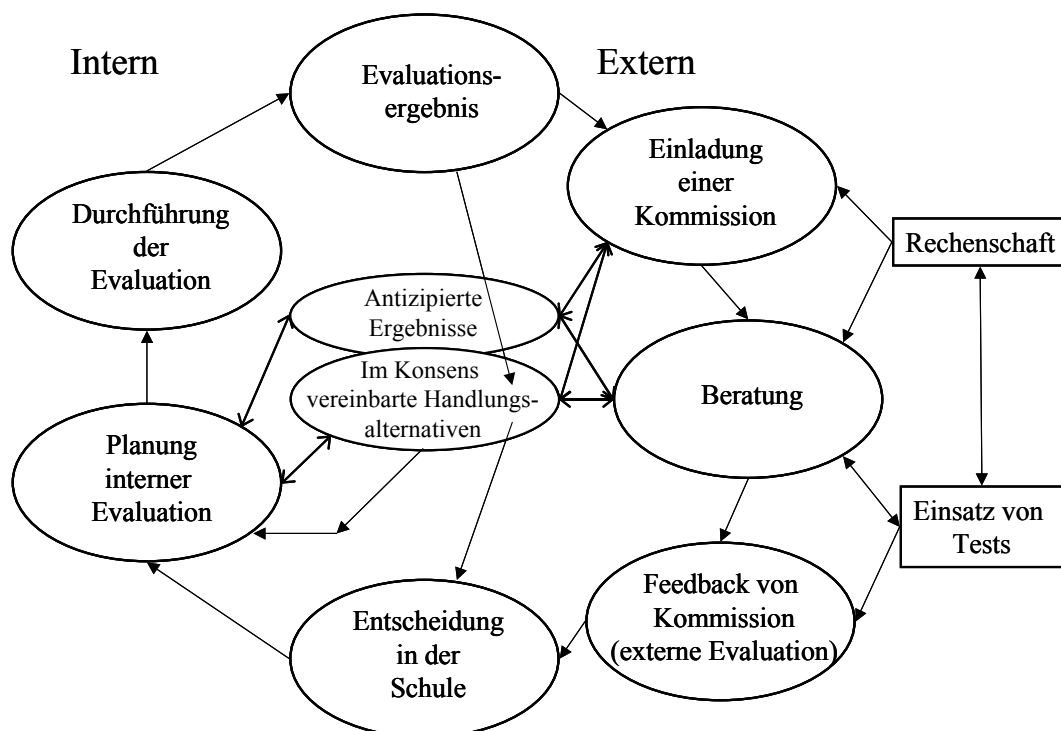
Unterschiede zwischen diesen beiden Evaluationsarten lassen sich aber nicht nur hinsichtlich des Vorgehens, sondern auch in Bezug auf die Folgen finden: Durch externe Evaluation kann auf den Schulen ein besonderer Druck lasten, insbesondere wenn die Befunde veröffentlicht werden und einzelnen Schulen schlechte Ergebnisse attestiert werden. Als Folge der im Rahmen der TIMS- und PISA-Studien erkannten

Defizite im deutschen Bildungswesen begannen einige Schulen deshalb recht bald, einzelne Bereiche der Schule selbständig und damit intern zu evaluieren. Mühlhausen (1999) und Altrichter und Messner (1998) berichten in diesem Zusammenhang von Beispielen, in denen das Unterrichtsgeschehen evaluiert wurde und Lehrer sich gegenseitig in der Interaktion mit den Schülern beobachteten. Als Vorteil wird bei internen Evaluationen von den Schulen vor allem die Möglichkeiten bewertet, dass sie selbst bestimmen konnten, was evaluiert werden soll und welche Ergebnisse an die Öffentlichkeit gegeben werden. Insofern ist es kaum verwunderlich, wenn innerhalb der Lehrerkollegien die interne Evaluation zumeist favorisiert wurde.

In der Folge wurden bei der Entscheidungsfindung bezüglich einer Evaluationsart beide Vorgehensweisen oftmals als konkurrierende Methoden wahrgenommen und das Für und Wider kategorisch und nicht selten heftig diskutiert. Im Folgenden soll vorgestellt werden, warum diese dichotome Sichtweise keinesfalls zwingend ist und inwiefern eine Verbindung beider Evaluationsmodelle sinnvoll sein kann.

Im Evaluationsmodell von Kempfert und Rolff (1999) und seiner Erweiterung von Bos und Postlethwaite (2000) werden beide Evaluationsformen nicht konkurrierend, sondern einander ergänzend verstanden. In dem von Bos und Schwippert (2002) erweiterten Modell (vgl. Abb. 1) werden explizit die antizipierten Ergebnisse und die im Konsens vereinbarten Handlungsalternativen berücksichtigt.

Abbildung 1: Erweitertes Modell der Verknüpfung interner und externer Evaluation



Hinter der Verknüpfung von interner und externer Evaluation steht der Gedanke, dass sich durch eine sinnvolle Verzahnung beider Ansätze eine Informationsmaximierung ergibt. Nachteile des einen Verfahrens können partiell durch die Vorteile des anderen kompensiert werden. So werden ‚blinde Flecke‘ und gemiedene Problemfelder interner

Evaluation im Rahmen einer externen Evaluation in den Blick genommen, und die Informationsentwicklung und Interpretation der Befunde wird durch externe Evaluationen ohne persönliche Vorbehalte offengelegt. Darüber hinaus kann die fehlende Tiefe bei der Beurteilung von Prozessen durch externe Beobachter durch die Detailkenntnis interner Beteiligter ergänzt werden.

Die Präsentation der Evaluationsergebnisse trägt maßgeblich zur Akzeptanz und zum Gelingen sowohl interner als auch externer Evaluation bei. Dass hierbei insbesondere auf eine angemessene und korrekte Darstellung der Ergebnisse zu achten ist, kann als selbstverständlich vorausgesetzt werden. Ein besonderes Augenmerk ist der Präsentation von statistischen Kennwerten zu widmen. Marginale bzw. relevante Befunde sind als solche angemessen den zum Teil nicht versierten Lesern der Evaluationsberichte vorzustellen (Klieme, Baumert & Schwippert, 2000). In diesem Zusammenhang ist schon bei der Planung einer Evaluation festzulegen, ob und an wen Rückmeldungen gegeben werden sollen. Die Form der Rückmeldung ist hierbei ebenso zu vereinbaren wie mögliche Unterstützungen zur Ergebnisinterpretation und zur Entwicklung von Interventionsmaßnahmen.

Bei der Vereinbarung von Handlungsalternativen ist zu überlegen, ob einige Ergebnisse als so persönlich eingestuft werden, dass diese nur den betroffenen Personen mitgeteilt werden oder ob diese im Rahmen des vereinbarten Prozesses allen Beteiligten zur Verfügung gestellt werden können. Durch Letzteres würde der gesamte Evaluationsprozess ein Maximum an Transparenz erreichen. Allerdings setzt eine offensive Diskussion der Ergebnisse ein professionelles Selbstverständnis bei der Reflexion des eigenen Handelns bei den Beteiligten voraus. Es ist zu beachten, dass rein persönliche Rückmeldungen neben der geringeren Transparenz auch die Durchführung der im Rahmen eines Qualitätsmanagements notwendigen Kontrollen erschweren. Hieraus ergibt sich, dass schon vor einer Entscheidung für eine Evaluation zu klären ist, wie die durch die Evaluation ausgelösten Prozesse im Rahmen weiterer Evaluationen überprüft werden.

Nach diesem kurzen Überblick betreffend interner und externer Evaluationsverfahren wird im Folgenden eine seit einiger Zeit praktizierte Mischform beider Evaluationsformen exemplarisch vorgestellt. Es handelt sich hierbei um die etablierte Evaluation von Privatschulen in Großbritannien durch das *Independent School Inspectorate* am Beispiel der Evaluation des *Plymouth College*.

4 Ein Evaluationsbeispiel aus England

4.1 Das englische Bildungssystem

Ähnlich wie in der Bundesrepublik Deutschland, war auch das englische Bildungswesen vor 1988 durch Dezentralismus gekennzeichnet, denn auch in England formulierte das Erziehungsministerium die übergeordneten Bildungsziele, während die Durchführung Sache der örtlichen Erziehungsbehörden, der *Local Education Authorities*, war. Letztere erarbeiteten in Zusammenarbeit mit den Schulbeiräten die Curricula, und zwar entsprechend der jeweiligen Besonderheiten der Schule (Sauxter, 1995). Dies

änderte sich durch den von der Regierung um Margaret Thatcher eingeführten *Education Reform Act* von 1988, der auch heute noch von entscheidender Bedeutung für das britische Schulsystem ist. Hintergrund für diese Reform war die Überzeugung, dass Wettbewerb zwischen Schulen zu einer Verbesserung der qualitativen Standards führen würde (van Ackeren, 2002). Der Wettbewerb zwischen den Schulen erfolgt dabei auf zwei Ebenen: Zum einen über zentrale Prüfungen wie z.B. die *General Certificate of Secondary Education* (GCSE-)Prüfung, die etwa der Mittleren Reife entspricht, und zum anderen über sogenannte *assessments*, durch die Schülerinnen und Schüler in ihrer Schulzeit bis zu sechs mal extern getestet und überprüft werden (ebd.). Eine direkte Folge dieses Systems ist die starke Orientierung der Schulen am *national curriculum*, dem landesweiten Lehrplan. Dieser ist in vier Stufen (*key stages*) aufgeteilt, an deren Ende jeweils Prüfungen abgelegt werden müssen. Innerhalb dieser *key stages* legen Fertigungsziele fest, was Schüler in den entsprechenden Altersgruppen wissen und verstehen sollen (Sauxter, 1995). Am Ende jeder Stufe werden vergleichende Leistungstest vorgenommen. Die Motivation für dieses Vorgehen ist offensichtlich, denn nur dann, wenn den Schülern das notwendige Wissen vermittelt wird, welches sie im Test benötigen (bzw. in den *assessments* verlangt wird), können sie diese auch gut bestehen. Es kommt somit zu einem doppelten Effekt: Zum einen gibt es landesweit einen umfassenden Standard; zum anderen bleibt weniger Raum für schuleigene Ziele und Lehrinhalte, da die Schulen vor allem bestrebt sind, in den Tests und *assessments* gute Ergebnisse zu erzielen. So sind durch die Erfüllung des *national curriculum* bereits ca. 80 Prozent der Unterrichtszeit minutiös verplant (Tooley, 1992). Die Notwendigkeit für ein erfolgreiches Abschneiden hat vor allem finanzielle Gründe und ergibt sich aus der Tatsache, dass mit Einführung des *Education Reform Acts* die Finanzressourcen der einzelnen Schulen viel stärker als zuvor indikatorenbezogen auf einer Per-Capita-Grundlage erfolgte (van Ackeren, 2002). Vergleichbare Systeme gibt es z.B. auch in den Niederlanden, Neuseeland, Schweden und mehreren Staaten der USA. Da in England, anders als in Deutschland, aber grundsätzlich freie Schulwahl besteht und Schul- und Wohnort räumlich getrennt sein können, ist zu vermuten, dass sich Schüler wie Eltern an den im Internet veröffentlichten Ergebnissen der *assessments* orientieren (vgl. z.B. die *performance tables* unter www.dfes.gov.uk/performance-tables/). Vorausichtlich werden Schulen, die in den *performance tables* Spitzenpositionen einnehmen, mehr Anmeldungen bekommen als solche, die sich am unteren Ende der *rankings* wiederfinden. Das wiederum bedeutet, dass nur kontinuierlich gute Ergebnisse in den *assessments* das Bestehen einer Schule sichern. Dieser Aspekt kann problematisch werden, wenn Kontextfaktoren in den *rankings* nicht ausreichend berücksichtigt werden, was zumindest nach Ansicht der Lehrgewerkschaft NUT (*National Union of Teachers*: www.teachers.org.uk/) teilweise der Fall ist.

Die so herbeigeführte Konkurrenzsituation ist ein entscheidendes Kennzeichen des britischen Bildungswesens und hat dazu geführt, dass viele Schulen ihre Schüler gezielt auf die Tests hin unterrichten. Dieses *teaching to the test* stellt einen häufig vorgebrachten Kritikpunkt an regelmäßigen Tests dar. Andere Unterrichtsinhalte, so die Folge, blieben zunehmend unberücksichtigt. Tatsächlich scheint eine derartige Orientierung der Schulen durch den *Education Reform Act* und das *national curriculum* ge-

wünscht. Während die Lehrer sich hauptsächlich dem Unterrichten widmen sollen, sehen sich die Eltern der Forderung ausgesetzt, selbst wieder mehr Verantwortung für die soziale und persönliche Entwicklung ihrer Kinder zu übernehmen (Sauxter, 1995). Dies gilt insbesondere für die öffentlichen Schulen, die ihre finanziellen Mittel allein aus der staatlichen Zuteilung beziehen, während Privatschulen durch unterschiedlich hohes Schulgeld ihren Haushalt sichern. Dies ist jedoch, wie im Folgenden gezeigt wird, nicht der einzige relevante Unterschied zwischen Privatschulen und staatlichen Schulen.

4.2 Privatschulen und staatliche Schulen

Das britische Bildungssystem kennt in stärkerem Maße als das deutsche eine Unterscheidung zwischen staatlich-öffentlichen und Privatschulen. Während der Besuch der ersteren kostenfrei ist, ist der Besuch von Privatschulen eine äußerst kostspielige Angelegenheit. Im *Plymouth College*, auf das hier exemplarisch eingegangen werden soll, betrug das Schulgeld 2001 ca. 19.000 € pro Jahr und Kind (ISI, 2002a). Mit diesem Betrag waren allerdings nicht nur die Kosten für die Beschulung, sondern gleichzeitig auch die Unterbringung in den *boarding houses*, Verpflegung und sportliche Aktivitäten an den Nachmittagen oder am Wochenende abgedeckt. Schüler, die außerhalb der Schule wohnten und täglich in die Schule kamen, zahlten dementsprechend weniger.

Eltern, die ihre Kinder Privatschulen besuchen lassen, wird vor allem die Rechtfertigung des hohen Schulgeldes interessieren. Sie wollen wissen, ob in der Privatschule der Unterricht, die Lernatmosphäre und die gesamte Erziehung besser sind. Eine der Kernfragen ist deshalb, ob ihre Kinder in den Privatschulen mehr und besser lernen, als dies an staatlichen Schulen der Fall wäre bzw. ist.

Auch die englische Regierung besitzt ein Interesse an der Arbeit der Privatschulen und untersucht deren Leistungsfähigkeit, wobei sie – anders als die Eltern – insbesondere das untere Ende der Werteskala in den Blick nimmt und danach fragt, ob der Unterricht in den Privatschulen mindestens das Niveau der staatlichen erreicht und die dort erworbenen Abschlüsse dieselbe Qualität besitzen. Aus diesem Grund müssen innerhalb der Privatschulen regelmäßig Evaluationen durchgeführt, in deren Rahmen Beobachter von außen in den Unterricht der Privatschulen gehen, um die Arbeit der Lehrer und der Schule zu evaluieren.

Es gibt in Großbritannien drei Organisationen, deren Evaluationen staatlich anerkannt sind. Dabei handelt es sich zum ersten um das *Office for Standards in Education* (OFSTED), deren Berichte ebenfalls im Internet einzusehen sind (OFSTED, 2002) und zum zweiten um das *Independent School Inspectorate* (ISI) und schließlich um *Her Majesty's Inspector of Schools* (HMI). Während OFSTED grundsätzlich für die Evaluation staatlicher Schulen zuständig ist, können Privatschulen frei wählen, von welcher Organisation sie sich evaluieren lassen. Zumeist bevorzugen sie das ISI, was vor allem auf finanziellen Überlegungen basieren dürfte, denn Evaluationen durch OFSTED sind um einiges teurer als die des ISI. Dies liegt im Wesentlichen daran, dass alle Schulleiter, deren Schulen Mitglied im *Independent School Council* (ISC) sind, regelmäßig kostenlos für Evaluationen zur Verfügung stehen müssen. Um ein Beispiel zu geben: Am *Plymouth College* kostete die 2001 durch das ISI durchgeführte Evalua-

tion ungefähr 22.000 £ (ca. 31.000 €). Dies ist um ein vielfaches billiger, als wenn OFSTED die Evaluation durchführen würde. Hier würde eine solche Evaluation mindestens 60.000 £ (ca. 96.000 €) kosten. Der Grund für diese hohen Kosten liegt darin begründet, dass die von OFSTED verpflichteten Inspektoren jeweils für eine ganze Woche vor Ort sind und entsprechend entlohnt werden müssen. Darüber hinaus fallen für die Evaluatoren gegebenenfalls auch Reise- und Übernachtungskosten an. Da die Kosten für die Evaluation von den evaluierten Schulen selbst getragen werden müssen, ist es verständlich, wenn die Schulen das ISI wählen, zumal die Evaluationen durch das ISI qualitativ zumindest nicht schlechter sind als die von OFSTED durchgeführten. OFSTED wurde und wird zudem vor allem von der Lehrgewerkschaft stark kritisiert. Dabei richtet sich die Kritik nicht primär gegen die Evaluation an sich, sondern gegen Teile der Durchführung. So wünscht die nationale Lehrgewerkschaft NUT beispielsweise, dass interne Evaluationsprojekte der Schulen von OFSTED stärker berücksichtigt werden und hat darauf aufbauend detaillierte Vorschläge gemacht, wie die Evaluationen zu verbessern wären.¹ Ein gewichtiger Punkt dabei ist, dass den Lehrern Verbesserungsvorschläge gemacht werden sollen, wenn Mängel festgestellt worden sind und dabei klarer zu umreißen sei, was genau von ihnen erwartet wird.

Wenn eine Schule sich weigert, Evaluationen durchzuführen, wird eine Zwangsevaluation durch OFSTED durchgeführt. Zusätzlich kann es jederzeit zu Kurzevaluationen durch *Her Majesty's Inspector of Schools* (HMI) kommen, dessen Berichte im Internet einzusehen sind (HMI, 2002). Auch hier werden zusammenfassende Berichte über den Zustand der Schule angefertigt. Zudem sind die Abschlussprüfungen der Schüler zentral angelegt, um die Ergebnisse mit denen anderer Schulen vergleichen zu können.

4.3 Das ISI

Das *Independent School Inspectorate* ist eine Unterorganisation des *Independent School Councils* (ISC), die sich mit der Evaluation der Mitgliedsschulen des ISC beschäftigt. Alle unabhängigen Privatschulen in England und Wales, die Mitglied des ISC sind, werden in einem Zeitrhythmus von 6 Jahren evaluiert. Die Evaluation stützt sich dabei sowohl auf die Ergebnisse der Schüler in zentral angelegten und extern evaluierten Tests (*GCSE*, *A-levels*, etc.) als auch auf Informationen, die ein Evaluationsteam innerhalb einer bestimmten Evaluationsperiode vor Ort sammelt. Das Unterrichtsgeschehen steht dabei im Zentrum der Evaluation; daneben werden zahlreiche andere Aspekte schulischer Arbeit untersucht. Um die Arbeitsweise der Teams genauer darzustellen, sei beispielhaft auf die 2001 am *Plymouth College* durchgeführte Evaluation eingegangen. Der entsprechende Bericht ist im Internet kostenlos einzusehen und enthält sowohl eine Auflistung der beteiligten Evaluatoren als auch ihrer Ergebnisse (ISI, 2002a).

Bei der Betrachtung der Evaluation des *Plymouth College* sind verschiedene Aspekte von Bedeutung: Zunächst einmal ist zu klären, welche Methodik das ISI bezüglich der Planung und Durchführung der Evaluationen anwendet. Des Weiteren ist zu beschreiben, welche Bereiche anhand welcher Kriterien evaluiert werden. Schließlich

ist die Art der Rückmeldung von Evaluationsergebnissen von Interesse und die Frage nach der staatlichen Anerkennung der Evaluationen.

4.3.1 Planung und Durchführung

Die Evaluationen durch das ISI weisen bezüglich der Planung und Durchführung einige Besonderheiten auf, die im Folgenden knapp dargestellt werden sollen.

Zunächst einmal ist festzuhalten, dass die jeweils betroffene Schule in die Planung der Evaluation miteinbezogen wird. Dies geschieht insofern als sie bei der Zusammensetzung des Evaluationsteams und der Auswahl der untersuchten Bereiche ein Mitspracherecht besitzt. Die Möglichkeiten der Schule, Einfluss auf den Evaluationsprozess zu nehmen, erschöpft sich jedoch nicht mit der Planung, denn auch im Verlauf der Evaluation kann es zu Korrekturvorschlägen seitens der Schule kommen, beispielsweise wenn die Betroffenen methodische Kritik anzumerken haben. Schließlich beginnt auch der Rückmeldeprozess in der Schule, und alle Beteiligten können somit Kritik an der Vorgehensweise und den Ergebnissen üben bevor diese veröffentlicht werden. Die vorgebrachte Kritik wird geprüft und kann unter Umständen zu einer erneuten Evaluation von Teilbereichen führen. Wenn eventuellen Bedenken oder Verbesserungsvorschlägen der Schule nachgekommen wurde, werden die Befunde schließlich veröffentlicht. Allerdings werden die an einer Schule festgestellten Ergebnisse vorher nochmals mit denjenigen Lehrpersonen, die bestimmten Fachbereichen vorstehen (*heads of department*), diskutiert. Dies geschieht in der jeweiligen Schule. Durch dieses Vorgehen wird gewährleistet, dass Schulleitung und Kollegium nicht erst aus Zeitung oder Internet mit den Ergebnissen und deren Auswertungen konfrontiert werden.

Wie die aufgeführten Punkte zeigen, gibt es eine umfassende Zusammenarbeit mit den evaluierten Schulen. In diesem Sinne gilt, dass die Schulen nicht passiv einer Evaluation unterzogen werden, sondern aktiv an deren Gestaltung beteiligt sind. Dies ist selbstverständlich nur innerhalb bestimmter Grenzen möglich, wenn die Kontrollfunktion erhalten bleiben soll. Aus diesem Grund können die Schulen nicht allein darüber entscheiden, welche Bereiche an ihrer Schule evaluiert werden sollen. Dass bestimmte Bereiche an allen Schulen evaluiert werden, gewährleistet schließlich den Vergleich der Arbeit verschiedener Schulen und somit die Einstufung in den *rankings*. Dabei steht das Unterrichtsgeschehen im Zentrum der Berichte, auf die nun einzugehen ist.

4.3.2 Evaluation des Unterrichts

1. Die Evaluation der Lehrer fand auf der Grundlage von Unterrichtsbeobachtungen in allen Fächern des Lehrplans statt. Im *Plymouth College*, einer Schule mit ca. 600 Schülern und knapp über 50 Lehrern, wurden im Jahr 2001 152 Unterrichtsstunden oder Teile derselben evaluiert, und zwar innerhalb eines Zeitraumes von 5 Tagen. Das bedeutet, dass im Durchschnitt jeder Lehrer dreimal innerhalb seiner Klasse beobachtet werden konnte. Zusätzlich wurden Lehrer zu ihrem Unterricht befragt.
2. Die Evaluation der Schüler erfolgte anhand verschiedener Kriterien, die sich zum einen mit Leistungsaspekten und zum anderen mit dem allgemeinem Verhalten beschäftigten. Dies geschah zunächst anhand der mündlichen Beiträge im Unter-

richtsgespräch; darüber hinaus wurden jedoch auch schriftliche Arbeiten (Hefte, Klausuren) durchgesehen, und die so gewonnenen Ergebnisse flossen in die Gesamtevaluation mit ein. Auch die Ergebnisse in zentralen Prüfungen wurden berücksichtigt. Doch ging es nicht nur um die Beurteilung der Leistungen: Mit Bezug auf anders ausgerichtete Evaluation führten die Inspektoren formelle und informelle Diskussionen mit Schülern. Auf diese Weise war es den Evaluatoren möglich, Einblicke in das Schülerverhalten zu gewinnen, die über den reinen Leistungsaspekt hinausgingen – ein ansonsten nicht nur in Deutschland häufig vorgebrachter Kritikpunkt an Evaluation. Bei der in Großbritannien praktizierten Form von Evaluation steht somit nicht nur Leistung im Fokus des Interesses. Die hierbei gewonnenen Erkenntnisse wurden ausgewertet und gestützt durch Beobachtungen des Schülerverhaltens in den allmorgendlichen Versammlungen, in denen Lehrer und Schüler Vorträge zu beispielsweise sozialen und gesellschaftlichen Problemen halten. Darüber hinaus wird auch der Schulalltag z.B. auf dem Schulhof durch die Evaluatoren des ISI und weitere Kontextfaktoren untersucht. So wird über die Evaluatoren und ihr Vorgehen gesagt (ISI, 2002b):

- (a) „Four of the inspecting team visited the two boarding houses on Tuesday evening to inspect the accommodation and provision for boarding, and some inspectors visited the sanatorium“;
- (b) „They observed many extra-curricular and sporting activities²“;
- (c) „They analysed the responses of 287 parents to a questionnaire distributed in advance of the inspection.“

Die zuletzt genannten Aspekte sowie die Gliederung des kompletten Evaluationsberichts verdeutlichen, dass es bei der Evaluation um mehr geht als die Feststellung der geleisteten Arbeit in den Unterrichtsfächern. Dieses Vorgehen macht vor allem unter dem Aspekt Sinn, dass Privatschulen großen Wert auf die Vermittlung von Sekundärtugenden legen. Die in diesem Zusammenhang untersuchten Bereiche (Schüler- und Schulumfeld, sportliche und sonstige Aktivitäten, sozialer Hintergrund der Schüler) können in Beziehung zu den Schülerleistungen gesetzt werden – und es kann ermittelt werden, ob die erzielten Ergebnisse auf der Grundlage der komplexen Hintergründe als überdurchschnittlich gut, schlecht oder normal zu bezeichnen sind. Relative Aussagen dieser Art beinhalten Evaluationsaspekte, die durch bloße Ergebnisdarstellung, wie zum Beispiel die durchschnittlich im *GCSE* erreichte Punktzahl (ISI, 2002c), nicht gewonnen werden können. Eine Orientierung am Durchschnitt und an Schulen mit vergleichbaren Bedingungen, sogenannten *bench-marks*, wird tatsächlich in den Inter- netberichten vorgenommen. Als Beispiel sei folgende Einschätzung des ISI bezüglich der Ergebnisse in den zentralen Prüfungen am Plymouth College wieder gegeben: „In public examinations GCSE results averaged over the last three years for grades A* to C, indicate that attainment is slightly below that of comparable schools, but well above the national average for maintained schools“ (ISI, 2002b).

Um die Vergleichbarkeit zwischen den Evaluatoren zu gewährleisten, nimmt die Objektivität der Evaluationen besondere Stellung ein. Das ISI versucht, diese durch die Verwendung von standardisierten Evaluationsbögen zu gewährleisten. Solche Kriterienkataloge existieren für jeden Evaluationsbereich. Als Beispiel wird im nachfol-

genden Abschnitt der Bogen zur Beurteilung des Unterrichtsgeschehens dargestellt. Die Wahl fiel auf diesen Bereich, weil im vorgestellten Evaluationsbericht von 2001 ein besonderes Gewicht auf den Unterricht gelegt wurde.

4.3.3 Evaluationskriterien für den Unterricht

Der hier beschriebene Evaluationsbogen dient der Erfassung von Daten zum Unterrichtsgeschehen. Er soll die Ergebnisse der Evaluatoren standardisieren und zu deren Objektivität beitragen. Dabei gibt es zusätzlich zum Evaluationsbogen ein sogenanntes *Manual*, in dem die einzelnen Abschnitte des Untersuchungsbogens genauer erläutert werden. In diesem werden die jeweils relevanten Kernfragen vorgestellt.

In jedem Unterrichtsfach stehen drei Bereiche im Zentrum des Interesses: *attainment*, *progress* und *teaching*. Der erste Bereich (*attainment*), innerhalb dessen es um die Erfassung der Schülerleistungen geht, wird mit verschiedenen Richtgrößen verglichen, nämlich (a) den Anforderungen des *national curriculum* und (b) dem durchschnittlichen Leistungsniveau von Schülern vergleichbaren Alters. Bei diesen Vergleichen werden die Ergebnisse der zentralen GCSE-Prüfungen ebenso berücksichtigt wie die Erwartungshorizonte, die die Evaluatoren aufgrund vorangegangener Evaluationen gewonnen haben. Als Durchschnitt im Sinne von (b) gelten dabei die Ergebnisse, die auf der Grundlage von Evaluation an vergleichbaren Schulen gewonnen wurden. Dabei werden die zu erwartenden Standards gestaffelt. Die Staffelung erfolgt in Abstimmung mit: (1) *ages, abilities and circumstances* und (2) dem sprachlichen Hintergrund der Schüler. Während der erste Punkt im Wesentlichen die sozialen und persönlichen Hintergründe berücksichtigt, bezieht sich der zweite explizit auf die Muttersprache. Insofern scheinen hier die oft wiederholten Forderungen der Abstimmung von Standards mit Charakteristiken der Schule und des Schülers zumindest ansatzweise erfüllt. Da sich die relevanten Aspekte im jeweiligen Fach unterscheiden, gibt es für jedes Unterrichtsfach einen speziell auf das Gebiet und den *national standard* abgestimmten Evaluationsbogen.

Die Schwerpunkte des Evaluationsbogens liegen auf folgenden sechs Punkten:

1. *Stundeninhalt und Zusammenhang*
(Content/Context of Lesson: topic, activities, organisation, special features);
2. *Erbrachte Leistungen/erworbene Fähigkeiten*
(Attainment: match to (a) pupils' ages, (b) pupils' capabilities – NC levels if appropriate –; evaluate knowledge, skills and understanding – with key examples);
3. *Erzielter Fortschritt der Lernleistung*
(Progress: in knowledge, understanding and skills – for each transitional phase –; progress by pupils of different gender, ability, home language, ethnic background);
4. *Lernverhalten und Lernatmosphäre*
(Learning & Behaviour (competence as learners; motivation; concentration; behaviour; relationships; independent and co-operative working);

5. *Lehrerverhalten*

(Teaching: planning/organisation; use of time and resources; control; understanding and catering for different needs; motivation; expectations; homework; knowledge of subject and of course requirements);

6. *Weitere Inhalte*

(Other Significant Evidence: +ve or -ve, e.g. language skills, numeracy, spiritual/moral/social/cultural, staffing, accommodation, resources).

Wie aus dem vorgestellten Design des Evaluationsbogens hervorgeht, stehen verschiedene Aspekte des Unterrichtsgeschehens im Zentrum der Aufmerksamkeit. Nach der Einordnung der beobachteten Leistungen (*attainment*) werden im zweiten Bereich die Lernzuwächse (*progress*) untersucht. Wie bereits erwähnt, stellt dieser Aspekt ein geeigneteres Kriterium zur Beurteilung des Lehrers dar, als die bloße Erfassung von Leistungsständen. Sinnvoll ist zudem, dass sich die Untersuchungen nicht allein am beobachteten Unterricht orientieren, sondern auch die schriftlichen Arbeiten der letzten Wochen und Monate berücksichtigt werden. Es ist zu vermuten, dass der Lernzuwachs (*added value*) eines Schülers unter Berücksichtigung seiner Leistungen im Unterricht und den schriftlichen Arbeiten der letzten Monate relativ genau beurteilt werden kann. Der dritte Bereich (*teaching*) beschäftigt sich mit der Evaluation des Lehrerverhaltens. Bezogen auf die hohe Anzahl an Unterpunkten, scheint hier das Hauptaugenmerk der Evaluation zu liegen. Es finden sich dabei sowohl Kriterien, die auf die fachliche Qualifikation des Lehrers zielen, als auch bezüglich seines Auftretens in der Klasse und der Organisation seines Unterrichts. Anhand der dargestellten Leitfragen lässt sich eine Einschätzung des Lehrerverhaltens und seiner Qualifikation vornehmen. Leider wird diese in den Evaluationsberichten nicht explizit in Zusammenhang mit den Schülerleistungen gebracht. Dieses Defizit wird auch durch die Fülle an anderen Informationen nicht kompensiert, denn diesbezügliche Rückmeldungen wären von hohem praktischem Nutzen. Dies gilt gerade vor dem Hintergrund der dialogisch angelegten Rückmeldeverfahren, auf die nun einzugehen ist.

4.4 Protokollierung der Ergebnisse

Die durch das ISI erstellten Evaluationsberichte zeigen eine sehr detaillierte Gliederung in verschiedene Evaluationsbereiche, über die im Folgenden lediglich eine grobe Übersicht gegeben wird.

Der Bericht (ISI, 2002a) beginnt mit einer Übersicht über die Evaluationsergebnisse. Innerhalb dieses Abschnittes kommt es zu einer ersten Einschätzung der schulspezifischen Stärken und Schwächen unter den Leitfragen 1. *What the school does well* und 2. *What the school could do better*. Dabei werden die Ergebnisse vor allem am beobachteten Unterrichtsgeschehen festgemacht, wobei sowohl die Leistungen der Schüler als auch das Verhalten des Lehrers explizit aufgegriffen werden. Der erste Abschnitt des Evaluationsberichts schließt mit einer Beschreibung und Beurteilung derjenigen Schritte ab, die die Schule auf der Grundlage vorangegangener Evaluationen vorgenommen hat. In den folgenden Abschnitten werden die verschiedenen Bereiche schulischer Arbeit genauer untersucht. Dies betrifft sowohl Aspekte des Unterrichts als

auch der Schulgestaltung und des Personals. Gleichermaßen wird die Integration der Schule in ihr soziales Umfeld untersucht und der Frage nachgegangen, inwiefern sie mit anderen Gruppen (Eltern, andere Schulen) zusammenarbeitet. Auch hier besitzt die Beschreibung von Entwicklungen Priorität gegenüber der Feststellung von Zuständen. Der Bericht schließt mit einem Verzeichnis der beteiligten Evaluatoren samt deren Adressen. So können interessierte Personen weitere Informationen zu von ihnen gewünschten Bereichen erhalten.

Das Hauptinteresse des Berichts liegt auf einer umfassenden Einschätzung der Situation der Schülerschaft. Innerhalb dieses Bereiches werden neben dem Unterricht weitere Aspekte wie *Personal Development, Pastoral Care, Support and Guidance, Pupils' Welfare (including Health & Safety)* und *Boarding* untersucht. Insofern stehen nicht nur die Schülerleistungen, sondern auch die materiellen und personellen Voraussetzungen im Zentrum des Interesses, unter denen die Leistungen zustande gekommen sind. Das bedeutet vor allem eine Darstellung der Ressourcen, die den Schülern zur Verfügung stehen. Interessanterweise werden dabei auch die Schüler befragt. Dies ist bei den meisten internen Evaluationen nicht der Fall und findet auch im Rahmen externer Leistungstests häufig nicht statt.

4.5 Rückmeldungen

Als Zielgruppen der Rückmeldeverfahren werden ausdrücklich verschiedene Gruppierungen genannt. Die primären Zielgruppen sind das ISC, die angegliederten Organisationen und die Eltern (ISI, 2002a). Diese Gruppen werden über das ISI-Handbuch darüber informiert, inwieweit die untersuchten Schulen die in England vorhandenen Standards – insbesondere den Anforderungen des *national standard* – erfüllen und/oder übertreffen. Die Schulen sind dabei verpflichtet, auch Eltern und weiteren interessierten Parteien den kompletten Evaluationsbericht zuzusenden und dabei die Kosten für Porto und Kopien selbst zu tragen. Zudem können über die Internetpräsenz des ISI genaue Informationen darüber abgerufen werden, wie die einzelnen Schulen bei vorangegangenen Evaluationen abgeschnitten haben, welche Missstände damals festgestellt wurden und wie die Schule darauf reagiert hat. Auf diese Art und Weise kann die Entwicklungsgeschichte einer Schule nachvollzogen werden. Darüber hinaus erfolgt auch eine Rückmeldung der Ergebnisse an unterschiedliche Gruppen innerhalb der Schule (Schulleiter, *heads of department* und Lehrer). Des Weiteren hat es sich das ISI zum Ziel gesetzt, auch den staatlichen Organisationen Informationen über das Leistungsniveau der untersuchten Schulen zukommen zu lassen. Innerhalb dieser Tätigkeit sieht sich das ISI vorrangig als Informationsvermittler und Ratgeber (ISI, 2002d). Die Berichte des ISI werden dabei seitens staatlicher Institutionen stichprobenweise überprüft. Dies geschieht durch OFSTED, das einige ISI-Inspektionen und -Berichte überprüft, um den Standard für Qualität und Unabhängigkeit zu gewährleisten (ISI, 2002d). Durch diese fortlaufende Kontrolle wird sichergestellt, dass die Evaluationen durch das ISI qualitativ hochwertig bleiben.

Ein weiteres Ziel, das nicht direkt mit der Veröffentlichung der Ergebnisse zu tun hat, aber nichtsdestotrotz Bedeutung besitzt, ist die selbst auferlegte Verpflichtung zur Evaluation des eigenen Vorgehens. Das beinhaltet das Ziel Möglichkeiten zu suchen,

die eigenen Methoden und Serviceleistungen kontinuierlich zu verbessern und auszuweiten (ISI, 2002a). Aus dieser Formulierung geht hervor, dass das ISI Evaluation als einen fortlaufenden Prozess betrachtet, der hinsichtlich seiner Qualität ständig neu untersucht werden muss. Leider wird jedoch nicht gesagt, wie dieses Ziel erreicht werden soll.

4.6 Verzahnung interner und externer Evaluation am Beispiel Englands

Wie bereits angedeutet vereinen die Evaluation durch das ISI Elemente externer und interner Evaluationsmethoden. An interne Evaluation erinnert zunächst einmal die genaue Kenntnis der inneren Struktur der Schule sowie des Lehrpersonals und der Schulentwicklung, die die Evaluatoren durch intensive Recherche, Schulung und aufgrund vorangegangener Evaluationen gewonnen haben. Bewusstsein und Kenntnis der schulinternen Besonderheiten sind zwar nicht so detailliert, wie dies bei schuleigenen Lehrern der Fall ist, aber dies muss nicht nachteilig sein, denn die Evaluatoren des ISI besitzen so eine gewisse Distanz zur untersuchten Schule – und zwar sowohl auf der emotionalen als auch der kognitiven Ebene. Sie sind so allein der Qualitätsentwicklung verpflichtet und nehmen als Außenstehende schulinterne Besonderheiten wahr, die den Lehrern der Schule eventuell nicht mehr auffallen. Zudem besitzen die Evaluatoren aufgrund ihrer eigenen Biographie bereits Einblicke in relevante Bereiche, denn zusätzlich zu ihren eigenen Erfahrungen an und mit Schule haben die Evaluatoren des ISI im *training course* einen wissenschaftlichen Hintergrund erworben. Bei einem derartig konstituierten Evaluationsteam kann ein hohes Maß an Kompetenz angenommen werden. Demgegenüber fehlen wie auch in Deutschland (Wottawa & Thierau, 1998) derart qualifizierte Evaluatoren bei den meisten internen Evaluationen.

Bezüglich der Evaluatoren des ISI sei hier auf ein Problem hingewiesen: Da das ISI vor allem auf Beobachtungen und Berichte setzt, stellt sich die Frage, in welcher Form die Beurteilungen normiert werden. Auf einer grundsätzlichen Ebene ist dies vor allem die Frage nach der Objektivität der Ergebnisse. Die Objektivität kann durch den bereits dargestellten standardisierten Fragebogen und den *training course*, in dem die jeweils relevanten Kriterien verdeutlicht und die Evaluatoren hinsichtlich ihrer Einschätzungen überprüft werden, zwar als weitestgehend gegeben angesehen werden, jedoch ist die Objektivität auch hier nicht absolut, weil die Bewertung auf der Grundlage subjektiver Beobachtungen erfolgt. Die Subjektivität muss gegenüber den meisten internen Evaluationen, bei denen entsprechend standardisierte Verfahren und Evaluationsbögen nicht existieren, dennoch als minimiert gelten; weil sich aus diesem Grund über interne Evaluationen wenig allgemeingültige Informationen bezüglich des methodischen Vorgehens finden lassen, werden diese hier nicht weiter untersucht.

Positiv erscheint bei den Evaluationen durch das ISI, dass die Schulleistungsmessung durch das ISI in den Schulen stattfindet und nicht – wie bei externer Evaluation – anhand von in der Schule gewonnenen Ergebnissen zentral analysiert wird. Dadurch können jeweils diejenigen Bereiche fokussiert und evaluiert werden, die von besonderem Interesse sind. Dem ISI geht es nicht nur um die in zentralen Tests erzielten Ergebnisse, sondern auch um zusätzliche Aspekte. Bei der Auswahl dieser Bereiche besitzen die zu evaluierenden Schulen ein Mitspracherecht. Sie können demzufolge Auf-

träge für Bereiche erteilen, die für sie von besonderem Interesse sind, wobei das Unterrichtsgeschehen als zentraler Punkt immer in der Evaluation enthalten ist.

Insgesamt lässt sich festhalten, dass der wesentliche Vorteil interner Evaluationen, nämlich die genaue Kenntnis der schuleigenen Bedingungen und Besonderheiten, mit dem zentralen Punkt externer Evaluationen, nämlich der Betrachtung innerer Besonderheiten durch einen Außenblick, in der vorgestellten Evaluationsform vereinigt werden. Im nun folgenden Teil geht es darum, die Situation der Evaluationsprozesse in Deutschland noch einmal zusammenzufassen und zu einer Einschätzung zu kommen, inwiefern das dargestellte Verfahren des ISI Impulse für die Weiterentwicklung von Evaluationen in Deutschland geben kann.

5 Entwicklungsmöglichkeiten der Evaluation in Deutschland

5.1 Vergangenheit und Gegenwart der Evaluationsvorhaben

Deutschland kann betreffend der Durchführung von Schulleistungsuntersuchungen auf keine lange Tradition zurückblicken und beteiligte sich zumeist, wenn überhaupt, nur partiell an den internationalen Projekten (Bos & Schwippert, 2002). Demzufolge waren deutsche Schulen lange Zeit in der Lage, sich kritischen Nachfragen zu entziehen und Strukturen beizubehalten, ohne diese in regelmäßigen Abständen zu evaluieren. Diese Haltung wurde zum einen dadurch begünstigt, dass die finanzielle Ausstattung der Schulen in Deutschland nicht an die Ergebnisse von Schulleistungsmessungen gekoppelt ist – ein entscheidender Gegensatz zum englischen System. Zum anderen ist in Deutschland die freie Schulwahl eingeschränkt: In der Regel müssen Schüler eine Schule innerhalb des Bezirkes besuchen, in dem sie wohnen. Demzufolge konkurrieren die Schulen untereinander nicht in dem Maße um Schüler wie in England. Des Weiteren wurden den deutschen Schulen im Laufe der letzten Jahrzehnte zunehmend mehr Autonomierechte eingeräumt – dies ging bis zur relativ eigenständigen Verwendung der finanziellen Mittel. Obwohl aber auch autonome Schulen unter öffentlicher Verantwortung bleiben (vgl. Lange, 1999), bestand lange Zeit keine Verpflichtung, die Effizienz der autonomen Arbeit zu evaluieren – und offenbar gab es auch kein inneres Bedürfnis der Schulen dies zu tun. Diese Haltung änderte sich vor allem mit Bekanntwerden der TIMSS-Ergebnisse, in deren Folge von der Öffentlichkeit pauschal verurteilt wurde, was lange Zeit kaum interessierte: Die Qualität des deutschen Bildungswesens.

Der positive Aspekt eines großen Interesses an der Bildungspolitik wurde durch die teilweise verheerenden Rückschlüsse und ‚Analysen‘ von Personen, die keine Experten auf dem Gebiet der Schulleistungsmessung waren, mehr als aufgehoben. Allzu oft wurde ohne einen zweiten (teilweise sogar ohne genauen ersten) Blick die Arbeit von Lehrern und Schulen pauschal verurteilt (vgl. Brügelmann, 2001). Zwangsweise fühlten sich viele Lehrer ungerecht beurteilt, weil ein System als Ganzes beurteilt wurde, dass so komplex ist, dass nur differenzierte Betrachtung und Untersuchung sinnvolle Evaluation ermöglichen. Dass die ablehnende, skeptische und zuweilen feindselige Haltung vieler Lehrer gegenüber äußerer Kontrolle eine direkte Folge dieses Prozesses

war, verwundert nicht. Der sogenannte ‚TIMSS-Schock‘ (Demmer, 2000), der in der Folge entstand, stellte allerdings eine ebenso unangemessene Reaktion auf TIMSS dar, wie der Versuch, die Ergebnisse zu bagatellisieren, indem behauptet wurde, besser abschneidende Länder hätten gezielt bessere Schulen ausgewählt (Kempfert & Rolff, 1999). Beide Reaktionen trugen nicht dazu bei, um von den Ergebnissen zu einem Prozess der Verbesserung zu gelangen.

Gerade im Zusammenhang mit dem zuletzt genannten Punkt sei noch einmal darauf hingewiesen, dass einer Evaluation die Aufgabe zukommt, Anregungen zum Handeln zu geben und nicht bei der Kritik stehen zu bleiben. Da es aber in der Geschichte deutscher Evaluationsvorhaben zumeist nur zu Kritik, aber nicht zur Ermutigung kam, verstärkte sich die Ablehnung externer Kontrollen weiter. Bei vielen Lehren entstand Angst vor äußerer Kontrolle und Beurteilung, die auf ein ‚schlechtes Gewissen‘ bezüglich mangelnder Unterrichtsqualität zurückgeführt wurde. Aber auch die angewandten Evaluationsmethoden sowie die sich anschließenden (zumeist recht einseitig orientierten) Auswertungen spielen eine entscheidende Rolle in diesem Prozess.

5.2 Vorbild England?

Die durch das ISI angewandte Evaluationsmethode greift viele der genannten Punkte auf. Zunächst einmal ist die verwendete Methode transparent, und die Betroffenen wissen, nach welchen Kriterien sie bewertet werden. Sie werden den Evaluationen nicht primär ausgesetzt, sondern nehmen aktiv an Gestaltung und Durchführung teil. Dies ist vor allem insofern wichtig, als es nicht darum gehen kann, Evaluationen von oben anzuordnen oder um das Vertrauen der Lehrkräfte zu buhlen. Hauptziel sollte deshalb sein, Akzeptanz von Evaluationen durch transparente und dialogisch angelegte Verfahren zu erzielen (Kempfert & Rolff, 1999) und Qualität anhand festgelegter und nachvollziehbarer Kriterien zu entwickeln und zu überprüfen. Zu diesem Ziel erscheint die Übernahme methodischer Elemente des ISI empfehlenswert, insbesondere die Integration der Schulen in den Prozess der Planung und Durchführung sowie die Betreuung von Schulen nach den Evaluationen. Eine dem ISI ähnliche Vorgehensweise wäre auch in diesem Punkt in Deutschland anzustreben. Zudem erscheint die Aufbereitung der Daten in den Texten der Evaluationsberichte sinnvoll, weil Evaluationsdaten erst dann ihre volle Nützlichkeit entfalten, wenn die Beteiligten eine gemeinsame Einschätzung der zu bewältigenden Probleme haben (Dalin, 1996).

Vor allem im Bereich der Untersuchung von Zusammenhängen zwischen Ursachen und Wirkungen im Bereich *Schule*, also dem Entstehen von Qualität, besitzt die Evaluation durch das ISI gegenüber den in Deutschland gängigen Methoden Vorteile. Dies gilt insofern, als sich das ISI hauptsächlich auf den Kernbereich von Schulqualität, den Unterricht konzentriert, ohne dabei andere Bereiche zu vernachlässigen. Die Vielzahl der untersuchten Bereiche ist wichtig, weil auch der Unterricht und seine Qualität von bestimmten Kontextfaktoren abhängig sind. Dies gilt sowohl für die Zusammensetzung und innere Struktur der Klasse als auch für Fragen der Lehrerqualifikation und Fortbildung. Auch von Lehrern kann nicht mehr erwartet werden, als ihnen beigebracht wird. Auch ihnen muss etwas angeboten werden (z.B. Lehrerfortbildung, Methodentraining), bevor etwas verlangt wird. Dieser Punkt ist allerdings beim ISI nur

unzureichend berücksichtigt, denn *wie* die Lehrer konkret unterstützt werden können, wird auch in den Berichten des ISI nicht erwähnt. Ebenso fehlt hier die Möglichkeit, anhand der Evaluationsergebnisse kreativen Wegen zur Verbesserung von Qualität Raum zu geben. Der hohe Druck, der durch die *action plans* und die möglichen Konsequenzen auf den Schulen lastet, macht dies schwer möglich.

6 Fazit

Zusammenfassend lässt sich in Bezug auf die Chancen und Gefahren von Evaluationsgeschichten folgendes festhalten: Interne und externe Evaluationsvorhaben stellen keine unvereinbaren Größen dar, sondern bieten die Möglichkeit der Verzahnung zum Ziel der Verfahrensoptimierung. Durch eine Übernahme von Elementen beider Methoden kann, wie am Beispiel der Evaluationen durch das ISI deutlich geworden ist, eine detaillierte Analyse der schuleigenen Leistungsfähigkeit vorgenommen werden und gleichzeitig ein Vergleich mit den nationalen Durchschnittsleistungen erfolgen. In diesem Zusammenhang sind die Leistungen der Lehrer und des Systems nicht nur besonders zu fokussieren, sondern es ist auch jeweils zu überprüfen, wann und wie Unterstützung gegeben werden kann. Ausschließlich externe Analysen eröffnen hierbei keinen ausreichenden Einblick in den schulischen Alltag.

Gleichzeitig ist festzuhalten, dass eine Öffnung der Schulen und Lehrer für den kritischen Blick von Außen unumgänglich ist. Nur dann scheint Qualitätsentwicklung möglich. Dabei könnte die Schulaufsicht in Deutschland neben der Wahrnehmung von Kontrollfunktionen bezüglich der Qualität der Evaluationen auch dafür sorgen, dass Wünsche, Ziele und Bedenken der Lehrer stärker berücksichtigt werden und eine Vermittlerposition einnehmen – wie die bereits zitierte Kritik der NUT an OFSTED zeigt, ist dies zumindest in England nicht immer ausreichend der Fall.

Zusätzlich sei bemerkt, dass Evaluation kein linearer Prozess ist, sondern den Charakter eines Kreislaufes besitzt. Das heißt im Einzelnen: Ob die aufgrund von Evaluationsergebnissen vorgenommen Veränderungen die gewünschten Erfolge erzielen, ist in jedem Fall zu prüfen. Dass etwas *anders* machen nicht unbedingt bedeutet, es besser zu machen, sollte ein im Evaluationsverlauf immer präsenter Gedanke sein. Daher sollte ein Leitgedanke ‚keine Reform ohne Evaluation‘ sein. Gerade deshalb ist zu berücksichtigen, dass Evaluationsvorhaben große Anforderungen an die Lehrer stellen: Es geht nicht darum, dass sie sich kritischen Blicken und Nachfragen aussetzen, sie sollen dann auch noch mögliche Verhaltensänderungen vornehmen, deren Erfolg ungewiss ist. Aus diesem Grund ist neben die Forderungen auch eine ausreichende Unterstützung zu gewährleisten. Erst die Kombination beider Schritte sorgt für die Professionalisierung des Lehrerberufs und macht Qualitätsentwicklung transparenter.

Insgesamt lässt sich festhalten, dass hinsichtlich der jüngeren Vergangenheit bereits positive Entwicklungen zu attestieren sind. Zwar hat sich die Verzahnung von Elementen externer und interner Evaluationen noch nicht flächendeckend durchgesetzt, es ist aber ein gesteigertes Interesse von Schulen an Ergebnissen externer Evaluationen zu beobachten. Diese Ergebnisse werden häufig in schuleigene Evaluationsvorhaben miteinbezogen. Aber nicht nur auf Seiten der Schulen zeichnet sich eine stetige Weiterentwicklung von Evaluationsprojekten ab: Auch innerhalb der Wissenschaftszweige,

die maßgeblich an Konzeption und Durchführung von Systemmonitoring Studien beteiligt sind, deutet sich eine intensivere Auseinandersetzung mit der Rückmeldepraxis aus diesen Untersuchungen an. Offenbar werden Forschungsdefiziten in diesem Bereich nunmehr deutlich wahrgenommen, und es erfolgt eine kritische Hinterfragung der eigenen Position – eine Grundvoraussetzung jeglicher Evaluationsvorhaben.

Für jene, die sich mit dem Gedanken einer Evaluation tragen, sei explizit empfohlen, diese als Ganzes und nicht modular zu planen. Vorbereitung des Prozesses, Involvierung der Beteiligten, Verständigung auf mögliche Handlungsalternativen, Umfang und Form der Rückmeldung und schließlich die Planung anschließender ‚Erfolgskontrollen‘ im Rahmen weiterer Evaluationen sollten von vornherein bedacht werden. An dieser Stelle wird nochmals unterstrichen, dass nach Auffassung der Autoren Evaluationen nicht einmalig, sondern fortlaufend in einem kontinuierlichen und kontrollierten Prozess erfolgen sollte, wenn diese nachhaltig zur Qualitätsverbesserung bzw. Qualitätssicherung beitragen sollen.

Anmerkungen

1. Die entsprechenden Stellungnahmen von NUT ist im Internet unter http://www.teachers.org.uk/resources/word/OFSTED_Conult.doc zu finden.
2. Aus meiner eigenen Erfahrung kann ich (H.F.) berichten, dass es diesbezüglich ein geradezu vorbildliches Angebot gibt. Die Schule organisiert Kajaktouren, Studienreisen in alle Welt, Kletterkurse, Wanderungen ins nahegelegene Dartmoor und vieles mehr. Allerdings sollte man – im Vergleich mit der BRD – nie vergessen, dass durch das Schulgeld in Privatschulen die finanziellen Mittel ungleich höher sind als in den staatlichen deutschen Schulen. Insofern muss die besondere Situation der Privatschulen in England berücksichtigt werden, bevor Forderungen an deutsche Schulen abgeleitet werden können.

Literatur

- Ackeren, I. van (2002). Zentrale Tests und Prüfungen im Dienste schulischer Entwicklung. Erfahrungen in Deutschland sowie Beobachtungen in England, Frankreich und den Niederlanden. *Bildung und Erziehung*, 55, 59–86.
- Ackeren, I. van (2003). *Evaluation, Rückmeldung und Schulentwicklung*. Münster: Waxmann.
- Altrichter, H. & Messner, E (1998). Wenn Schulen sich den Spiegel vorhalten – fünf Schulen lernen aus und für ihre Weiterentwicklung. *Pädagogik*, 6, 30–33.
- Becker, G., Ilsemann, C. von & Schratz, M. (2001). Evaluation. Erwartungen – Erfahrungen – Herausforderungen. *Friedrich Jahresheft*, 19, 4–7.
- Bos, W. & Postlethwaite, T. N. (2000). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. In H.-G. Rolff, W. Bos, K. Klemm, H. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung (Bd. 11)* (S. 365–386). Weinheim: Juventa.
- Bos, W. & Schwippert, K. (2002). TIMSS, PISA, IGLU & Co. Vom Sinn und Unsinn internationaler Schulleistungsuntersuchungen. *Bildung und Erziehung*, 55, 5–23.
- Brügelmann, H. (2001). Kontroversen um die Schulleistungsmessung In F. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 33–44), Weinheim: Beltz.
- Dalin, P. (1996). Ansätze schulinterner Evaluation im internationalen Vergleich. In Landesinstitut für Schule und Weiterbildung (Hrsg.), *Evaluation und Schulentwicklung. Ansätze, Beispiele und Perspektiven aus der Fortbildungsmaßnahme Schulentwicklung und Schulaufsicht* (S. 311–333). Bönen: Verlag für Schule und Weiterbildung, DruckVerlag Kettler.
- Demmer, M. (2000). TIMSS- Schock und PISA-Wellen. *Die Deutsche Schule*, 92, 262–265.

- Frisch, M. (1965). *Tagebuch 1946–1949*. Frankfurt a.M.: Suhrkamp.
- Helmke, A. (2003). *Unterrichtsqualität*. Seelze: Kallmeyer.
- HMI. (2002). Verfügbar unter: <http://www.archive.official-documents.co.uk/document/hoc/157/157-00.htm> [28.08.2001].
- ISI. (2002a). Verfügbar unter: <http://www.isinspect.org.uk/frreports.htm> [16.07.2002].
- ISI. (2002b). Verfügbar unter: <http://www.isinspect.org.uk/frreports.htm> [21.07.2002].
- ISI. (2002c). Verfügbar unter: <http://www.isinspect.org.uk/frreports.htm> [22.07.2002].
- ISI. (2002d). Verfügbar unter: <http://www.isinspect.org.uk/fraboutp.htm> [16.07.2002].
- Kempfert, G. & Rolff, H.-G. (1999). *Pädagogische Qualitätsentwicklung*. Weinheim: Beltz.
- 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, H. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung (Bd. 11)* (S. 387–419). Weinheim: Juventa.
- Lange, H. (1999). Qualitätssicherung in Schulen. *Die Deutsche Schule*, 91 (2), 144–160.
- Mühlhausen, U. (1999). Gegenseitige Hospitation im Unterricht – ein unheimlicher Fortbildungswunsch von Lehrerinnen und Lehren. *Die Deutsche Schule*, 91 (2), 199–214.
- OFSTED. (2002). <http://www.ofsted.gov.uk> (16.07.2002).
- Sauxter, B. (1995). Der Education Reform Act aus dem Jahre 1998 – eine kritische Annäherung. *Die Deutsche Schule*, 87, 228–237.
- Tooley, J. (1992). The ‘Pink Tank’ on the Education Reform Act. *Journal of Educational Studies*, 40 (4), 163–172.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Bern: Huber.