

Klemm, Klaus

Large scale assessments in einem modernisierten Bildungssystem. Über die bildungspolitische Bedeutung von Schulleistungsstudien

Die Deutsche Schule 92 (2000) 3, S. 329-338



Quellenangabe/ Reference:

Klemm, Klaus: Large scale assessments in einem modernisierten Bildungssystem. Über die bildungspolitische Bedeutung von Schulleistungsstudien - In: Die Deutsche Schule 92 (2000) 3, S. 329-338 - URN: urn:nbn:de:0111-pedocs-276175 - DOI: 10.25656/01:27617

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-276175>

<https://doi.org/10.25656/01:27617>

Nutzungsbedingungen

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

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

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

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

Kontakt / Contact:

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

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Large scale assessment in einem modernisierten Bildungssystem

Über die bildungspolitische Bedeutung von Schulleistungsstudien

Die Bearbeitung des mir vorgegebenen Themas ‚Large scale assessments in einem modernisierten Bildungssystem‘ möchte ich so aufbauen, dass ich einleitend zwei Modernisierungstrends aufzeige, die unser Bildungssystem jetzt schon beeinflussen und die es nach meiner Einschätzung in Zukunft eher noch verstärkt prägen werden. Daran anschließend will ich in einem zweiten Schritt versuchen, aus diesen von mir beschriebenen Trends heraus die Aufmerksamkeit, die large scale assessments derzeit erreichen, zu verstehen. Darauf aufbauend beschäftige ich mich im dritten Teil meiner Ausführung mit dem Leistungspotential von Schulleistungsstudien für ein modernisiertes Bildungssystem. Beschließen möchte ich mein Referat mit einer vorsichtigen Warnung vor einer Überbürdung der Schulen und ihrer Lehrerschaft.

1. Globalisierung und Dezentralisierung als mächtige Trends

Bei der Annäherung an mein Thema bin ich über den Begriff ‚modernisiertes Bildungssystem‘ gestolpert. Weder verfügen wir in Deutschland über eine einigermaßen präzise, womöglich auch noch konsensfähige Vorstellung darüber, was ein modernisiertes Bildungssystem kennzeichnet, noch habe ich den Eindruck, dass es jemanden gäbe, der dem deutschen Bildungssystem eine wie immer auch definierte Modernität attestierte. Der damit verbundenen Schwierigkeit entziehe ich mich teilweise durch eine Umformulierung: Ich streiche das Partizip der Vergangenheit ‚modernisiert‘ und rede in meinem Vortrag über ‚Large scale assessments in einem Bildungssystem auf dem Wege der Modernisierung‘. Danach bleibt aber immer noch die Unbestimmtheit des Begriffs ‚Modernität‘. Beim Umgang mit dieser Unbestimmtheit will ich, nicht nur, weil es nicht dem Zeitgeist entspräche, darauf verzichten, eine klug ausgedachte und wohl begründete Definition eines ‚modernen Bildungssystems‘ zu entwickeln, um meine Überlegungen zu large scale assessments an ihr abzuarbeiten. Ausgehend von der Erkenntnis, dass Zielsetzungen von Systemen sich in ihrem Entwicklungsprozess erst ergeben, möchte ich mich darauf beschränken, zwei Entwicklungslinien herauszuarbeiten, die unser Bildungssystem in diesen Jahren prägen und die ich – ohne persönliche Wertung – in Folge ihrer Prägekraft als ‚modern‘ ansetze. Dabei verzichte ich auf die auch nur näherungsweise Gleichsetzung von ‚modern‘ und ‚fortschrittlich‘.

Die beiden Entwicklungslinien, die ich hierbei im Blick habe, sind die der Globalisierung und die der Dezentralisierung. Meine These, dass damit zwei im

hohen Maße das Schulsystem prägende Trends angesprochen sind, möchte ich im Folgenden belegen. Dann erst, aber dann schon, lasse ich mich auf large scale assessments ein.

Ich möchte mit dem *Prozess der ‚Globalisierung‘* beginnen. Aus dem vielfältigen Chor derer, die sich mit diesem Thema befassen, lassen sich zwei Feststellungen zu Merkmalen der Globalisierung herausarbeiten. Zum einen wird gesagt, das Zeitalter der Globalisierung werde von einem in seinem Ausmaß bisher nicht gekannten Austausch von Gütern, Dienstleistungen und Kapital auf den Weltmärkten, besser noch: auf einem Weltmarkt gekennzeichnet. Und weiter scheint unumstritten zu sein, dass der Austausch von Menschen, Gütern, Dienstleistungen und Kapital in Zeiten der Globalisierung von einem in seinem Tempo bisher unvorstellbarem Austausch von Wissen und vom Aufbau einer weltweiten Wissensgesellschaft, die den Wert von Wissen an dessen weltweiter Nützlichkeit bemessen wird, begleitet sein wird.

Hinsichtlich der *Folgen*, die dieser Prozess für die beteiligten Menschen hat, gehen die Einschätzungen weit auseinander. Aber auch, wenn man den Schreckensszenarien von Autoren wie *Martin und Schumann* (‚Die Globalisierungsfalle‘) nicht folgen mag und der eher vorsichtigen Analyse des Franzosen Daniel Cohen (‚Richesse du monde, pauvreté des nations‘) folgt, bleibt als Resümee die Erwartung einer Aufteilung der Erwerbstätigen in Globalisierungsgewinner und Globalisierungsverlierer. Cohen formuliert: „Die Arbeit unterliegt einem Prozess wachsender Professionalisierung, der sämtliche Handlungsträger, die sich nicht in ihn eingliedern können, an den Rand der Gesellschaft drängt“ (1998, S. 89). Er hält es für wahrscheinlich, „dass die unqualifizierten Arbeiter in den reichen Ländern zu den Verlierern der Übergangszeit gehören, die mit aller Gewalt über sie hereinbricht – eine Zeit, in der die Nachfrage nach unqualifizierter Arbeit abrupt zusammenbricht.“ (1998, S. 101) Alle Arbeitsmarktanalysen und auch die aktuell vom Nürnberger IAB vorgelegten Prognosen zum Qualifikationsbedarf in Deutschland stützen die Einschätzung, dass in Deutschland auf mittlere und lange Sicht nicht die Arbeit, wohl aber die Arbeit für gar nicht oder wenig Qualifizierte ausgeht. Dies ist der objektive Hintergrund des breiten Interesses, das der Frage nach der Qualität des deutschen Bildungswesens neuerdings entgegengebracht wird und das den Wechsel der Blickrichtung fort von Kontext und Prozess und hin zu Wirkungen von Bildung und Ausbildung gelenkt hat.

Zeitgleich mit dem Prozess der Globalisierung vollzieht sich – weniger umfassend, aber gleichwohl mächtig – in Deutschland, aber nicht nur hier, ansatzweise eine *Verlagerung von Kompetenzen* aus jeweils übergeordneten Machtzentren in kleinere dezentrale Einheiten: in Regionen, in Kommunen, in Stadtteile. Es wäre reizvoll, darüber nachzudenken, ob in dieser Tendenz zu verstärkter Dezentralisierung das Bedürfnis zum Ausdruck kommt, der Unübersichtlichkeit der globalisierten Welt mit der Vertrautheit der näheren Umgebung ein Gegengewicht entgegen zu setzen. Da dies aber den Rahmen meines Themas sprengen würde, beschränke ich mich darauf, die Tendenz zu Dezentralisierung für den Bereich der öffentlichen Verwaltung zu skizzieren. Dort hat sich (vgl. Lange 1999) ein Paradigmenwechsel angebahnt: Unter der Überschrift ‚Neue Steuerungsmodelle‘ wird seit einigen Jahren eine Abgabe von Kompetenzen von höheren und mittleren auf untere Ebenen verfolgt. Mit

der damit verbundenen Zurücknahme von Input- und Prozesssteuerung geht eine Akzentuierung der Output-Steuerung einher. Der Rückzug auf das budgetierte Bereitstellen von Ressourcen, die Vorgabe von zu erreichenden Zielen sowie die Kontrolle der Zielerreichung erfolgt aus der Einschätzung heraus, dass die einzelne Einheit die Wege der Zielerreichung aufgabenadäquater herausfinden und kostengünstiger beschreiten kann.

Im Schulbereich wird dieser Paradigmenwechsel unter der Überschrift der *'Teilautonomisierung'* verhandelt. Das funktionale Äquivalent der Zurücknahme staatlicher Regelungsmacht ist dort, im Schulbereich, die Verstärkung aller Formen der Evaluation, aus der Sicht des Staates insbesondere auch der der externen Evaluation. Sein Interesse daran, sicher zu stellen, dass vorgegebene Ziele erreicht und nicht durch institutionelle Ziele einzelner Einrichtungen verdrängt werden, begründet sich aus der öffentlichen Verantwortung des Staates für das Bildungswesen und zusätzlich aus dem Anspruch der Vergleichbarkeit: Solange das Berechtigungswesen aufrecht erhalten bleibt, müssen die Länder, schon gar in ihrer Einbindung in das föderale System, Sorge dafür tragen, dass die Wirkung schulischer Arbeit überall da, wo gleiche Zertifikate vergeben werden, vergleichbar ist.

2. Large scale assessments an der Schnittstelle von Globalisierung und Dezentralisierung

Beide hier aufgeführten Strömungen, die der Globalisierung in einer Weltökonomie und die der Dezentralisierung, der Verlagerung von Kompetenzen auf untere Ebenen also, fügen sich – obwohl einerseits global und andererseits lokal angesiedelt – im Bereich der Schulentwicklung zusammen: Qualitätssicherung als Antwort auf *weltweite Herausforderung und Vergleichbarkeitssicherung* als Reaktion auf lokal gewählte autonome Lösungswege treffen sich in ihrem Steuerungsinteresse. Vor diesem Hintergrund übernehmen large scale assessments – ich nenne sie im Folgenden auch Schulleistungsstudien – in der deutschen Schulentwicklung am Ausgang des vergangenen und am Beginn des neuen Jahrhunderts eine schon fast zentrale Funktion im schulischen Modernisierungsprozess zwischen Globalisierung und Autonomisierung. Wie hoch die Bedeutung von Schulleistungsstudien derzeit zumindest im schulpolitischen Handlungsfeld eingeschätzt wird, ist an der Serie von Studien erkenntlich, die neben der internationalen Studie zum Leseverständnis, den TIMS-Studien und PISA sowie den PISA-Erweiterungen durchgeführt wurden oder werden bzw. in Vorbereitung sind: Nach meinem unvollständigen Überblick sind dies, sieht man einmal von BIJU ab, allesamt für die jeweiligen Länder flächendeckend für alle oder einzelne Schularten angelegte Leistungsstudien in Bayern, im Saarland, die Hamburger Studie sowie Studien in Bremen, Brandenburg und Rheinland-Pfalz. Diese sicherlich nicht abschließende Auflistung macht – zumal in Verbindung mit der in einer Reihe von Ländern beabsichtigten Einführung von zentralen Abschlussprüfungen – deutlich, wie hoch die Bedeutung von large scale assessments eingeschätzt wird.

Nun ist es nicht zulässig, aus einer hohen (bildungs-)politischen Bedeutungszumessung von Schulleistungsstudien auf deren Funktionalität für die Schulentwicklung zu schließen. Dazu bedarf es systematischerer Überlegungen, die ich im Folgenden anstellen möchte. Ich werde dabei so vorgehen, dass ich da-

nachfrage, was Schulleistungsstudien in einem Schulsystem leisten können, das sich den Herausforderungen von Globalisierung und Dezentralisierung gegenüber sieht.

3. Zum Leistungspotential von large scale assessments

Meinen Versuch der Beschreibung des Leistungspotentials von large scale assessments habe ich vierfach untergliedert: Ich spreche über das Aufräumen von Mythen, über einen neuen Blick auf den Ertrag von Schule, über die Bedeutung eines zu gewinnenden Referenzrahmens für die Schulentwicklung und über das large scale assessment als Dauerveranstaltung.

3.1. Mit Mythen wurde aufgeräumt

Je mehr die Studie zum Leseverständnis, das TIMSS-Material und die nun zwei Hamburger Studien ausgebreitet und – was TIMSS angeht – durch ausdifferenzierende Studien unterfüttert werden, umso deutlicher wird, dass die Konfrontation der alltäglichen Wahrnehmung des deutschen Schulsystems mit empirischen Befunden zum Aufgeben von Mythen führt. Dies macht den Kopf frei, Perspektiven der Schulentwicklung neu zu denken. Drei dieser wackelnden Mythen möchte ich stellvertretend anführen:

- Regionen mit stark *expandierten Bildungsbeteiligungsquoten* sind, was die Schulleistungen ihrer Schüler angeht, nicht schwächer als solche mit eher niedrigen Quoten. Was immer die Bildungsexpansion der vergangenen Jahrzehnte erbracht haben mag: ein Leistungsabfall, wenn es ihn denn geben sollte, war ursächlich mit ihr nicht verbunden.
- Die unterstellte *leistungsmäßige Hierarchie* von den höheren Schulen über Mittelschulen hin zu den gleichwertigen Schulen wird durch das Leistungsprofil konkreter Schulen in Frage gestellt; sie gilt zwar für die Durchschnittswerte der einzelnen Schulen, nicht aber für die getestete Kompetenz einzelner Schüler in diesen Schularten. Die Überlappung zwischen den Testergebnissen der Schüler unterschiedlicher Schularten sind in beiden Hamburger Studien ebenso wie in der TIMS-Mittelstufenstudie beachtlich. Von einer durchgängigen Verteilung der Schüler auf Schularten entlang ihrer individuellen Leistungsfähigkeit kann keine Rede sein.
- *Gegliederte* Schulsysteme sind *integrierten* bei der Erbringung von Leistungen im kognitiven Feld nicht überlegen. Diese Feststellung gilt dann, wenn integrierte Systeme an Stelle von und nicht zusätzlich zu gegliederten Systemen implementiert wurden. Dem, was Helmke und Weinert ‚Optimalklassen‘ nennen (Helmke 1989), Klassen, in denen der Leistungszuwachs eher stark und die Spreizung der Leistungsfähigkeit innerhalb der Klasse vergleichsweise gering ist, kommen integrierte Systeme im Vergleich zu gegliederten Systemen genauso gut oder sogar stärker nahe. So resümieren Baumert und Köller unter Bezug auf BIJU und TIMSS: „Schüler im Einheitsschulsystem der ehemaligen DDR erreichten in allen vergleichbaren Untersuchungsfächern (Mathematik, Naturwissenschaften, Deutsch) tendenziell oder deutlich bessere Leistungsergebnisse als Schüler des gegliederten Systems in den alten Bundesländern. Dies war überwiegend auf Förderungserfolge bei leistungsschwächeren Schülern zurückzuführen.“ (1998, S. 13f.) Was in diesem Zusammenhang noch schwerer wiegt: Erwar-

tungswidrig führt die Konzentration eher leistungsstarker Schüler in für sie eingerichteten Schulen, in Gymnasien also, nicht zu einer besonders starken deutschen Gruppe im Bereich der Spitzenleistungen.

Mit Blick auf die Frage nach dem Beitrag, den large scale assessments im Modernisierungsprozess von Schule leisten können, zeigen die drei hier angeführten Beispiele schon jetzt: Anders als es der Mythos vom Qualitätsverfall durch Bildungsexpansion und Integration suggeriert, muss der Prozess der Bildungsexpansion, der aus der Sicht der Abnehmer des Bildungssystems unverzichtbar ist, *nicht qualitätsmindernd* wirken; die Aufteilung von Schülern auf drei Schulstände entspricht ihrem Leistungsprofil nicht; ungegliederte Systeme, die auf derartige Aufteilungen verzichten, bezahlen dies nicht mit Qualitätseinbußen.

3.2 Der Ertrag von Schule rückt in die Mitte des allgemeinen Interesses

Der vielleicht bemerkenswerteste Beitrag, den die aktuellen Schulleistungsvergleiche zum Modernisierungsprozess der Schule leisten, besteht darin, innerhalb der Qualitätsdebatte zu Schule und Unterricht eine *Akzentverlagerung* eingeleitet zu haben. Während insbesondere in der westdeutschen Tradition der Qualitätsdebatte Kontext- und Prozessvariablen im Mittelpunkt und Wirkungen eher am Rande standen, konzentriert sich der Blick jetzt sehr stark auf die Wirkung von Schule und insbesondere von Unterricht. Ablesen lässt sich diese Verschiebung der Blickrichtung nicht nur an den Zentralthemen der schulpolitischen Debatte, sondern auch an Schwerpunkten der Forschungsförderung, an thematischen Akzenten in Fachzeitschriften, an Tagungsthemen und – nicht zuletzt – am neu gewonnenen Interesse an fachdidaktischen Themen in der universitären Lehrerbildung. Klaus Tillmann hat das von mir hier angesprochene neue Interesse unlängst so charakterisiert: „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, Schulzufriedenheit, Lehrerverbundenheit etc. befasst. Alles wichtige Merkmale, gewiss: Doch die Frage nach der Güte von Fachleistungen lässt sich offensichtlich nicht tabuisieren, wenn es um Schulqualität geht. Dies zeigt die massive öffentliche Diskussion der Leistungsvergleichsstudien, die auch die reformorientierte Schulforschung ‚kalt‘ erwischt hat. Ob und welche Zusammenhänge zwischen ‚unseren‘ Prozessvariablen von Schulqualität und der fachlichen Leistung besteht, haben wir nämlich so gut wie nie untersucht...“ (Tillmann 1998, S. 11). Schulleistungsvergleiche lenken, so lässt sich resümieren, das Interesse, auch das der Forschung, auf eben den Bereich, der in einer sich globalisierenden Arbeitswelt einen enormen Bedeutungsgewinn erfährt: auf den Ertrag von Schule und auf dessen Qualität.

Mit dieser Verlagerung der Aufmerksamkeit, weg vom Kontext und vom Prozess schulischer Arbeit, hin zu ihren Wirkungen sind aber zugleich zwei *Nebenwirkungen* verbunden: die der – wie ich es nennen möchte – ‚curricularen Globalisierung‘ und die der Engführung bei den Lernzielen. Das gilt es auszuführen. Den internationalen Vergleichsstudien liegt ein internationales Curriculum zugrunde, das faktisch durch die Testaufgaben definiert wird (vgl.

Ramseier 1997, S. 57) und das – wie etwa bei der SII-TIMS-Studie – nicht in jedem Fall auf curriculare Validität überprüft wurde. In den Ländern, in denen das Abschneiden bei derartigen internationalen Studien eine hohe Bedeutung für die weitere Schulentwicklung erhält (wie ganz offensichtlich in Deutschland, weniger jedoch z.B. in Schottland), könnte die Schulpolitik insgesamt oder die Schulentwicklungsarbeit an einzelnen Schulen versucht sein, im Bestreben um bessere Ergebnisse beim „nächsten Mal“ den Unterricht und die Überprüfungsverfahren international auszurichten. Die Standards setzende Kraft internationaler Studien konfliktiert dann mit landesweit ausgerichteter Lehrplan- und mit regional autonomer Schulprogrammarbeit. Die damit beschriebene Entwicklungsmöglichkeit wird weniger für die Bereiche der Primar- und der Sekundarstufe I und auch weniger für die von TIMSS und PISA in den Blick genommenen Unterrichtsfächer gelten; wohl aber lässt sich vorstellen, dass im Gefolge internationaler Tests z.B. im Gebiet der Fremdsprachen auf dem Umweg dieser Tests eine fachdidaktische Wende eingeleitet oder zumindest verstärkt würde: weiter fort von kultureinführenden Lernzielen und stärker hin zum Lernziel der Kommunikationskompetenz. Unabhängig davon, ob dies begrüßt oder beklagt werden sollte, gilt: Die faktische Entscheidungskompetenz darüber verschiebt sich hin zu den internationalen Testkonstrukteuren; Tests entfalten hier eine beachtliche Eigendynamik.

Neben diesem Effekt der ‚curricularen Globalisierung‘ ist der der *Engführung bei den Lernzielen* nicht auszuschließen: Die öffentliche Debatte über die international vergleichenden Studien, insbesondere über TIMSS, hat den Blick nahezu ausschließlich auf Wirkungen im Bereich fachlicher Leistungen gerichtet. Sie ist damit hinter die Qualifikationsforschung zurückgefallen, die ja – ohne fachliche Kompetenzen gering zu schätzen – ein deutlich weiteres Qualifikationsverständnis herausgearbeitet hat. Auch wenn es richtig ist, dass man aus der Abwesenheit mathematischer nicht automatisch auf die Anwesenheit sozialer Kompetenzen schließen kann, gilt doch auch, dass die Qualität von Schule bei einer Engführung hin zu ausschließlich fachlicher Kompetenzvermittlung leiden und Modernität verlieren würde.

3.3 Schulentwicklung gewinnt einen zusätzlichen Referenzrahmen

Mit den Ergebnissen der vorliegenden und für die nächsten Jahre zu erwartenden Schulleistungsstudien gewinnt die Schulentwicklung einen Referenzrahmen. Ein solcher Rahmen ist sowohl für die Mikro- wie auch für die Makroperspektive bedeutsam. Für die Standortbestimmung der einzelnen Klasse und der Einzelschule, für die Mikroperspektive also, gilt, was Kempfert und Rolff mit Blick auf die Bedeutung von large scale assessments für die interne Evaluation der einzelnen Schule schreiben: „Selbsteinschätzungen sind nahezu zwangsläufig einseitig und möglicherweise zu wenig selbstkritisch. Das gilt für alle Berufe, nicht nur für Lehrer. Wenn die Selbstbeobachtungen nicht durch einen fremden Blick ergänzt werden, dann wird die eigene Wahrnehmung zu selektiv.“ (1999, S. 137) Hinsichtlich der Makroperspektive, bei der es um die Funktion, die Schulleistungsvergleichsstudien des Typs TIMSS oder PISA für ein Land übernehmen können, geht, lässt sich mit Ramseier feststellen: „Diese Vergleiche stellen das nationale Leistungsniveau in bestimmten Fächern fest und liefern damit einen Indikator für das Funktionieren des Bildungssystems bzw. für dessen Wirksamkeit. Der Vergleich mit anderen Ländern gibt dem ein-

zelen Land die Möglichkeit, seinen Stand anhand dieser 'sozialen Norm' zu bewerten." (1997, S. 51)

Ein solcher Referenzrahmen ist allerdings nur dann hilfreich, wenn er die Entstehungsbedingungen, also *Kontext- und Prozessvariablen*, einschließt. Dadurch erst werden Ansatzpunkte für Verbesserungen herausgearbeitet und der Konfrontation der Einzelschule mit 'unfairen' Vergleichsdaten vorgebeugt. Gerade das Einbeziehen von Kontext und Prozess erlaubt die Identifizierung von Gruppen vergleichbarer Schulen und den Vergleich ihres Ertrages im Feld fachlichen Lernens. Vor diesem Hintergrund war und ist ein zentraler Aspekt der Kritik, die gegenüber der Studie zum Leseverständnis und gegenüber den TIMS-Studien vorgebracht wurde, der Verweis darauf, dass die Kontext- ebenso wie die Prozessvariablen, die besondere Schwächen und Stärken erklären könnten, nicht miterhoben bzw. nicht oder nicht rechtzeitig mitgeteilt worden seien.

Der Vorwurf, hier sei black-box-Forschung betrieben worden, wurde ernst genommen. Die zweite Generation von large scale assessments in Deutschland – und nicht nur PISA – arbeitet mit einem *aufwendigen Instrumentarium*, das zu einem Datenschatz führen wird, mit dessen Hilfe ein Teil der Testergebnisse besser zu verstehen sein wird. Die Kombination von Daten zu Kontext, Prozess und Wirkung wird es zulassen, Ergebnisse der einzelnen Schulen – eingebettet in solche aller teilnehmenden Schulen – so an Schulen zurückzugeben, dass sie dort in den Dienst der Qualitätsverbesserung gestellt werden können. Dies ist, soweit ich weiß, bei TIMSS nicht geschehen, dies ist allerdings für PISA und z.B. auch im Rahmen der Leistungsstudie in Rheinland-Pfalz geplant.

Der *Rückmeldung von Testergebnissen* und eines damit verbundenen Interpretationsangebotes hinsichtlich ihres Zusammenhangs mit Kontext- und Prozessvariablen kommt bei der Arbeit mit Schulleistungsstudien eine zentrale Bedeutung zu. Sie stellt die Brücke dar zwischen den large scale assessments und der teilautonomen Schulentwicklung, zwischen landesweiten Zielvorgaben für Schulen und der Schulprogrammarbeit in der Einzelschule, zwischen interner und externer Evaluation. Nur wenn dieser Brückenschlag gelingt, kann vermieden werden, dass large scale assessments mit ihrem Ziel, zur Qualitätssicherung beizutragen, den anderen Strang der Modernisierung, der in den Schulen durch verstärkte Autonomie geprägt ist, konterkarieren.

3.4 Large scale assessments haben die Tendenz, zur Dauerveranstaltung zu entarten

Das hier angemahnte Zusammenführen von large scale assessments und Schulentwicklung wird in dem Maße schwieriger, in dem Schulleistungsstudien zu alltäglichen Ereignissen werden. So hilfreich sie – dosiert eingesetzt – für das Wegräumen von Mythen, für das Justieren der Schule auf ihr Kerngeschäft und für eine Standortbestimmung einzelner Schulen und der Schulen eines Landes insgesamt sein mögen, so kontraproduktiv werden sie im Modernisierungsprozess, sobald sie zur Dauerveranstaltung entarten. In einer Phase, in der Deutschland nach Jahren der Abstinenz in den neunziger Jahren erstmalig wieder an internationalen Studien teilgenommen hat und dies jetzt mit PISA fortsetzt und in der einzelne Länder in einzelnen Fächern auf einzelnen Altersstufen derartige Studien durchführen, kann von dauernden Tests wahrlich keine Rede sein.

Aber gleichwohl ist eine Entwicklung erkennbar, in deren Verlauf large scale assessments zu einer *regelmäßigen Übung mutieren*: Gibt es doch gute Gründe, die Aussagekraft von Schulleistungsuntersuchungen durch Längsschnittstudien mit mindestens zwei Messungen qualitativ überproportional zu steigern (vgl. Baumert/Köller 1998, S. 18). Ist es doch nachvollziehbar, dass auf die Kritik an der Konzentration der eingesetzten Tests auf die Kulturtechniken mit einer Ausweitung auf weitere Fächer, z.B. auf geistes- und sozialwissenschaftliche Kompetenzbereiche, reagiert wird (vgl. Ramseier 1997, S. 57 und Lange 1999, S. 156 f.). Kann es doch nur schwer zurückgewiesen werden, dass die Schulpolitik in einem föderalen Staat aus Gründen der Sicherung der Vergleichbarkeit ein starkes Interesse an regelmäßigen innerdeutschen Vergleichsuntersuchungen hat.

Würde man jedoch diesen Weg gehen, so würde man large scale assessments mit all ihrem Entwicklungspotential von der *Schulentwicklung* abkoppeln: Schulen, Schulaufsicht und Forschung wären überfordert, regelmäßige Vergleichsstudien aufzunehmen und produktiv zu verarbeiten. Dass aus der Fülle von Tests nicht zwangsläufig Qualität erwächst, zeigt uns das Beispiel der USA: Dort können statt steigender Schulleistungen eher Abnutzungserscheinungen bei den Testanwendern beobachtet werden (Kempfert/Rolff 1999, S. 17). Wer large scale assessments unter der Hand dazu nutzen will, eine Deutschland umfassende Form regelmäßiger zentraler Leistungsüberprüfung zu etablieren, der sollte diese Nebenwirkungen kalkulieren und zur Kenntnis nehmen, dass uns gerade die TIMS-Studie gezeigt hat, dass zentrale Tests vielleicht der Vergleichbarkeit von Zertifikaten dienen, sicher aber keinen eigenständigen Beitrag zur Sicherung von international hoher Qualität darstellen.

4. Zum Schluss die Warnung vor einer Überbürdung von Schulen und Lehrerschaft

Ich habe versucht, die Bedeutung von large scale assessments für den Modernisierungsprozess eines Bildungssystems (das meint in meiner vorab gewählten Definition, für die Sicherung hoher und vergleichbarer Qualität in einem durch Autonomie geprägten System) zu skizzieren. Ich wollte dabei zeigen, dass Schulleistungsstudien einen entwicklungsförderlichen Beitrag leisten können, wenn sie Kontext- und Prozessvariablen einbeziehen, wenn sie eine von Fairness geprägte Rückmeldungskultur entwickeln, wenn ihre Rolle im Kern auf das Gewinnen eines Referenzrahmens für die Entwicklung der teilautonomen Schule begrenzt bleibt und wenn sie dadurch Teil eines Qualitätsmanagements werden.

Wer large scale assessments die damit umrissene Funktion zuweist, darf jedoch nicht übersehen und darf auch nicht übersehen machen, dass das damit angesteuerte Ziel nur in dem Maße erreichbar ist, in dem es *von der Gesellschaft und ihrer Kultur getragen wird*. Mit drei abschließenden Hinweisen möchte ich dies konkretisieren:

- Die Gesellschaft muss ihre *Schulen in die Lage versetzen*, den an sie herangetragenen Qualitätsansprüchen gerecht zu werden. Schulleistungsstudien müssen so angelegt sein, dass sie bei den Lehrerinnen und Lehrern nicht als Belastung wahrgenommen werden, die auf die in den letzten Jahren ohnedies angewachsene Last zusätzlich aufgesattelt wird. Je mehr die Kollegien in den

Schulen aus der Rückmeldung ihrer Schulergebnisse Hilfen für die Analyse und für die Verbesserung ihrer täglichen Arbeit gewinnen und je stärker sie dabei auch personell gestützt werden, umso besser sind die Chancen dafür, dass large scale assessments zum produktiven Bestandteil eines umfassenden Prozesses der Schulentwicklung werden.

- Die Zielsetzung, die mit dem Einsatz von Schulleistungsstudien verbunden ist, nämlich die Sicherung und Steigerung der Qualität der Arbeitsergebnisse der Schulen, darf nicht einfach zu einer Angelegenheit der Schulen und nur der Schulen gemacht werden. Es muss gelingen, im unmittelbaren gesellschaftlichen Umfeld von Schule die Zielsetzung ‚Qualität‘ viel stärker als bisher zu verankern. Das damit Gemeinte hat viele Dimensionen, eine davon will ich mit einem Beispiel verdeutlichen: Auf Elternabenden in deutschen Schulen habe ich Eltern erlebt, die sich darüber beklagten, dass von Freitag auf Montag Hausaufgaben erteilt worden seien. Auf Elternabenden in französischen Schulen habe ich Eltern gehört, die die Lehrenden aufforderten, von Donnerstag nachmittags zu Freitag vormittags mehr Hausaufgaben zu stellen. Ohne dass ich für die aus meiner Sicht zu belastende französische Schule Partei ergreifen möchte, will ich mit diesem Beispiel daran erinnern, dass die Qualität, die in Schulen erreicht wird, eng zusammenhängt mit dem, was eine Gesellschaft den Schulen an Unterstützung zu geben bereit ist.

- Das *kulturelle Umfeld der Schule* muss die ihr zugewiesene Rolle im Modernisierungsprozess tragen. Den damit angesprochenen Zusammenhang möchte ich nach dem ökonomischen Einstieg in mein Thema mit einem kulturhistorischen Rückblick illustrieren. In der 1995 vorgelegten OECD-Studie zum Ausmaß von Analphabetismus bei der 16- bis 64-jährigen Bevölkerung erwies sich Polen als das schwächste, Schweden aber als das stärkste Land. Nun wissen wir aus der historischen Alphabetisierungsforschung (vgl. Block 1995 und 1996), dass Schweden das erste europäische Land war, dessen Bevölkerung bereits zu Beginn des 18. Jahrhunderts komplett literalisiert war. Gestützt auf den lutherischen Protestantismus und ein Kirchengesetz von 1686 erreichten die Ortsgeistlichen eine außerschulische Massenerziehung. Das setzte die Bevölkerung in die Lage, religiöse Schriften selbständig zu lesen. Auf der anderen Seite war die Bevölkerung Polens – das Land zerstückelt, ökonomisch wenig entwickelt, katholisch – in weiten Landstrichen selbst in der zweiten Hälfte des 19. Jahrhunderts noch wenig alphabetisiert, war in Europa Schlusslicht. Ich möchte mit diesem Hinweis auf Kontinuitäten darauf aufmerksam machen, wie abhängig Schulen davon sind, bei der Wahrnehmung ihres kulturellen Auftrags in übergeordnete gesellschaftliche Strömungen eingebettet zu sein.

Eine abschließende *Warnung vor einer Überbürdung der Schulen* und ihrer Lehrerschaft durch das Einfordern von mehr Qualität, für die die Gesellschaft insgesamt aber nicht einsteht, erscheint mir daher angebracht.

Literatur

- Baumert, Jürgen / Köller, Olaf 1998: Nationale und internationale Schulleistungsstudien. In: Pädagogik. 50, 1998, 6, S. 12-18
- Block, Rainer 1995: Der Alphabetisierungsverlauf im Preußen des 19. Jahrhunderts. Quantitative Explorationen aus bildungshistorischer Perspektive. Frankfurt a.M.

- Block, Rainer 1996: Determinanten der preußischen Alphabetisierung im 19. Jahrhundert. In: Historical Social Research. 1996, 1, S. 94-121
- Cohen, Daniel 1998: Fehldiagnose Globalisierung. Frankfurt
- Helmke, Andreas 1989: Unterrichtsqualität und Schulleistungen. In: Tillmann, Klaus-Jürgen (Hg.): Was ist eine gute Schule? Hamburg, S. 77-94
- Kempfert, Guy / Rolff, Hans-Günter 1999: Pädagogische Qualitätsentwicklung. Weinheim
- Klemm, Klaus: Steuerung der Schulentwicklung durch zentrale Leistungskontrollen. In: Rolff, Hans-Günter u.a. (Hg.): Jahrbuch der Schulentwicklung, Band 10. Weinheim, S. 271-294
- Lange, Hermann 1999: Qualitätssicherung in Schulen. In: Die Deutsche Schule. 91, 1999, 2, S. 144-159
- Martin, Hans.-Peter / Schumann, Harald 1996: Die Globalisierungsfalle – Der Angriff auf Demokratie und Wohlstand. Reinbek
- Ramseier, Erich 1997: Naturwissenschaftliche Leistungen in der Schweiz. Bern
- Tillmann, Klaus-Jürgen 1998: „Schulqualität sichern“ – Erziehungswissenschaftliche Anmerkungen zur aktuellen bildungspolitischen Diskussion – Vortrag beim Hearing der Fraktion der ‚Grünen‘ im NRW-Landtag in Düsseldorf am 5.9.1998 (Manuskript)

Klaus Klemm geb. 1942, Dr. phil, Professor für Erziehungswissenschaft (Bildungsforschung/Bildungsplanung)

Anschrift: Hektorstr. 14, 45131 Essen

E-mail: k.klemm@uni-essen.de