

Boenicke, Rose

Leistung und Qualitätssicherung. Gegen Verengungen der Zielsetzung von Schule

Die Deutsche Schule 92 (2000) 2, S. 139-151



Quellenangabe/ Reference:

Boenicke, Rose: Leistung und Qualitätssicherung. Gegen Verengungen der Zielsetzung von Schule - In: Die Deutsche Schule 92 (2000) 2, S. 139-151 - URN: urn:nbn:de:0111-pedocs-276023 - DOI: 10.25656/01:27602

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-276023>

<https://doi.org/10.25656/01:27602>

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

Leistung und Qualitätssicherung

Gegen Verengungen der Zielsetzung von Schule

Skeptische, insbesondere durch TIMSS geschürte Nachfragen nach der *Qualität von Schule und Unterricht*, Forderungen nach verstärkter Qualitätsentwicklung und Qualitätssicherung – all dies verdeckt, dass der *Qualitätsbegriff selbst höchst unbestimmt ist* und von unterschiedlichen Interessengruppen völlig verschieden definiert wird. Ihn an quantifizierenden, auf Schülerleistung bezogenen Aussagen festzumachen, bietet nur scheinbar einen Ausweg, denn solche Zahlen lassen sich nur schwer zum produktiven Ausgangspunkt für die Diskussion von *Prozessqualitäten* und damit der inhaltlichen Bestimmung von Qualität machen – eher wirken sie stigmatisierend. Vor allem im Ausland wurde deshalb eine Reihe von Ansätzen entwickelt, diese Klippe zu umschiffen, indem man versucht, quantitative und qualitative Verfahren zu integrieren. Insofern versammelt sich zurzeit unter dem Begriff des Leistungsvergleichs ein Spektrum sehr verschiedener und zum Teil gegenläufiger Ansätze.

1. Schülerleistung als „Produktqualität“?

In der Diskussion der Frage, was eine gute Schule sei, standen lange Zeit die *programmatischen Entwürfe* im Vordergrund. Gab es an Schulen Anstrengungen, sie in die Tat umzusetzen, so galt dies selbst schon als Qualitätsausweis. Schon gar nicht wurde nach dem Verhältnis von investierten Mitteln und Ertrag im Sinne der Effizienz gefragt. Dies hat sich mit TIMSS schlagartig geändert – wobei die Ergebnisse dieser Studie wohl nur der Anlass, nicht der Grund für den plötzlichen Fokus auf messbare Ergebnisse sind: Vorangegangene Studien mit ähnlichen Ergebnissen waren bekanntlich weitgehend unbeachtet geblieben.

Das vergleichsweise schlechte Abschneiden deutscher Schülerinnen und Schüler in der TIMS-Studie hat mit einem Schlage wieder unser Bildungssystem in die öffentliche Diskussion gebracht. Dass *im Zuge der Sparpolitik* danach gefragt wird, ob die Resultate der Arbeit an Schulen zu den aufgewendeten Mitteln in angemessener Relation stehen, ist das Eine, dass *im Zuge der Ökonomisierung aller gesellschaftlichen Bereiche* diese Ergebnisse als in Zahlen ausdrückbare Lernprodukte definiert werden, ungleich problematischer. Zweifellos ist es legitim, Nachweise dafür zu verlangen, „dass in der Sache auch wirklich erreicht wird, wofür Mittel gefordert und bewilligt werden“ (Lange 1999, 145) Nur: Um welche „Sache“ geht es?

Die Diskussion hat gezeigt, dass sich aus den Ergebnissen der Studie *höchst unterschiedliche Folgerungen* ziehen lassen, und welcher jeweils das Wort geredet wird, hat weniger mit der Studie selbst zu tun als mit dem jeweiligen zugrunde gelegten Leistungsbegriff und Bildungsverständnis und damit der je-

weiligen Vorstellung, worin Qualität von Schule und Unterricht vorrangig zu suchen ist. Dass diese Vorstellungen weit auseinander gehen, zeigt sich in dem Umstand, dass aus TIMSS gleichzeitig abgeleitet wird, „mehr Selbständigkeit und Kreativität im Lehren und Lernen zu fordern und andererseits eine landesweite ‚Testkultur‘ entwickeln zu wollen“ (Hagemeister 1999, S.160), zwei Forderungen, die nur sehr schwer zu vereinbaren sein dürften.

Qualität von Schule und Unterricht zu fordern, bedeutet die Wahl eines Begriffs, der einen hohen Grad an Plausibilität und Zustimmungsfähigkeit hat und insofern diese unterschiedlichen Auffassungen eher verdeckt. Denn der Qualitätsbegriff, der derzeit eine so beachtliche Konjunktur hat, erweist sich bei näherer Betrachtung als wenig hilfreich. Eher droht er zu einer „mächtigen Leerformel“ (Böttcher 1999, 20) zu werden. Seine Unbestimmtheit, aber grenzenlose rhetorische Ausbeutbarkeit verdeckt, dass „Qualität von Schule“ unter der Hand in den letzten Jahren eine folgenschwere Umdeutung erfahren hat. Ging es Ende der achtziger, Anfang der neunziger Jahre vor allem darum, die Vielfalt von Dimensionen zu bestimmen, deren Zusammenspiel von Schulqualität sprechen lässt, Qualität also an bestimmten Prozessmerkmalen festzumachen, so wird gegenwärtig in den öffentlichkeitswirksamen Diskussionen mehr und mehr der „output“ zum ausschlaggebenden Qualitätskriterium, gemessen als Schülerleistung oder als Leistung der Schule im Systemvergleich. Gemessen an den Kompliziertheiten, die mit dem Evaluationskonzept einhergehen, hat diese Strategie der Qualitätssicherung den Charme des Einfachen und Vertrauten, da sie ihre Logik aus der Analogie zum Warentest schöpft (vgl. Brügelmann 1999, 39).

Während sich an die Frage, was unter Leistung und Bildung verstanden werden soll, in der Vergangenheit kaum beizulegende Kontroversen knüpften, wird der Qualitätsbegriff, der derzeit im Vordergrund steht, wie eine *einfach bestimmbare Messgröße* gehandhabt, über die weitgehend Konsens besteht. Offensichtlich liegt seine Attraktivität vor allem in der Analogie zur Güterproduktion, für die sich Qualitätskriterien ja recht eindeutig bestimmen lassen. Entsprechend werden Qualitätsstandards vor allem auf die so genannten output-Merkmale bezogen, auf Testleistungen und das Abschneiden in Leistungsvergleichen. Der Auftrag der Schule geht aber bekanntlich weit über dies hinaus. An einem Qualitätsbegriff zu arbeiten, der die komplexen Anforderungen, denen Schule in unserer Gesellschaft gerecht werden muss, adäquat erfasst, hat sich seit den achtziger Jahren der Arbeitskreis „Qualität von Schule“ (vgl. Steffens/Bargel 1987ff, 1993) zur Aufgabe gemacht.

Die Gründe für diesen Versuch, Schulleistung an „*Produktqualitäten*“ festzumachen, führen Posch und Altrichter auf zwei Entwicklungen zurück: „Derzeit besteht international ein starker Druck auf Verwendung zentraler Testsysteme für die Qualitätsevaluation, nicht zuletzt deshalb, weil sie einem marktwirtschaftlichen Denken entsprechen und ähnlich den Bilanzen der Unternehmen einfache quantitative Vergleichsmöglichkeiten zu eröffnen scheinen. Sie kommen auch den Autonomisierungstendenzen im Bildungswesen entgegen, weil sie dem mit der Verringerung an Input-Kontrolle verbundenen Machtverlust der Verwaltung entgegenwirken“ (Posch/Altrichter 1997, 40)

Qualität von Schule und Unterricht in erster Linie an *messbaren und in Tests erbrachten Schülerleistungen* festmachen zu wollen, wirkt auf den ersten Blick

und vor allem für den nicht mit pädagogischen Fragen Befassten selbstverständlich, verkürzt den Qualitätsbegriff aber auf dreifache Weise:

- Der Produktaspekt verdrängt die Frage nach den *Prozessqualitäten*, damit aber vor allem nach der pädagogischen Gestaltung der Schule. Weder geraten dabei die Bedingungen, unter denen Schülerinnen und Schüler lernen, in den Blick, noch die unterschiedlichen Ebenen und Zielsetzungen ihrer Lernprozesse, etwa auf der Ebene sozialen Lernens oder des Selbstkonzepts. Dass Schule sowohl bilden als auch erziehen soll, droht insofern zum bloßen Lippenbekenntnis zu werden, zumal Schülerinnen und Schüler ja sehr deutlich zwischen den bewerteten und nicht bewerteten Anteilen des Unterrichts unterscheiden. Werden die abtestbaren Leistungen vollends zum wesentlichen Ziel des Unterrichts erhoben, so bedeutet dies eine Verödung der pädagogischen Kultur. Wie wenig sie ein verzichtbarer Luxus ist, wird deutlich, wenn man erfolgreiche Schulen näher betrachtet: „Schulen, die gerade im überfachlichen Bereich gute Arbeit leisten und dabei Lern- und Erziehungsbedingungen sicherstellen, in denen sich die Schüler aufgehoben fühlen, in denen man ihnen Beachtung schenkt, in denen ein positives Schul- und Lernklima sichergestellt wird und in denen auch darauf Wert gelegt wird, dass die Kinder und Jugendlichen etwas lernen – solche Schulen sind nach wissenschaftlichen Untersuchungen zugleich auch Schulen, die in den klassischen Fachleistungen (in Tests ermittelt) gut abschneiden. Umgekehrt gilt jedoch nicht, dass Schulen, die ausschließlich Wert auf gute Fachleistungen legen, damit auch positive Wirkungen im Verhaltensbereich und Bereich überfachlichen Lernens mit erzeugen. Hier waren häufig auch durchwachsene und schlechte erzieherische Verhältnisse anzutreffen.“ (Steffens/Markstahler 1999, S. 217/18)

- Eine Konzentration auf Tests bringt für den *Leistungsbegriff selbst eine starke Verengung* mit sich. Nur der schmale Ausschnitt der mit Papier und Bleistift zu erbringenden Kompetenzen ist erfassbar. Es geht um Fachleistungen und hier wiederum um kognitive Leistungen, da sie am ehesten bewertbar sind. Hierbei zählt allein das hervorgebrachte Produkt; Prozessaspekte wie etwa unterschiedliche Lösungswege oder die Stellung dieser Leistung innerhalb der bisherigen Lerngeschichte finden keine Beachtung. Auf Klausuren und Tests basierende Leistungsmessung muss Messbarkeit und damit Vergleichbarkeit herstellen; zwangsläufig muss sie dazu Schülerleistungen standardisieren. Kreativität im Umgang mit Problemen, individuelle, neue Lösungen sind hier eher unerwünscht. Leistungsmessung kann in der Regel nicht umhin, von vornherein festzulegen, was eine richtige und falsche Antwort ist. Schüler lernen, dass Erfolg bedeutet, Antworten zu geben, die bereits bekannt sind, denn nur im vornherein als richtig festgelegte Antworten sind bewertbar. Dies war so lange funktional, so lange Schule eine starke gesellschaftliche Selektions- und Platzanweiserfunktion hatte, wie groß die Zweifel an der Funktionalität gegenwärtig sein müssen, wird deutlich im nächsten Punkt.

- Lernen und Situationen der Leistungsmessung unterliegen so unterschiedlichen Bedingungen, dass sie sich eher behindern als gegenseitig stützen, zumindest wenn man von nachhaltigem und verständnisorientiertem Lernen ausgeht. Die *kurzzeitige Abfolge von Wissensaneignung und -überprüfung* behindert den langfristigen Aufbau von kumulativ aufeinander bezogenen Wissenssegmenten, denn dafür ist ein Erkennen übergreifenden

der Zusammenhänge die Voraussetzung. Darauf, wie abträglich eine Einbettung in Quasi-Prüfungssituationen (wie bei manchen Formen des fragend-entwickelnden Unterrichts) für die Lernfähigkeit ist, hat Weinert wiederholt aufmerksam gemacht. „Obwohl Leistung in der Regel eine Funktion vorausgehenden Lernens ist, unterliegen Leisten und Lernen völlig unterschiedlichen psychologischen Gesetzmäßigkeiten... Wer sich subjektiv in einer Leistungssituation wähnt, bemüht sich in erster Linie darum, Erfolge zu erzielen und Misserfolge zu vermeiden. Man denkt und handelt in einem motivational gespannten Feld, weil man sich in Leistungssituationen notwendigerweise bewähren kann oder persönlich versagen muss. Niemand ist in dieser Situation motiviert, Neues zu lernen, Wissenslücken zu schließen oder unklar Gebliebenes doch noch zu verstehen; subjektiv kommt es vielmehr darauf an, das Gewusste zu aktivieren, mangelndes Wissen nicht preiszugeben, Fehler zu vermeiden und sich selbst in einem günstigen Licht zu präsentieren. Im Vergleich dazu sind Lernsituationen in der Regel entspannter, offener, sach-, informations- und problemzentrierter. Man möchte etwas wissen, entdecken, erfassen; Fehler brauchen nicht vermieden zu werden, wenn man aus ihnen lernen kann; andere Schüler sind nicht Konkurrenten, sondern Lernpartner; der Lehrer erscheint nicht als Prüfinstanz, sondern in seiner Funktion als pädagogisch-psychologischer Berater. (...) Aufgabe eines guten Unterrichts ist es, Lern- und Leistungssituationen im Bewusstsein der Schüler so zu separieren, dass eine produktive Lernkultur entstehen kann.“ (Weinert 1998, S.109/10)

Während schulisches Lernen traditionell immer Gefahr lief, sich darauf zu reduzieren, wie Leistungssituationen am besten, d.h. unter Einsatz aller Mittel, gemeistert werden können, wurde in der jüngsten Vergangenheit verstärkt darüber nachgedacht, wie die von der Schule initiierten Lernerfahrungen *selbst als sinnvoll* erlebt werden können. Sie sollen für jeden Schüler und jede Schülerin als substantieller Zuwachs an Kompetenzen wahrnehmbar und damit *anschlussfähig für weitere Lernprozesse* werden. Hintergrund dieser Forderung ist freilich weniger ein Interesse an der Mündigkeit der Lernsubjekte, sondern ein objektiv-gesellschaftlicher Grund: „die Unbestimmbarkeit einer sich beschleunigt entwickelnden Wissensgesellschaft“ (BLK 1997, S.9), in der Problemlösekompetenz, Lern- und Innovationsfähigkeit zum Standortfaktor wird. Wo unter Qualität von Schule vor allem die Qualität der von ihr ermöglichten Lernprozesse bedeutet, geraten denn auch neue, sehr viel differenziertere Formen der Leistungsrückmeldung und der Selbstevaluation durch die Lernenden ins Blickfeld.

Selbstkritisch gibt jedoch Tillmann zu bedenken, dass auch die „reformorientierte Schulforschung in der Bundesrepublik“ nicht unschuldig ist an den beschriebenen Verengungen. „Forschung zur Qualität von Schule betreiben wir seit etwa 20 Jahren mit durchaus respektablen Ergebnissen. Dabei haben wir allerdings *stets einen großen Bogen um das Kriterium ‚Fachleistung‘* gemacht, uns umso intensiver aber mit Schulklima, Schulfriedenheit, Lehrkooperation etc. befasst. (...) Ob und welche Zusammenhänge zwischen ‚unseren‘ Prozessvariablen von Schulqualität und der fachlichen Leistung besteht haben wir ... so gut wie nie untersucht (Ausnahme: Fends Vergleichsstudien in den 70er Jahren). Deshalb können wir auch kaum empirisch gesicherte Argumente ins Feld führen, wenn polemisch zwischen ‚Kuschelecken-Pädagogik‘ und ‚Leis-

tungsförderung' unterschieden wird.“ (Tillmann 1999, S. 82) Insofern, als „das große und teure Schulsystem auch auf dieser Ebene rechenschaftspflichtig ist“, sieht Tillmann durchaus eine Berechtigung der Leistungsvergleiche – allerdings weist er darauf hin, dass sie dieser Funktion erst dann gerecht werden würden, wenn ihnen nicht mehr an Aussagekraft aufgebürdet würde, als sie leisten können. „Leistungsvergleichs-Untersuchungen sind als Diagnose brauchbar, sie können auf eine – gelegentlich sehr unbequeme – Realität verweisen. Für die Schulentwicklung bedeutsam werden sie allerdings erst, wenn sich daran Überlegungen und Maßnahmen zur Qualitätsverbesserung anschließen. (...) Dies schließt ein, Verfahren der Selbstevaluation so auszubauen und zu konkretisieren, dass sie von Schulen auch real genutzt werden können. (...) In dem Maße, in dem mehrdimensionale schulinterne Evaluationen auch real durchgeführt und veröffentlicht werden, wird die Öffentlichkeit nicht mehr allein auf Leistungsvergleichs-Studien verwiesen sein“ (ebda, S.82/83).

All dies zusammengefasst geht es darum, ...

... einen *entfalteten Qualitätsbegriff* aufrechtzuerhalten, der Bildung und Erziehung, Prozess und Produkt, input und damit Voraussetzungen ebenso wie output und damit Wirkungen gleichermaßen umfasst, und dieses Qualitätsverständnis vor Verengungen zu schützen;

... Leistungsvergleiche nur als *einen von mehreren Indikatoren* zu gebrauchen, da sonst ihrer Aussagekraft zu viel zugemutet wird, und gleichzeitig an einer Verbesserung dieses Instruments zu arbeiten, damit es überhaupt messen kann, was es zu messen vorgibt;

... die Forderung nach Rechenschaft unseres Schulsystems nicht dazu zu missbrauchen, Unterricht mehr und mehr unter das *Diktat kurzfristig zu erzielen der Testleistungen* zu stellen und damit Ansätze zu einer Lernkultur wieder zurückzufahren, die auf komplexere Leistungen zielt, als sich in Tests erfassen lassen;

... die Perspektive von Schülerleistungen auf die *Leistungsfähigkeit einer Schule insgesamt* zu erweitern und dazu Möglichkeiten der Selbst- und Fremdevaluation weiterzuentwickeln. Damit geht einher, das Bewusstsein der komplexen Leistungen von Schule in der Öffentlichkeit zu stärken.

2. Strategien der Qualitätsüberprüfung

Diese Forderung nach Rechenschaft wird gegenwärtig *auf ganz unterschiedliche Weisen beantwortet*. Sie alle zielen auf Leistungsvergleiche – sei es auf Systemebene, sei es auf der Ebene der individuellen Schülerleistung.

- *Zentrale Abschlussprüfungen* finden sich in Baden-Württemberg, Bayern, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen. Im internationalen Maßstab überwiegen sie ohnehin und haben hier z.T. die Gestalt einer externen Leistungskontrolle durch staatliche und/oder private Testbüros im Auftrag der „Abnehmer-Institution“.

- In der Mehrzahl der Bundesländer werden zurzeit erste Erfahrungen mit *Vergleichsarbeiten* in unterschiedlichen Klassenstufen und Fächern gemacht. Dabei reicht das Spektrum von der verbindlichen Vorschrift über das freiwillige Engagement von Schulen im Rahmen interner Evaluationsvorhaben bis hin zu punktuellen Forschungsprojekten. Zum Teil wird mit vorgegebenen Aufgaben gearbeitet, zum Teil werden Musteraufgaben zugrunde gelegt. Schwerpunkt

für ihren Einsatz ist die Sekundarstufe und die Fächer Mathematik, Deutsch und Englisch.

- Zunehmende Bedeutung erlangen unterschiedliche *Ranking-Verfahren*, zu denen auch die internationalen Vergleichsuntersuchungen, etwa TIMSS und PISA gehören. Schulleistungen einer Region, eines Landes oder sogar international über die Vergabe von Listenplätzen zu dokumentieren, ist von hoher Öffentlichkeitswirksamkeit; zudem besteht kaum Dissens darüber, dass die Aufgaben tatsächlich voraussetzende Kompetenzen abfragen. Darüber, wie die Lösung zustande kam – „ob sie (mechanisch auswendig gelernt) abgerufen oder (selbständig erarbeitet) mit Verständnis entwickelt wurde“ (Brügelmann 1999, 43) – sagt z.B. TIMSS gar nichts aus, während PISA gerade diese Dimension zu erfassen beansprucht. Da hier zugleich eine große Zahl sozialer und kultureller Kontextvariablen erfasst werden sollen, versteht sich diese Studie als komplexes Diagnose-Instrument, aus dem sich keine einfachen Aussagen über Ranking-Plätze ableiten lassen.

- *Externe Evaluation* wird in Deutschland fast nur im Rahmen von Modellversuchen, Schulversuchen und anderen Forschungsprojekten eingesetzt und liegt dann in der Hand der begleitenden Wissenschaftler/innen. Dies geschieht in der Regel in Zusammenarbeit mit der Schulaufsicht, während eine Evaluation allein durch die Schulaufsicht von keinem Bundesland als praktikabel eingeschätzt wird.

- *Inspektionsmodelle* entweder ausschließlich als externe Evaluation oder in Zusammenspiel mit internen Evaluationsmaßnahmen finden sich vor allem in England, Schottland, den Niederlanden und den Skandinavischen Ländern. Schulinspektion ist in einigen Ländern Aufgabe staatlicher Institutionen, in England „wurde das traditionelle Her Majesty's Inspectorate (HMI) abgeschafft und durch eine neue nach marktwirtschaftlichen Kriterien operierende halbstaatliche Institution ersetzt: das Office for Standards in Education (OFSTED).“ (Posch/Altrichter 1997, 35)

3. Wirkungen und Nebenwirkungen internationaler Leistungsvergleiche

Im Folgenden möchte ich zunächst kurz auf *internationale Leistungsvergleiche*, wie sie durch TIMSS und PISA repräsentiert werden, eingehen, bevor ich dann Modelle externer und interner Evaluation, die auf die konkrete Schule bezogen sind, darstelle.

In der Diskussion der stattgefundenen und der zu erwartenden Leistungsvergleiche – und hier spielt kaum eine Rolle, ob sie lokalen oder internationalen Zuschnitts sind – wird deutlich, dass sie nur *schwer für Schul- und Unterrichtsentwicklung nutzbar* zu machen sind. Genauer: Während direkte Konsequenzen für Unterrichtsentwicklung nicht ableitbar sind – und auch die Konzeption des BLK-Modellversuchs „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ ergibt sich nicht schlüssig aus den Daten und lag übrigens schon vorher bereit – sind mittelbare Auswirkungen auf Unterricht durchaus zu erwarten und werden erheblich sein, wenn sich Vergleichsuntersuchungen vom Sonderfall zur Dauererhebung auswachsen. In diesem Zusammenhang wird vor allem auf die folgenden Kritikpunkte immer wieder hingewiesen:

3.1 Leistungstests erfassen ein zu schmales Segment

- Leistungsvergleichen wird vorgeworfen, dass sie nur Resultate, „das Kreuzchen an der richtigen Stelle“, nicht aber die *Qualität von Lösungswegen und Denkprozessen* erfassen. Diese Gefahr scheint bei großen, wissenschaftlich geleiteten Vergleichsuntersuchungen weitgehend ausgeschaltet, da sich hier die Testkonstrukteure explizit zur Aufgabe gemacht haben, insbesondere Problemlösefähigkeiten zu erfassen. Hingegen ist dies kaum sichergestellt bei ad hoc entwickelten Tests.

- Dennoch erfassen auch die internationalen Survey-Studien in der Regel *nur Lernresultate, nicht aber Lernausgangslagen* und den von einer Schule ermöglichten *Wissenszuwachs* und gehen damit an der eigentlichen Lern- und Lehrleistung vorbei. PISA versucht auf diese Kritik zu reagieren, indem zumindest eine große Anzahl soziokultureller Kontextvariablen der einzelnen Schülerinnen und Schülern mit den Testleistungen in Verbindung gebracht werden.

- Die erbrachte Testleistung lässt wenig *Prognosen* über die Anwendbarkeit dieses Testwissens in schulfremden, anspruchsvolleren Situationen zu. Ihre Transferierbarkeit ist in der Regel gering. PISA versucht deshalb den Ausschnitt der gemessenen Testleistungen auf verständnis-, problem- und anwendungsorientierte Aufgaben zu verschieben – wie weit dies in einer Testsituation gelingen kann, bleibt abzuwarten.

3.2 Damit laufen Leistungstests Gefahr, zu einem verkürzten Verständnis von Unterricht und Schule beizutragen

- Leistungstests definieren einen bestimmten *Ausschnitt der Ergebnisse schulischen Lernens* als Leistung. Dabei handelt es sich um fachliches, kognitives paper-and-pencil-Wissen. Andere Lernleistungen bleiben nicht nur unerfasst, sondern verlieren in der Konsequenz an Wertschätzung. In der Tendenz setzen so eher die Testkonstrukteure als die pädagogisch Handelnden Standards dafür, was relevantes Wissen ist. Als Rückwirkung auf den Unterricht hieße dies: Vernachlässigung von nicht-testrelevanten Lernerfahrungen oder auch ein direktes „teaching to the test“.

- Damit befördern Leistungstests die ängstliche Fixierung auf eine vorab feststehende „*einzig richtige Lösung*“; Fehler werden als Versagen wahrgenommen, nicht als Lernanlass. In einer von raschen Veränderungen, Komplexität und Uneindeutigkeit geprägten Gesellschaft hat dies schon fast den Charakter der Lernverhinderung, vermittelt auf jeden Fall wichtige Kompetenzen nicht.

- Anders gesagt: Leistungsvergleiche legen nahe, sich eng an *vorgeschriebene Lehrpläne* zu halten; selbständiges, interessengeleitetes Lernen wird zum Risiko. Extra-curriculare Aktivitäten geraten vollends ins Hintertreffen. Auch hier versucht PISA einen anderen Weg einzuschlagen, indem die Aufgabenstellungen Distanz zu den nationalen Curricula waren, und eher Eichsicht und Verständnis überprüfen, jedoch dürfte dieser komplexe Ansatz kaum übertragbar auf den „Normalfall“ des Leistungstests sein.

3.3 Von Leistungstests gehen damit kaum konkrete Anstöße zu Schulentwicklung aus

- Da Leistungsvergleiche nichts über die *Bedingungen* aussagen, unter denen Wissen erworben wurde, geben sie wenig diagnostische Hinweise zur Qualität des Unterrichts. Dadurch bieten sie auch wenig Handhabe für diesbezügliche Verbesserungen.
- Alle Qualitätsentwicklung geschieht in der *einzelnen Schule* und geht von den Lehrerinnen und Lehrern aus. Für sie enthalten die groß angelegten Leistungsvergleiche kaum motivierende Impulse und noch weniger konkrete Hinweise auf Veränderungsmöglichkeiten. „Aus der Lehrerperspektive bleibt widersprüchlich, dass man einerseits die eigene Schule eigenständig weiterentwickeln soll, andererseits aber mit einem zentral definierten Leistungstest gemessen wird.“ (Tillmann 1999, S.81)

PISA versucht wie gesagt aus dieser vor allem auf TIMSS gemünzten Kritik Konsequenzen zu ziehen, indem *Kontextvariablen stärker berücksichtigt* werden und die Testaufgaben beanspruchen, „Verständnis, Problemlöse- und Anwendungsfähigkeit“ (Schweitzer 1999, S. 135/6) zu erfassen. Mit Hilfe von Kontextfragebögen sollen Daten zu den Lern- und Lebensbedingungen mit fachlichen Kompetenzen in Beziehung gesetzt werden. Damit soll gleichzeitig deutlich werden, „in welchem Maße Schulsysteme und Schulen in der Lage sind, Unterschiede im sozio-ökonomischen Status der Schülerinnen und Schüler auszugleichen oder welche anderen Faktoren das Leistungsniveau beeinflussen“ (ebda., S.136). Doch auch für diese Studie gilt, dass sie die entscheidenden Fragen, die eine Qualitätsentwicklung des Unterrichts betreffen bzw. in welcher Richtung sie zu suchen ist, nicht beantworten können. Dies ließe sich nur in Erfahrung bringen durch „Begleitforschungsstudien ..., in denen tatsächlich bestimmte Innovationen auf der Ebene materieller Ausstattung, auf der Ebene des Schullebens, auf der Ebene des Unterrichtens oder der Lehrerqualifikation systematisch variiert und in ihrer längerfristigen Wirkung untersucht würden.“ (Preuss-Lausitz 1999, S.57/8)

4. Das englische Inspektionsmodell

Während die Schwäche der Leistungstests gerade darin liegt, über *Prozessqualitäten* wenig Aufschluss zu geben, ist dies das Hauptanliegen jener anderen groß angelegten Qualitätssicherungsstrategie, der *Evaluation von Schule und Unterricht*. Unter diesem Titel lassen sich eine Reihe sehr unterschiedlicher Überprüfungsverfahren subsumieren – von Strategien ausschließlich schulinterner Evaluation, die Lehrerinnen und Lehrer selbst entwickeln, durchführen und auswerten, bis hin zu Maßnahmen der Schulinspektion als durchgängig externer Evaluation, wie sie ab 1992 in England etabliert wurde. Dazwischen liegt ein breites Spektrum an Mischverfahren – repräsentiert etwa durch Schottland, Dänemark oder Schweden. Vermutlich enthalten diese Mischungsverhältnisse sehr viel mehr Anregungen für unsere Praxis als das englische Beispiel, das auf den ersten Blick dem Tagtraum eines preußischen Verwaltungsbeamten entsprungen zu sein scheint und auf den zweiten Blick durch seinen Widerspruch zwischen einer sehr weitgehenden Deregulierung im Sinne der Marktorientierung und andererseits „absoluter staatlicher Kontrolle“ (Aspland 1999, S.293) verwirrt.

Ausgangspunkt des rigorosen *Interventions-Systems der englischen Schulinspektion* war offensichtlich die Feststellung in den achtziger Jahren, dass 40% der Absolventen des englischen Schulsystems in ihren Leistungen nicht einmal Mindestanforderungen erfüllten. Die Thatcher-Regierung beantwortete dies mit einem Maßnahmenkatalog, der von einer Curriculumreform über neue Standards der Beurteilung von Lehrkräften und ein neues Schulinspektionsmodell bis zur Öffnung der Schule für den Markt reichte. Die Neumodellierung der Schulinspektion war geknüpft an die Auflösung der alten regionalen Her-Majesty's-Inspectorate-Büros. An seine Stelle trat 1993 ein zugleich halbstaatliches, aber über genaue Verfahrensvorgaben sehr viel direkter der Regierung unterstelltes „Office for Standards in Education“ (OFSTED). Es operiert nach marktwirtschaftlichen Kriterien und seine Angestellten sind mit Ausnahme des Leiters eines Teams nicht fest angestellt. Sie wechseln häufig und wurden z.T. nur über den Besuch eines dreitägigen Kurses für diese Aufgabe qualifiziert. Auch „besteht für die Schule die Möglichkeit, aus einem Markt von Angeboten selbst ein Inspektionsteam auszuwählen.“ (Posch/Altrichter 1997, S. 35) Andererseits ist das Vorgehen staatlicherseits bis in die kleinste Detail in einem Handbuch fixiert: „Das Handbuch schreibt ein hoch formales und detailliertes Verfahren vor ohne Flexibilität in der Anwendung oder Interpretation durch das Inspektorenteam.“ (Aspland 1997, 280, m.Übers.R.B.) Es legt fest, was zu prüfen ist, welche Beurteilungsstandards anzuwenden sind und wie der Prüfungsprozess zu gestalten ist. Diese Handbücher sind im Buchhandel für jeden erhältlich und erfreuen sich hoher Auflagenziffern.

Da jede staatliche Schule im 4-Jahres Rhythmus inspiziert wird, bedeutet dies die Inspektion von 6000 Schulen pro Jahr und dies wiederum einen Stab von 5000 Inspektoren, was das OFSTED zu einer *schwerfälligen und kostspieligen Institution* macht („a massive and unwieldy bureaucracy which is unlikely to be sustainable.“ Aspland 1997, S.283). Die Schule wird von einem Inspektorenteam von bis zu 16 Personen unter Leitung eines Registered Inspectors ca. sechs Tage lang besucht. Die große Personenzahl ist nötig, da alle Fächer im Team vertreten sein sollen. Die Vorbereitungen im Vorfeld des Besuchs sind umfangreich. Die vorab von der Schule angeforderten Unterlagen umfassen Daten zur Schüler- und Lehrerschaft, zur finanziellen und räumlichen Ausstattung, zu Curriculum, Schülerleistungen und Lehrer-Schüler-Relation, Entwicklungsplan, Stundenpläne, Protokolle und weitere Dokumente. Nach einem Erstbesuch der Schule durch den Registered Inspector und vorbereitende Sitzungen des Teams gibt es eine Besprechung mit Eltern, von der Lehrer ausgeschlossen sind. In der folgenden Inspektionswoche gibt es neben fortlaufenden Unterrichtsbeobachtungen Gespräche mit Schülern, Begutachtungen von Schülerarbeiten und Analyse der Testergebnisse, Einsicht in Unterrichtsvorbereitungen und in weiteres dokumentarisches Material. Die Inspektion endet mit einer mündlichen Berichterstattung unmittelbar nach Ablauf und einem schriftlichen Bericht. In beiden Fällen wird dazu das Kollegium gehört, um ihm Gelegenheit zu geben, Fakten richtig zu stellen, ohne dass Urteile verhandelbar wären oder Ergebnisse diskutiert werden können. Der Bericht enthält neben in Noten ausgedrückten Fachprofilen und einer Bewertung der Schüler sowie der Effektivität des Managements Angaben zu den „key issues for action“. Die Schule ist verpflichtet, innerhalb von 40 Tagen nach Eingang des Berichts einen Aktionsplan zu veröffentlichen und danach über eigene Maß-

nahmen zu berichten, bzw., falls sie als „school at risk“ oder „failing school“ eingestuft wird, die Überwachung dieser Maßnahmen bis hin zur Einsetzung eines vom Ministerium eingesetzten Leitungsteams hinzunehmen.

Lässt man die Erfahrungen Revue passieren, die mit dieser Art der Qualitätskontrolle gemacht wurden, so scheinen *positive Wirkungen* am ehesten von dem erwähnten Handbuch auszugehen, da Schulen dies auch unabhängig von Inspektionen der eigenen Evaluation zugrundelegen. Allenfalls ist als weitere positiver Effekt ein größerer Zusammenschluss innerhalb der Kollegien anzusehen, um sich und die Schule zu schützen, jedoch nur in dem Sinne, sich in einem möglichst guten Licht zu präsentieren – Wirkungen in Richtung einer tatsächlichen Weiterentwicklung gehen davon nicht aus. Die Liste der *Negativeffekte* ist ungleich länger:

- Da das Verfahren als starr, unpersönlich und an abstrakten Normen orientiert wahrgenommen wird, erscheint es als Kontrolle und nicht als Entwicklungs-Chance.
- Da das soziale Umfeld kaum berücksichtigt wird und Schwächen mehr Aufmerksamkeit erhalten als Stärken, eignet sich der Bericht nicht als Ausgangspunkt für die konkrete Weiterarbeit.
- Da Lehrkräfte nur als Ausführende von curricularen Vorgaben („teachers as part of a delivery system for a curriculum imposed upon them.“ Aspland 1997, S.284) und Objekte aufsichtlicher Massnahmen gesehen werden, entwickeln sie kein eigenes Interesse an Verbesserungsinitiativen.
- Die über die Schule publizierten Daten bestimmen ihre Stellung auf dem Bildungsmarkt und damit die Mittelzuweisungen. Zum Risikofaktor werden damit für die Schulen alle nicht in Daten erfassbaren Bildungsziele, ebenso aber auch ein Schülerwahlverhalten, das nicht zu bestmöglichen Leistungen führt und vor allem auch sozial benachteiligte Schüler. All dies führt zu einem schlechteren Abschneiden und wird vermieden.

5. „How Good is Our School?“ – Selbstevaluation an schottischen Schulen

Gegen das englische Inspektionssystem ist immer wieder vorgebracht worden, dass es tatsächlich nicht zu Qualitätsentwicklung führt, ja im Grunde nicht einmal zu Qualitätssicherung. Auch in anderen europäischen Ländern, etwa in Dänemark oder Schweden, spielen Inspektionsmodelle eine Rolle, die hier aber *zusammenwirken mit Verfahren der schulinternen Evaluation*, und eine ungleich stärkere Betonung von Gespräch, gegenseitiger Anerkennung, Überantwortung der Handlungsinitiativen an die Akteure selbst bedeuten. Wie so etwas aussehen kann, möchte ich abschließend am Beispiel Schottlands darstellen, wobei ich mich vor allem auf die Unterschiede zum englischen Modell konzentriere.

Abgesehen davon, dass Schulinspektionen in Schottland nur alle 6 – 8 Jahre durchgeführt werden, unterscheiden sie sich von der englischen Variante vor allem dadurch, dass in ihnen *Diskussionen mit Lehrern und Lehrerinnen, Schulleitung und Eltern* eine tragende Rolle spielen. Der abschließende Bericht ist denn auch eher ein Protokoll dieser Gespräche, zumindest enthält er „für das Schulpersonal nichts Neues oder Überraschendes“ (Posch/Altrichter 1997,

S.90). Die Schule erhält Gelegenheit zur Stellungnahme, so dass die endgültige Fassung in Verhandlungen mit der Schule entsteht. Innerhalb von vier Monaten soll die Schule auf dieser Basis einen Aktionsplan entwickeln, dessen Umsetzung ein bis anderthalb Jahre später überprüft wird.

Unabhängig davon sind schottische Schulen gehalten, eine jeweils auf drei Jahre bezogene *Schulentwicklungsplanung* vorzulegen. Sie „soll ...

... zur laufenden Überprüfung der pädagogischen Ziele anregen,
... eine Übersicht über die Leistungen der Schule bieten,
... die Schule veranlassen, Prioritäten zu setzen und zu realisieren,
... eine möglichst zweckmäßige Verwendung der ... Ressourcen sicherstellen,
... die Schule veranlassen, ihre Leistungen öffentlich sichtbar zu machen“ (ebda, S.91).

In einem internen Diskussionsprozess formuliert das Kollegium die *Ziele* seiner Arbeit. Diese sind die Grundlage, auf der die Schule in einem *internen Evaluationsverfahren* den Erfolg ihres Handelns überprüft. In einer *ersten Phase* werden sie nach Priorität geordnet und die Ressourcen zu ihrer Verwirklichung spezifiziert. In einer *zweiten Phase* werden die daraus folgenden schulischen Aktivitäten in sieben Schlüsselbereichen von der Schule selbst evaluiert: Curriculum, Schülerleistung, Lehren und Lernen, Beratung und Unterstützung, Ethos, Ressourcen, Management incl. Qualitätssicherung. Dazu werden zunehmend vorab festgelegte Indikatoren herangezogen, die vom Inspektorat auf der Grundlage von Gesprächen mit Lehrern, Schülern und Eltern entwickelt wurden („How Good is Our School?“). Diese „Leistungsindikatoren waren ursprünglich offenbar eher für die Hand der InspektorInnen gedacht... Inzwischen scheinen sie jedoch primär der Unterstützung von Reflexionsprozessen an den Schulen zu dienen.“ (ebda, S.99) Mit anderen Worten sind sie sozusagen von der Hand der Aufsicht in die der Schulen übergegangen. Sie sollen einerseits eine gewisse Vergleichbarkeit zwischen den Evaluationsergebnissen zwischen den Schulen ermöglichen, andererseits aber auch über eine Auseinandersetzung mit diesen Normen die interne Diskussion der eigenen Qualitätsstandards an den Schulen anregen. Dieser Evaluationsprozess soll nicht einfach nur die Defizite in der Verwirklichung der angestrebten Ziele aufdecken, sondern zur Überprüfung anregen, „ob es sinnvoll erscheint, das ursprüngliche Ziel weiter anzustreben oder ob neue Umstände andere Ziele erfordern.“ (ebda, S. 92) Deutlich soll werden, auf welchen Stärken die Schule aufbauen kann und wo sie sich Verbesserungen wünscht.

Phase drei bedeutet, dass auf dieser Grundlage Entscheidungen darüber getroffen werden, ...

... wo eine Veränderung bisheriger Verfahren nötig wird,
... wo Ressourcen von Bereichen geringerer Priorität zu solchen höherer Priorität umgeschichtet werden müssen,
... wo versucht werden muss, zusätzliche Ressourcen zu erhalten.

Diese drei Phasen müssen dokumentiert und in einem *10- bis 12-seitigen Bericht* der Schulaufsicht vorgelegt werden. Dabei soll zu jedem der sieben Schlüsselbereiche ein Stärken-/Schwächen-Profil vorgelegt und die Entwicklungsvorhaben der nächsten Jahre unter Angaben eines Zeitplans und der Strategien, Ressourcen, Erfolgskriterien und Evaluationsmaßnahmen beschrieben werden. Gerade die Auflage an die Schulen, einen solchen Bericht zu verfassen, ver-

deutlicht, dass hier in gewisser Weise das englische Verfahren genau umgedreht wird – Fremdevaluation wird zur Selbstevaluation. Darüber, wie weit dies den schottischen Schulen ermöglicht, tatsächlich in eine kollegiale Reflexion der eigenen Ziele einzutreten und Entwicklungsprozesse in die eigene Hand zu nehmen oder ob unter dem Druck der Legitimation vor Schulaufsicht und Öffentlichkeit dies umgebogen wird zu einem bloßen „window-dressing“, d.h. zur möglichst ansprechenden Drapierung des Schaufensters „Schulentwicklungsplan“ – darüber gibt es bislang keine verlässlichen Informationen. Aspland schreibt in Bezug auf das englische System: „Beachtlicher Erfindungsreichtum wird deutlich in der Art, wie Schulen sich um die Manipulation ihrer Leistungsdaten bemüht haben, um sich besser zu vermarkten.“ (Aspland 1999, S.290, m.Übers.RB) Es ist deutlich geworden, welcher Druck sie in diese Defensive zwingt.

6. Fazit

Leistungsvergleiche – und dies gilt sowohl für internationale Survey-Studien wie für interne Vergleiche auf Schulebene, etwa Vergleichsarbeiten und Evaluationsvorhaben – sind nur so viel wert, wie von ihnen *motivierende Impulse* ausgehen, in einen *Diskussionsprozess über die eigene Arbeit* einzutreten. Wie man dies erfolgreich verhindern kann, ist an dem einen oder anderen Beispiel deutlich geworden; welche Bedingungen dafür förderlich sind, ist dagegen nur in Umrissen sichtbar. Unverkennbar ist die Gefahr groß, dass Lehrpersonen durch Formen externer Evaluation oder die summarische Erhebung von „Lernendständen“ sich in ihrer Arbeit nicht genügend gewürdigt finden, und allein schon deshalb von solchen Ansätzen kaum Impulse für Schul- und Unterrichtsentwicklung ausgehen. Abzuwarten bleibt, ob es dem versprochenen differenzierten Blick von PISA auf unser Bildungssystem gelingt, seinen Anspruch einzulösen: Eher zu diagnostizieren als zu bewerten, und damit bei der Klärung zu helfen, ob es in unserer Gesellschaft konsensfähige Erwartungen an Schule und Unterricht gibt und welche Rahmenbedingungen vorliegen müssen, damit in diesem Sinne gute Leistungen möglich werden.

Literatur

- Aspland, R# 1997: Managing Quality in the UK Education System. In: Peter.Posch /Herbert Altrichter: Möglichkeiten und Grenzen der Qualitätsevaluation und Qualitätsentwicklung im Schulwesen. Innsbruck, Wien, S. 263-299
- Baumert, Jürgen; Rainer Lehmann u.a. 1997: TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen
- BLK (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung) 1997: Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“. Bonn
- Böttcher, Wolfgang. 1999: Für eine Allianz der Qualität. In: Die Deutsche Schule. 91, 1999, 1, S.20-37
- Brügelmann, Hans.1999: Qualität des Unterrichts und pädagogische Verantwortung: ein Selbstgespräch. In: ders. (Hg.): Was leisten unsere Schulen? Zur Qualität und Evaluation von Unterricht. Seelze-Velber, S.32-46
- Hagemeister, Volker 1999: Was wurde bei TIMSS erhoben? Über die empirische Basis einer aufregenden Studie. In: Die Deutsche Schule. 91, 1999, 2, S.160-177
- Lange, Hermann 1999: Qualitätssicherung in Schulen. In: Die Deutsche Schule. 91, 1999, 2, S.144-159

- Posch, Peter; Herbert Altrichter 1999: Evaluation und Entwicklung von Schulqualität – Dimensionen, Modelle und strategische Vorschläge. In: Dies. Möglichkeiten und Grenzen der Qualitätsevaluation und Qualitätsentwicklung im Schulwesen. Innsbruck, Wien, S. 1-155
- Preuss-Lausitz, Ulf. 1999: Demokratische Selbstvergewisserung anstelle von Black-Box-Messungen. In: Hans Brügelmann (Hg.): Was leisten unsere Schulen? Zur Qualität und Evaluation von Unterricht. Seelze-Velber, S. 54-58
- Schweitzer, Jochen 1999: Keine Angst vor PISA! Perspektiven einer neuen Bildungsoffensive. In: Die Deutsche Schule. 91, 1999, 2, S.134-143
- Steffens, Ulrich; Tino Bargel (Hg.) 1987ff: Beiträge aus dem Arbeitskreis Qualität von Schule. Wiesbaden-Konstanz
- Steffens, Ulrich; Tino Bargel 1993: Erkundungen zur Qualität von Schule. Neuwied, Kriftel, Berlin
- Steffens, Ulrich; Jürgen Markstahler (1999), Beitrag zur Tagung. In: G.Grogger, W.Specht (Hg.), Evaluation und Qualität im Bildungswesen. Problemanalyse und Lösungsansätze am Schnittpunkt von Wissenschaft und Bildungspolitik. Dokumentation eines internationalen Workshops. Blumau/Steiermark, 18.-21.2.1999, S.215-225
- Tillmann, Klaus-Jürgen 1998: „Schulqualität sichern“ – erziehungswissenschaftliche Anmerkungen zur aktuellen bildungspolitischen Diskussion. In: Hans. Brügelmann (Hg.): Was leisten unsere Schulen? Zur Qualität und Evaluation von Unterricht. Seelze-Velber
- Weinert, Franz E. 1998: Neue Unterrichtskonzepte zwischen gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (Hg.): Wissen und Werte für die Welt von morgen, München, S. 101-125

Rose Boenicke, geb. 1948, PD Dr., Lehrerin, dann wissenschaftliche Mitarbeiterin am Institut für Pädagogik, TU Darmstadt, seit 1997 am Hessischen Landesinstitut für Pädagogik, Wiesbaden;
 Anschrift: Elisabethenstr. 40, 64283 Darmstadt