

Schirp, Heinz

## **Zentrale quantitative Leistungsmessungen und qualitative Schulentwicklung. Die Wirkungen von High Stakes Tests in den USA**

*Die Deutsche Schule 98 (2006) 4, S. 422-435*



Quellenangabe/ Reference:

Schirp, Heinz: Zentrale quantitative Leistungsmessungen und qualitative Schulentwicklung. Die Wirkungen von High Stakes Tests in den USA - In: Die Deutsche Schule 98 (2006) 4, S. 422-435 - URN: urn:nbn:de:0111-pedocs-273560 - DOI: 10.25656/01:27356

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-273560>

<https://doi.org/10.25656/01:27356>

### **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](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Digitalisiert

Mitglied der

  
Leibniz-Gemeinschaft

---

Heinz Schirp

## Zentrale quantitative Leistungsmessungen und qualitative Schulentwicklung

Die Wirkungen von High Stakes Tests in den USA\*

---

Lieben Sie skurrile und vielleicht sogar makabre Geschichten? Wenn ja, dann ist Ihnen der Name Edgar Allan Poe sicherlich ein Begriff. Vielleicht haben Sie sogar seine Geschichte „Der Sturz in den Mahlstrom“ gelesen. Für alle, die sich vielleicht nicht mehr so genau daran erinnern, eine knappe Zusammenfassung:

Zwei Fischer, zwei Brüder, werden an der Küste Norwegens mit ihrem Boot in einen gewaltigen Strudel, in einen furchterlichen Mahlstrom gerissen. In ihrer Angst und Panik klammern sie sich an den Planken ihres Bootes fest und müssen mit ansehen, wie ihr Boot in rasender Fahrt allmählich immer tiefer sinkt und sich immer schneller dem zerstörerischen Schlund des Mahlstroms nähert. Nach einigen Minuten der Panik setzt bei dem älteren der beiden Brüder allmählich wieder etwas klareres Denken ein. Er beginnt, genauer hinzusehen, zu beobachten und allmählich fällt ihm auf, dass nicht alle Gegenstände in gleicher Geschwindigkeit im Strudel verschwinden. Die leichten Gegenstände etwa schwimmen offensichtlich länger oben, während die schweren Boote und Wrackteile schneller nach unten gezogen werden. Er ruft seinem Bruder zu, dass es vernünftig sei, jetzt das Boot zu verlassen, sich an ein leeres Fass zu binden und auszusteigen. Sein jüngerer Bruder aber ist so von Angst und Panik gelähmt, dass er nicht auf die guten Ratschläge seines Bruders hört und das Boot nicht verlassen will. Also bindet sich der ältere Bruder alleine an ein Fass und springt über Bord in das tosende Meer. Und tatsächlich: während das schwere Fischerboot im Strudel immer schneller nach unten gezogen wird, treibt das Fass mit dem daran gebundenen Fischer länger am oberen Rand des Mahlstroms. Kurz nachdem das schwere Boot endgültig vom Strudel verschlungen wird und auf dem Meeresgrund zerschellt, beruhigt sich das Meer allmählich, der Mahlstrom löst sich auf und der an das Fass gebundene Fischer ist gerettet.

*Norbert Elias* verwendet diese Metapher vom Mahlstrom, um zweierlei damit zu thematisieren. Zum einen geht es ihm darum, die Eigendynamik und die Sogwirkungen gesellschaftlicher Entwicklungen zu verdeutlichen.

Zum zweiten sieht er in den unterschiedlichen Verhaltensweisen der beiden Brüder zwei paradigmatische Zugänge, sich mit gesellschaftlichen Entwicklungen auseinanderzusetzen. Er nennt diese beiden unterschiedlichen Zugänge *Dis-tanz* und *Engagement* (Elias 1983, S. 74ff.).

Ich möchte beide Zugänge von Norbert Elias aufnehmen und auf die Entwicklung und Effekte quantitativer, zentraler Leistungsmessungen übertragen. Am Beispiel der bestehenden Modelle zur Leistungsmessung lässt sich zum

---

\* Antrittsvorlesung an der Universität Bielefeld am 30.05.2006

einen gut aufzeigen, wie groß die Eigendynamik und die Sogwirkungen solcher Verfahren sind und welche Auswirkungen sie auf das Schul- und Bildungssystem haben können. Zum zweiten sind aber auch die beiden von Norbert Elias benutzten Leitbegriffe „Distanz“ und „Engagement“ hilfreich, um von unterschiedlichen Perspektiven aus auf das Gesamtkonzept quantitativer Leistungsmessungen zu schauen: einmal eher *kritisch*, auch *ideologiekritisch* und zum anderen aber auch *apologetisch*, mindestens aber positiv engagiert.

Hier soll vorerst der kritische Zugang im Mittelpunkt stehen. Zunächst möchte ich deshalb ein bildungspolitisches Konzept vorstellen, dass es eher ratsam erscheinen lässt, sich von bestimmten Modellen zentraler Testpraxis zu distanzieren. Dabei soll auch deutlich werden, dass wir uns – wie der Fischer im Boot – aus einer gewissen Duldungsstarre befreien müssen und genauer hinschauen sollten, wohin uns der Strudel der Leistungsmessungen möglicherweise treiben könnte. Auf dem Hintergrund dieser kritischen Folie soll dann aufgezeigt werden, unter welchen Bedingungen es durchaus *sinnvolle* Modelle und Verfahren von Leistungsmessung geben könnte, die z.B. pädagogischen, entwicklungspsychologischen und lerntheoretischen Begründungen standhalten.

Erinnern wir uns zunächst noch einmal kurz an die Verarbeitung der Ergebnisse der PISA-Studie. In der ersten Phase ging es vielen Schulforschern, Bildungspolitikern und Pädagogen wie den beiden Fischern im Mahlstrom. Es gab tatsächlich einen PISA-Schock, mit all den Wirkungen und Nebenwirkungen, die ein solcher Schock mit sich bringt. Erst eine gewisse Sprach- und Ratlosigkeit, dann eine Art bildungspolitischer Torschlusspanik nach dem Motto: „Es muss endlich etwas geschehen!“ Jetzt sind wir – wie der Fischer im Boot – in einer Phase, in der wir uns etwas ruhiger und genauer umschauchen können, was zentrale international wie national vergleichende Schulleistungstests bedeuten. Dass sie Auswirkungen – gewollte wie ungewollte – haben, ist ja unverkennbar.

Alle internationalen und nationalen Vergleichstudien haben nun mal aufgrund ihrer Vergleichs- und Ranking-Struktur etwas gemeinsam, was gerade für Schulentwicklungsprozesse von besonderer Tragweite ist. Sie produzieren in den jeweils untersuchten Bereichen immer Sieger und Verlierer. Schon die Länder, die sich im Mittelfeld einer solchen Rangtabelle befinden, geraten unter Erklärungsdruck, weil ihre Ergebnisse eben suboptimal sind, – oft sogar unabhängig davon, wie groß die jeweiligen Punktabstände der Länder in den Rankings sind. Es wird eigentlich nur noch der Tabellenplatz gesehen; die Schlagzeilen in den Zeitungen tun dann ihr Übriges, um Testergebnisse in die Nähe von Bildungskatastrophen zu rücken.

Die nächste internationale OECD Studie wird übrigens 2007/08 stattfinden. Sie wird die sozialen Kompetenzen der Schüler und Schülerinnen testen. Man muss nach den Ergebnissen der IEA – „Civic Studie“ nicht die Gabe der Prophetie besitzen, um prognostizieren zu können, dass Deutschland wohl wieder nicht zu den Siegerländern gehören wird.

Und so hetzen wir mit unseren Bildungsbemühungen den jeweiligen suboptimalen Ergebnissen immer wieder ein Stück hinterher. Eine mühselige und wenig erfolgreiche Strategie. Viele Lehrerinnen und Lehrer, Schulleiter und Schulaufsichtskollegen artikulieren denn auch ihr Unbehagen an dieser Dynamik

von Leistungstests. Sie fühlen sich z.T. regelrecht überrollt von den Anforderungen, die sich aus solchen outputorientierten Leistungstests ergeben.

TIMSS, PISA, PIRLS, IGLU, DESI, MARKUS, QUANTUM & Co sind ja nicht folgenlos geblieben. Sie haben z.B. dazu geführt, dass es in der Grundschule und in der Sek. I mit VERA und ELSE Lernstandserhebungen gibt. Um diese lehrplanvalide planen zu können, mussten neue Kernlehrpläne mit klar definierten Standards und Kompetenzerwartungen entwickelt werden. Ähnliches gilt für die zentralen Prüfungen und das Zentralabitur, das ab 2007 in NRW durchgeführt werden soll. Hier wird erkennbar, dass und wie die internationalen und nationalen Leistungsüberprüfungen eine deutliche Sogwirkung auf die Schulen ausgeübt haben.

Woher kommt nun diese starke Fixierung auf Leistungsmessungsverfahren und deren Ergebnisse? Wir verdanken dies einer neuen *Steuerungsphilosophie*, die durch mindestens drei miteinander verbundene Konzepte geprägt wird.

*Erstens* lässt sich eine deutliche Ökonomisierung der Bildungssysteme beobachten. Unter Stichworten wie *accountability*, *output-Orientierung* und *Rechenschaftslegung* sollen Schulen daran gemessen und danach bewertet werden, was sie wirklich leisten; es wird daher versucht, die Ergebnisse schulischer Arbeit, also etwa die Lernergebnisse der Schülerinnen und Schüler quantitativ zu definieren und zu messen.

*Zweitens* – und in Konsequenz aus einer solchen ökonomischen Betrachtung – ergibt sich die Notwendigkeit von Vergleichen. Im Mittelpunkt stehen dabei auf komplizierten methodischen Verfahren beruhende empirische Untersuchungen. Erst durch Vergleich objektivierbarer Leistungsergebnisse der Schülerinnen und Schüler – so glaubt man – wird es möglich, Stärken und Schwächen einzelner Bildungssysteme zu erkennen und – orientiert an den Vorbildern offensichtlich leistungsfähigerer Systeme – ggf. umzusteuern, wenn die eigenen Ergebnisse sich als suboptimal herausstellen. Zentrale Prüfungen und Testergebnisse lassen sich in einer solchen bildungsökonomischen Betrachtungsweise als eine neue Art von „Währung“ verstehen, die Auskunft gibt über die Leistungsfähigkeit einer Schule oder – bei internationalen Vergleichsstudien – über die Leistungsfähigkeit eines nationalen Schulsystems.

*Drittens* schließlich wird damit das Leitbild einer selbstständigen Schule entworfen, die als lernende Institution ihren Freiraum nutzt, um die eigene Schul- und Unterrichtsqualität weiterzuentwickeln. Solche selbstständigen Schulen benötigen – so die neue Steuerungsphilosophie – nur noch ein Mindestmaß an Vorgaben. Sie müssen sich allerdings permanent messen (lassen) etwa an den Ergebnissen, die sie bei zentralen Tests und Lernstandserhebungen erzielen. Im Zuge der entstehenden Konkurrenz werden die Schulen dann je nach Ergebnissen besser oder schlechter dastehen, mehr oder weniger attraktiv sein für Eltern, die sich überlegen, auf welche Schule sie ihr Kind schicken sollen. Eine solche Konkurrenz – so das neoliberale Credo – hebt das pädagogische Geschäft und sorgt dafür, dass die Schulen selbst an ihrer qualitativen Weiterentwicklung arbeiten.

Nun hat jeder dieser drei neuen Steuerungsstrategien ja durchaus etwas für sich und niemand wird ernsthaft etwas dagegen haben können, wenn es etwa darum geht, ...

... die *Selbstständigkeit* der Schulen und ihr pädagogisches Profil zu verbessern,  
... ihre *Leistungsfähigkeit* zu überprüfen und  
... ggf. durch Vergleiche mit erfolgreichen Schulen und Bildungssystemen die eigene Bildungslandschaft zu verbessern und damit letztlich zu verdeutlichen, dass auch hohe *Bildungsinvestitionen* sinnvoll und wirksam eingesetzt worden sind.

Ob und wie sich zentrale quantitative Leistungsmessungen mittel- und langfristig auf die Schulen auswirken können, dieser Frage möchte ich zunächst mit der angekündigten *kritischen Distanz* nachgehen. Ein Blick über den Zaun in die USA soll zeigen, welche systemischen Konsequenzen dort zu beobachten sind.

Es liegen ganz aktuelle, gut belegte empirische Untersuchungen darüber vor, wie sich zentrale Leistungsmessungen auf ein Bildungssystem insgesamt auswirken. Auch wenn wir bei uns noch weit von US-amerikanischen Testverfahren und Testkonzepten entfernt sind, ist eine kritische Aufarbeitung vielleicht doch ganz hilfreich, um frühzeitig und warnend auf Probleme aufmerksam zu machen, mit denen wir uns demnächst vielleicht auch auseinandersetzen müssen.

## **1. Wie wirken sich zentrale Schulleistungs-Tests auf das Schul- und Bildungssystem aus? Ein Blick über den Zaun in die USA**

Das amerikanische Bildungssystem setzt seit Jahrzehnten auf zentral entwickelte Standards und zentrale, standardisierte Tests, um die Leistungen der Schüler und Schülerinnen zu messen.

Die Produktion von Tests und von Materialien, die darauf vorbereiten sollen, ist in den USA für die Verlage zu einem Milliardengeschäft geworden. Im Sinne der bereits erwähnten Outputorientierung lautet die dahinter stehende bildungspolitische Philosophie: Wenn man die Schulen mit den Ergebnissen ihrer Arbeit konfrontiert, dann entsteht eine Konkurrenz- und Wettbewerbssituation, in der die Schulen selbst sich um ihre qualitative Weiterentwicklung bemühen.

1983 nach einer vernichtenden Analyse des amerikanischen Schulsystems wurden standardisierte Tests die Grundlage für bildungspolitische Reformbemühungen. Diese Leistungstests werden *High Stakes Tests* (HST) genannt, weil sie ausgeklügelte, hoch standardisierte und validierte standardorientierte Testaufgaben enthalten. *High Stakes Testing* (HST) ist ein spezifischer Reformansatz, der die Vergabe von Belohnungen und Sanktionen für Schulen an die Ergebnisse solchen Vergleichstests bindet. Sie werden staatenübergreifend eingesetzt und sollen vergleichende Daten zu den Leistungen der Schulen in den Bundesstaaten liefern. Darüber hinaus gibt es in den meisten amerikanischen Bundesstaaten noch jeweils selbst entwickelte und eingesetzte Leistungstests zu einzelnen Fächern oder am Ende bestimmter Bildungsgänge, etwa am Ende der Junior High School.

Die schulischen Leistungsergebnisse und das damit verbundene Schulranking werden veröffentlicht. Sie werden z.B. für die Schulaufsicht Grundlage für Interventionen, für die Eltern eine wichtige Entscheidungsgrundlage dafür, in welche Schule sie ihr Kind schicken wollen.

Wie hat sich nun dieser testbasierte Reformansatz bewährt? Im Jahre 2001, also 18 Jahre nach der Einführung dieses standard- und testorientierten Reformkonzeptes, kommt der damals amtierende amerikanische Erziehungsminister Roderik Paige zu der Feststellung: „Die vom ‚Department of Education and the National Centre for Educational Statistics‘ erstellten statistischen Ergebnisse zeigen, dass heute 70 % der Viertklässler, die Innenstadtschulen besuchen, im Bereich des Lesens nicht einmal über grundlegende Kompetenzen verfügen. Unsere Oberstufenschüler (Senior High School) hinken in ihren Leistungen bei internationalen Mathematik Tests hinter den Schülern fast aller Industrienationen her. Ein Drittel der Schüler und Schülerinnen, die das College besuchen, müssen Nachhilfekurse belegen, bevor sie überhaupt am regulären Unterricht teilnehmen können. Alle diese beunruhigenden Zustände bestehen nach wie vor, trotz der seit 20 Jahren eingeführten so genannten ‚Schulreform‘ und trotz ständig ansteigender Bildungsausgaben in den Kommunen, in den Bundesstaaten und auf der Bundesebene.“ (Paige 2001, p. 1)

Und welche Konsequenz zieht der Erziehungsminister aus dieser desaströsen Situation? Er kommt zu dem höchst überraschenden Schluss: „Der einzige sinnvolle Weg, die Schülerleistungen zu verbessern, liegt darin, klarere und höhere Standards festzulegen!“ (Paige 2001, p. 1).

Eine interessante Philosophie: Die Ergebnisse der Schüler sollen also dadurch verbessert werden, dass man die Maßstäbe höher setzt. Im Sport wäre das etwa so, als würde man feststellen, dass die meisten Schüler/innen beim Hochsprung nicht die auf 1 Meter 10 liegende Messlatte überspringen können. Also legt man sie auf 1,30. Ob das der richtige Weg ist, muss ernsthaft in Zweifel gezogen werden.

Im Jahre 2002 hat der jetzige Präsident George W. Bush das sog. „No Child Left Behind“ Gesetz (NCLB) „Kein Kind wird zurück gelassen“ verkündet, in dem es – mindestens programmatisch – darum geht, sich um jedes einzelne Kind zu kümmern und jedes einzelne Kind entsprechend seinen individuellen Voraussetzungen zu fördern. In diesem Konzept geht es ebenfalls im Kern darum, Standards festzulegen und mit Hilfe zentraler Tests zu überprüfen, ob die Schüler und Schülerinnen diese erreicht haben.

### *1.1 Campbells Law: Zur Aussagekraft zentraler Leistungstests (High Stakes Tests)*

Sharon Nichols (University of Texas at San Antonio) und David Berliner (Arizona State University) haben im März 2005 eine Forschungsstudie vorgelegt, die in den USA für einiges Aufsehen gesorgt hat. David Berliner gehört dabei zu den renommiertesten empirischen Erziehungswissenschaftlern Amerikas. Schon der Titel der Studie hat es in sich: „The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing“ also: „Die unvermeidbare Korrumpierung von Indikatoren und Lehrern durch zentrale Leistungstests.“

In ihrer Studie belegen Nichols and Berliner das von dem amerikanischen Sozialpsychologen Donald Campbell formulierte und nach ihm benannte Gesetz. Dieses Campbell'sche Gesetz lautet:

„Je stärker ein einzelner quantitativer sozialer Faktor dazu benutzt wird, soziale Entscheidungen zu begründen, desto stärker ist er verzerrenden Einflüssen ausgesetzt und

je mehr führt er selbst dazu, die sozialen Prozesse zu verzerren und zu verfälschen, die eigentlich untersucht und verbessert werden sollen.“ (Lyons 1975, p. 35)

Das Campbell'sche Gesetz gilt übrigens für alle sozialen Institutionen, nicht nur für Schulen und Bildungseinrichtungen.

Etwas einfacher ausgedrückt heißt das für unseren Zusammenhang: Wenn quantitative Leistungsergebnisse – etwa aus Schulvergleichen – zum alleinigen Bezugspunkt für die Beurteilung von Schulqualität gemacht werden, dann besteht die Gefahr, dass ein verzerrtes Bild von den tatsächlichen Leistungen von Schulen entsteht. Schulen unternehmen offensichtlich alles Mögliche, um ihre Ergebnisse zu „schönen“, zu mogeln und zu täuschen.

Das klingt auf den ersten Blick höchst erstaunlich. Wenn man aber bedenkt, dass es vom guten oder schlechten Abschneiden einer Schule abhängt, ob Eltern ihre Kinder weiterhin dort anmelden, ob und in welchem Maße Sanktionen der Schulaufsicht oder der Kommune drohen, ob weiterhin finanzielle Ressourcen zur Verfügung gestellt werden oder nicht, dann lässt sich diese Begründung durchaus nachvollziehen.

Nichols und Berliner belegen die negativen Auswirkungen von Tests mit einer Fülle von Beispielen aus empirischen Schulforschungsstudien, Schulberichten, Artikeln in pädagogischen Zeitschriften und Berichten von und Interviews mit Lehrern und Schulaufsichtsbeamten. Die Autoren der Studie machen insgesamt 10 schwerwiegende Faktoren aus, die aufzeigen, wie sich zentrale Testergebnisse negativ auf Lernen, Schulentwicklung und Leistungsentwicklung auswirken.

Ich konzentriere mich hier auf drei zentrale Aspekte:

- auf die Aussagekraft solcher Leistungstest,
- auf ihre Auswirkungen auf Lehrplanvorgaben, Lerninhalte und Lernprozesse,
- auf die Auswirkungen zentraler Tests auf das Lehrerhandeln.

Die Ergebnisse der Studie machen zunächst einmal nachdrücklich deutlich, wie groß der Druck der Schulen ist, bei zentralen Tests einigermaßen gut abzuschneiden, und wie groß die Gefahr ist, die Ergebnisse der eigenen Schule auch mit nicht zulässigen Mitteln zu verbessern.

Viele Schulen und Lehrer mogeln bei den Tests, indem sie z.B. unkorrekte Punktzahlen vergeben, um nicht unter den Durchschnitt ihres Schulaufsichtsbezirks zu rutschen und mit entsprechenden Sanktionen rechnen zu müssen. Fast 50 % der in einer Studie befragten Schülerinnen und Schüler geben zu, bei den Tests zu mogeln. Dabei wird vermutet, dass die Dunkelziffer deutlich höher liegt und dass einige Lehrerinnen und Lehrer dies offensichtlich dulden, gelegentlich sogar unterstützen, weil sie sich davon bessere Testergebnisse versprechen.

Schüler/-innen mit schlechten Leistungsergebnissen und mit dürftigen Leistungsprognosen werden systematisch von den Tests ausgeschlossen, um damit den gesamten Leistungsschnitt zu verbessern und im Vergleich mit anderen Schulen bessere Rangplätze zu bekommen.

Schüler/-innen mit schlechten Noten und geringen Erfolgsaussichten in bestimmten Bildungsgängen werden dahingehend beraten, andere Bildungsgänge zu wählen und die Schule zu verlassen. Damit zählen sie nicht als Schulab-

brecher und verschlechtern auch nicht mehr die Schulbilanz. Die Schulen, die sich dagegen brav an die Testvorgaben gehalten haben, stehen letztlich als die Dummen da und finden sich, weil sie eben nicht getrickst und gemogelt haben, zu schlecht und ungerecht beurteilt. Aus den Lehrer-Interviews wird erkennbar, dass einige Schulen sich deshalb auch vornehmen, beim nächsten Mal das zu tun, was die anderen Schulen ja schließlich auch machen: beim Test zu mogeln.

Damit pervertiert dieses System der quantitativen Leistungsmessung aber nun völlig. Campbells Gesetz wird klassisch bestätigt: Die Aussagekraft des gesamten Tests und der ermittelten Ergebnisse wird letztlich nachhaltig verfälscht und entwertet.

Zentral entwickelte und zentral eingesetzte Tests müssen darüber hinaus letztlich so angelegt werden, dass sie z.B. für alle Bundesstaaten „passen“, unabhängig davon, welche spezifischen Schwerpunkte, Profile, pädagogische oder curriculare Programme dort jeweils eingeführt sind. Die Testsaufgaben müssen deshalb zwangsläufig von den Lernbedingungen und Lernvorgaben „vor Ort“ abstrahieren. Sie können damit zwar überall eingesetzt werden, aber sie treffen die tatsächliche Lernwirklichkeit der Schulen nur bedingt.

Aus der Studie von Nichols und Berliner lassen sich die folgenden systemischen Effekte belegen und zusammenfassen:

- *High Stakes Tests geben kein realistisches Bild der Schülerleistungen wieder*
- *Wenn Testergebnisse mit Rankings und Sanktionen gekoppelt werden, leisten sie offensichtlich einem institutionalisierten Betrug Vorschub.*
- *Je höher der Leistungsdruck ist, bei Tests gut abzuschneiden, desto weniger sagen die Testergebnisse etwas aus über die tatsächlichen Leistungen der Schüler oder der Schule.*
- *Je ausgeklügelter und statistisch abgesicherter quantitative Leistungstests angelegt werden, desto weniger treffen sie die tatsächlichen Bedingungen unterrichtlichen Lehrens und Lernens.*

## *1.2 Zentrale Tests und ihre Auswirkungen auf Lehrplanvorgaben, Lerninhalte und Lernprozesse*

Die Studie belegt weiterhin nachdrücklich, dass durch die starke Fokussierung auf die zentralen Tests und durch den Druck, möglichst gute Ergebnisse in der eigenen Klasse, der eigenen Schule zu erzielen, das Schulcurriculum immer enger, sozusagen immer testkonformer wird. Lehrerinnen und Lehrer überbetonen dabei geradezu zwangsläufig schnell und einfach antrainierbares und abfragbares Wissen und die damit verbundenen multiple choice-Aufgaben. Die selbst vorgenommene Verengung des Curriculums führt letztlich dazu, dass Lehr- und Lernprozesse sich zunehmend an den erwarteten Testaufgaben orientieren. Komplexere Aufgabenstellungen werden eher gemieden. Sie kommen in den meistens Tests auch nicht vor, da deren Entwicklung und standardisierte Auswertung ziemlich kompliziert sind.

Mit der deutlichen Verengung des Curriculums, seiner Themen, Inhalte und Methoden geht ein zunehmendes Test-Training einher. Gelernt und geübt werden die Fähigkeiten, mit Testaufgaben möglichst clever umzugehen. Darauf verwenden Lehrer/-innen offensichtlich überproportional viel Zeit.

Als Resultat daraus kann ein Prozess beobachtet werden, der als „Teaching to the test“ längst bekannt ist. Es wird nicht getestet, ob Schülerinnen und Schüler das gelernt und verstanden haben, was gelehrt und bearbeitet worden ist, es wird vielmehr das unterrichtet, was getestet werden soll.

Dadurch, dass überproportional viel Zeit in testvorbereitende Verfahren investiert wird, werden andere Fächer und Lernbereiche in ihrer didaktischen Bedeutung und in ihrem Stellenwert im Schulcurriculum deutlich zurückgedrängt. Dieses „Narrowing the curriculum“ hat entsprechende Effekte für die Wahrnehmung von Unterricht und Schule durch die Schülerinnen und Schüler und die Rhythmisierung des Unterrichtsalltags.

Eine Studie des Center on Educational Policy bestätigt diese Einschätzung nachdrücklich. Befragt wurden 15.000 Schulbezirke. 71 % berichteten davon, dass sie angesichts der zentralen Tests und der Notwendigkeit, dabei gut abzuschneiden, Fächer wie Geschichte, Musik, Kunst radikal zusammengestrichen haben. In einigen Schulen wird nichts anderes mehr unterrichtet als Mathematik und Englisch. William Reese, Autor einer Studie über das öffentliche Schulsystem in den USA, verweist darauf, dass die Schulen ihre Lehrpläne ganz pragmatisch den Tests der staatlichen Qualitätskontrolle angepasst haben. „Sie verfahren nach dem Prinzip: Warum soll ich etwas unterrichten, das nicht getestet wird?“ (zit. nach Moll 2006, S. 39).

Solche Effekte hat in der US-Bildungsadministration offensichtlich niemand vorausgesehen, obwohl diese eigentlich seit einigen Jahren schon mit Händen zu greifen waren. Entsprechend blauäugig oder vielleicht sogar ignorant mutet die Stellungnahme des Sprechers des US-Bildungsministeriums an, wenn er darauf verweist, dass der Staat den Schulen ja kein Curriculum vorschreibe und dass diese ja „frei“ seien, ihre Lehrinhalte und Fächervorgaben zu gestalten.

Angesichts der Bedeutung von Testergebnissen ist der Freiraum der Schulen eben doch stark eingeengt. Fassen wir zusammen:

- *Zentrale Leistungstests führen tendenziell zu einer Verengung des Schulcurriculums.*
- *Lernprozesse werden weniger auf die Entwicklung von Kompetenzen und komplexere Verstehensprozesse ausgerichtet, sondern mehr auf abfragbares Test-Wissen.*
- *Unterricht entwickelt sich bezogen auf Lerninhalte und -verfahren in Richtung höherer Testkonformität.*
- *Die Verengung auf die Testanforderungen führt dazu, dass z.B. sozial-interaktive und musisch-kreative Lerninhalte und -verfahren deutlich unterrepräsentiert sind.*

### *1.3 Die Auswirkungen zentraler Tests auf das Lehrerhandeln*

„Wenn man einen guten Test hätte, wäre es sogar angemessen, auf einen solchen Test hin zu unterrichten. Lehr- und Lernprozesse sowie die dazugehörenden Überprüfungsformen hätten dann ihren angemessenen Platz. Aber es gibt einen feinen Unterschied zwischen „teaching to the test“ and „teaching the test itself“ (Nichols/Berliner 2005, p. 89). Es ist offensichtlich so, dass sowohl Lehrerinnen und Lehrer als auch die Verlage, die die Tests entwickeln,

diese Linie oft überschreiten. Das führt zum einen zu der bereits skizzierten Verfälschung und Verzerrung der Leistungs- und Ergebnisindikatoren; zum anderen aber auch dazu, dass Lehrer/-innen immer weniger kreative, didaktisch anspruchsvolle und Schüler motivierende Lernsituationen und Lernzugänge nutzen, sondern sich gezwungen sehen, den Test selbst zu üben, z.B. dadurch, dass sie die Tests der vergangenen Jahre zu Übungszwecken einsetzen.

Nichols und Berliner haben festgestellt, dass 80 % der Lehrer im Elementarbereich in North Carolina mehr als 20 % der gesamten Unterrichtszeit dazu verwenden, für die überstaatlichen High Stakes Tests zu üben und noch deutlicher: ein Drittel dieser Lehrer/-innen gibt an, mehr als 60% der Unterrichtszeit für den Test aufzuwenden, der im Staat North Carolina eingesetzt wird.

Wenn Tests und ihre Ergebnisse eine derart überdimensionierte Bedeutung bekommen, wenn sie wie ein Damoklesschwert über den Köpfen der Lehrer/-innen hängen, dann werden solche evaluativen Instrumente offensichtlich nicht als Hilfen zur Verbesserung des Unterrichts wahrgenommen, sondern eher als eine immer wiederkehrende Bedrohung.

Lehrer/-innen fühlen sich in die Rolle von Testvorbereitern gedrängt. Sie können sich weniger als es eigentlich ihrem Selbstverständnis und ihrem professionsbezogenen Aufgabenspektrum entspricht, um individuelle Lernprozesse, Lerninteressen, Begabungen und individuelle Lernwege ihrer Schüler kümmern, auch wenn sie dies möchten und für notwendig erachten. Auf Dauer besteht die Gefahr, dass gerade die didaktischen und lerntheoretischen Aspekte einer differentiellen Begabungs-, Interessens- und Verstehensförderung zunehmend in den Hintergrund geraten.

Nicht zufrieden stellende Testergebnisse erhöhen dabei den Druck, beim nächsten Mal besser abzuschneiden. Die Neigung, zu standardisierten, testvorbereitenden Materialien und Verfahren zu greifen, wird deutlich höher. Für die testaffine Vorbereitung werden dabei zunehmend selbst wieder vorgetestete und standardisierte Übungsmaterialien eingesetzt.

Das hat noch einen weiteren Nebeneffekt. Da die Testergebnisse keine oder nur sehr geringe diagnostische Qualitäten haben, bekommen Lehrer/-innen auch keine Hinweise, wie sie die Leistungsfähigkeit ihrer Schüler kontinuierlich verbessern können. Es bleibt vielmehr offen, warum die Schülerleistungen unbefriedigend sind und wo man ggf. ansetzen müsste, um sie zu verbessern. Zusammenfassend:

- *Lehrer/-innen werden in die Rolle möglichst erfolgreicher Testvorbereiter gedrängt.*
- *Gestaltungsformen und Lernarrangements, die auf differentielle Förderung angelegt sind, werden zunehmend nachrangig.*
- *Beides führt ganz offensichtlich aus Sicht vieler Lehrer/-innen zunehmend zu einem Prozess der Entprofessionalisierung, weil ihre didaktischen und pädagogischen Kompetenzen immer weniger bedeutsam werden.*

Diese aufgezeigten negativen Effekte sind systemisch mit der Realisierung zentraler quantitativer Leistungstests verbunden. Sie sind „inevitable“ wie Campbell es formuliert, also unvermeidbar. Anstatt realistische und für die Rechenschaftslegung notwendige analytische Daten zu erheben, verzerren und verfälschen sie den evaluativen Blick auf die tatsächlichen Leistungen von Schu-

len und Schülern. Weder für die betroffenen Schulen noch für die verantwortlichen Bildungsadministrationen sind solche Ergebnisse wirklich hilfreich.

Ronald Wolk, Herausgeber der renommierten Zeitschrift *Education Week* und lange Zeit ein vehementer Unterstützer standardorientierter Tests in den USA zieht eine düstere Bilanz:

„Die durch das ‚Kein Kind soll zurück bleiben‘-Gesetz vorangetriebene standardbasierte Reform verstärkt in der Tat die am wenigsten wünschenswerten Merkmale traditioneller Schulen, nämlich die Fixierung auf Tests und Testvorbereitung, die Unflexibilität und Ineffizienz bei der Nutzung von Zeit, die Überbetonung der Erfüllung von Lehrplanvorgaben, das Auswendiglernen sowie eine überzogene Standardisierung, die den Schülern Bildung nach dem Motto anbietet: ‚Eine Größe passt für alle!‘ – unabhängig davon, welche Lernunterschiede die einzelnen Schüler tatsächlich haben.“ (Wolk 2006, p. 52)

Soweit zu den wichtigsten Untersuchungsergebnissen aus der Nichols/Berliner Studie und ein kleiner Nachtrag dazu:

In den USA haben inzwischen zweiundachtzig (!) nationale Verbände für Bildung, Bürgerrechte, religiöse Belange und Bürgervereinigungen in seltener Einmütigkeit eine gemeinsame Erklärung unterzeichnet, die eine grundlegende Überarbeitung des „No Child Left Behind“-Gesetzes fordern. Wichtigste Kritikpunkte:

- *Überbetonung von standardisierten Tests;*
- *Einengung des Lehrplans und des Unterrichts auf Test-Vorbereitung;*
- *zu große Zahl an Schulen, die als zurückgeblieben ausgewiesen werden;*
- *Gebrauch von Sanktionen, die zu keiner Schulverbesserung führen;*
- *unangemessener Ausschluss von Schülern von der Schule, um Testwerte nach oben zu drücken;*
- *unangemessene finanzielle Ausstattung des Gesetzes.*

## **2. Zentrale Leistungsmessung und Schulqualität**

Nach dieser kritisch-distanzierenden Bilanzierung der Testphilosophie und der Testpraxis in den USA nun der Versuch herauszufinden, ob und unter welchen Bedingungen zentrale Leistungsmessungen in unseren bildungspolitischen Kontexten einen Beitrag zur Entwicklung von Schul- und Unterrichtsqualität leisten können.

### *2.1 Nachhaltige Lernergebnisse sind auf breite Lernangebote und vielfältige inhaltliche Lernanreize angewiesen*

Wir wissen aus empirischen Studien, dass auf kurzfristige Testleistungen zielende Lernverfahren, also etwa Lernen und Büffeln für die nächste Klassenarbeit, die nächste Klausur, den nächsten Test ... sich als wenig gehirnfreundlich, wenig erfolgreich und als nicht nachhaltig erwiesen haben. Die Vergessensrate bei solchen testorientierten Lernformen ist besonders hoch. Gute Behaltensleistungen der Schüler hängen vielmehr von dem individuell erfahrenen Sinn von Lernanstrengungen, von intrinsischen Motivationen, von Vernetzungsmöglichkeiten mit bereits bekannten Sachverhalten, von positiven Leistungsrückmeldungen, von Beziehungen zur eigenen Lebenswelt, von ihrem erkennbaren Nutzen und von individuellen Fördermöglichkeiten und Lern-

zeiten ab. Man könnte auch verkürzend sagen: „Lernqualität hängt von der individuellen Lernentwicklung der jeweiligen Kinder und Jugendlichen ab.“

Das bedeutet z.B., dass es beim Lernen eben auch „Umwege“, eigene Interessen, Beziehungen und Vernetzungen mit anderen Wissensbereichen gibt. Für erfolgreiche Lernprozesse sind dabei gerade solche Prozesse bedeutsam, in denen Schüler in interaktiven Prozessen sich mit Fragen auseinandersetzen, die sie sachlich/inhaltlich interessieren. Genau diese Entwicklung ignorieren zentrale Leistungstests. (Comer 2006, p. 1) Statt möglichst breite, inspirierende und motivierende Lernangebote zu machen, die sich die Schülerinnen und Schüler in gemeinsamer Arbeit selbsttätig erschließen können, reduzieren viele Leistungstests Lernen auf kurzfristiges, auswendig gelerntes Faktenwissen, abgefragt in multiple choice Manier.

## *2.2 Individuelle Leistungsmessungen und zentrale quantitative Leistungstests müssen voneinander getrennt werden*

Individuelle Lernentwicklung und zentrale Tests sind zwei verschiedene Paar Stiefel. Tests dürfen die Originalität von Lernentwicklungen nicht eingrenzen. Deshalb sollten zentrale quantitative Schulleistungstests in regelmäßigen Abständen anonym und nur bei einer kleinen Anzahl von Schülerinnen und Schülern durchgeführt werden. Auf diese Weise erhält die Bildungsadministration im Sinne eines notwendigen Monitoring-Wissens Hinweise auf den Leistungsstand und die Leistungsentwicklung von Schulen und Schulformen. Das wird z.B. in Finnland erfolgreich praktiziert. Dort können Schulen freiwillig an diesen anonymisierten Tests teilnehmen, müssen aber sogar dafür bezahlen. Viele Schulen nehmen diese Möglichkeit wahr; sie – und nur die jeweiligen Schulen allein – bekommen die ausgewerteten Schülerergebnisse.

Aus solchen – an kleinen anonymen Schülerpopulationen erhobenen – empirischen Leistungsstudien lassen sich Referenzwerte ermitteln und auf deren Hintergrund auch Leistungs- und Kompetenzerwartungen formulieren. Solche Referenzwerte können wichtige Orientierungsmarken setzen. Sie verdeutlichen kriteriell, fachlich und überfachlich, was Schüler und Schülerinnen etwa am Ende eines Bildungsganges in einem Fach an Kompetenzen erworben haben sollten.

Eine zweite Bedingung ist eine schulische Adjustierung, die dafür sorgt, dass die Leistungserwartungen die jeweils schulspezifischen Bedingungen wie etwa das soziale, ökonomische, interkulturelle Umfeld der Schüler, der Eltern und der Gemeinde berücksichtigen.

## *2.3 Aus Testaufgaben müssen sich Hinweise auf Lernförderung entwickeln lassen*

„Vom Wiegen wird die Kuh nicht fett.“ Ich bleibe bei diesem oft als populistisch gescholtenen Satz. Erfolgreiches Lernen geht nicht primär über Leistungsmessung, sondern über Unterstützung und Förderung!

Testaufgaben müssen so konzipiert sein, dass sich aus den Auswertungen Hinweise zur spezifischen Lernförderung entwickeln lassen. Dies stellt an die Testentwickler hohe Anforderungen. Sie können solche qualitativen Aufgaben eigentlich nur entwickeln, wenn sie dabei didaktisch-methodische Modelle einbeziehen.

## 2.4 Outputorientierung bei den Lernergebnissen muss durch Inputorientierung bei der Lehrerqualifizierung gestützt werden

Diagnostische Materialien, Umgang mit Test-Ergebnissen, datengestützte Lernentwicklung, standardorientierte Unterrichtsgestaltung ... das alles ist für viele Kolleginnen und Kollegen Neuland. Sie benötigen professionsbezogene Unterstützung. Sehr viel stärkere Investitionen in die inputs von Schulentwicklung, in die Erweiterung von Lehrerqualifikationen werden dazu benötigt. Eine entsprechende Lehrerfortbildung kann man ganz sicher nicht nur privaten Institutionen und Verlagen überlassen. Solange Lehrer/-innen nicht praktisch erfahren können, dass und wie sie Testergebnisse zur Verbesserung des Unterrichts nutzen können, solange werden sie auch Lernstandserhebungen eher als unnötige Belastung und Kontrolle nicht aber als Hilfe ansehen. Das gilt ganz besonders dort, wo bereits zarte Pflänzchen einer innovativen Lernkultur entstanden sind. In den vergangenen Jahren hat es wichtige und interessante Veränderungen in Richtung einer schülerorientierten Lern- und Schulkultur gegeben.

Wochenplanarbeit, Freiarbeit, Stationen-Lernen, Projektlernen, Öffnung von Schule, Gestaltung des Schullebens, Kooperation mit außer-schulischen Partnern in der Kommune, Selbstständiges Lernen, Lernstudios ... stehen exemplarisch für eine sich entwickelnde neue Schulkultur, die sich auch in Schulprogrammen und Schulprofilen niedergeschlagen hat. Gerade die Schulen, in denen solche Ansätze besonders intensiv entwickelt wurden, müssen befürchten, dass durch zentrale Tests und die Fokussierung auf Testergebnisse diese innovativen Ansätze wieder ins Abseits gerückt werden. Keine unbegründete Annahme wie bereits aufgezeigt. Gerade für solche Ansätze muss gezeigt werden, ob und ggf. wie sich solche innovativen Ansätze mit testbasierten und standardorientierten Leistungserwartungen vertragen.

Dazu aber ist es unbedingt erforderlich, die *Kompetenzen der Lehrerinnen und Lehrer zu erweitern*. Sie müssen z.B. lernen,

... wie man sinnvolle und tragfähige Schlüsse aus den Ergebnissen der Lernstandserhebungen zieht und welche Bedeutung dies für die Unterrichtsgestaltung hat;

... wie sich Standardorientierung und die Gestaltung individueller Lernwege miteinander verbinden lassen;

... welche Lernarrangements benötigt werden, die Kompetenzen der Schüler in unterschiedlichen Fächern und Lernbereichen zu fördern, um so einer Verengung des Curriculums entgegenzuwirken;

... wie man diagnostische Lernmaterialien einsetzen und nutzen kann;

... welche Formen der schul- und unterrichtsinternen Evaluation und der Rückmeldungen zu Schülerergebnissen geeignet sind, die Lern- und Leistungsfreude der Schüler/-innen zu fördern.

Dazu wäre nicht nur eine verstärkte Lehreraus- und -fortbildung notwendig, es müssten auch entsprechende *Schul- und Modellversuche* durchgeführt werden, um geeignete Antworten von der und in der Praxis selbst entwickeln zu lassen.

## 2.5 Wertschätzung und Kooperation statt Kontrolle und Konkurrenz

Einige Fragen zum Abschluss: Mit der empirischen Wende und dem daran sich orientierenden Paradigma der Outputorientierung sind – gewollt oder unge-

wollt – die bildungspolitischen Schlüsselbegriffe Konkurrenz, Kontrolle und Tests verbunden. Ist das aber wirklich das Grundmodell, das unsere Schulen voranbringt? Geht die Entwicklung von Unterrichts- und Schulqualität wirklich über Konkurrenz und Kontrolle. Oder ist nicht vielleicht ein besseres Modell denkbar, das sich an den Leitideen Kooperation, Wertschätzung und Unterstützung orientiert? Ist vielleicht auch eine sinnvolle Balance beider Ansätze denkbar?!

Gerade wenn man von den Siegern der internationalen Vergleichsstudien lernen will, dann sollte man sich etwa die Philosophie finnischer Bildungspolitiker zu eigen machen: Sie lässt sich auf drei einfache aber entscheidende Grundsätze reduzieren:

- *Durch Tests keinen Schüler beschämen und keine Schule an den Pranger stellen.*
- *Lehrer für differentielle Förderaufgaben und für den Umgang mit Heterogenität qualifizieren.*
- *Möglichst viele Personen und Institutionen der Kommune an der Schulentwicklung beteiligen.*

Das heißt z.B. Verzicht auf jede Form eines innerschulischen oder schulvergleichenden Rankings. Die Gefahr ist einfach zu groß, dass dies zu Effekten führt, wie sie in den USA zu beobachten sind.

Und nachdem mit einer etwas makaberen Geschichte begonnen wurde, biete ich Ihnen mit dem nachfolgenden Zitat eine eher idyllische Abschlussanmerkung an, die von Hartmut von Hentig stammt. Er formuliert die Spannung zwischen Distanz und Engagement bei der Beurteilung zentraler Prüfungen so:

„PISAs ungewollte Folgen sind ambivalent. Die deutsche Bildung könnte Glück gehabt haben, wie die beiden Hasen im Kinderlied: Sie sind beim unbekümmerten Grasens vom Jäger niedergestreckt worden, aber, nachdem sie sich aufgerappelt haben und sie sich besannen, dass sie noch ihr Leben hatten, hupften sie von dannen. Der Schreck könnte uns aufgeweckt haben – wachsam gemacht gegen die Gefahren und wach für die Chancen der Vermessung von Schulleistung. Der Schreck muss – wie bei den Hasen – der Besinnung weichen, dem Wissen, dass Bildung und Erziehung langsame Vorgänge sind. Nehmen wir uns Zeit und prüfen wir genau – auch die Prüfungen.“ (von Hentig 2003, S. 233)

## Literatur

- Archer, Jeff: The Road Less Travelled. In: Education Week: Quality counts at 10: A Decade of Standards-Based Education, Vol. 25, Issue 17, Jan. 2006, p. 34-37
- Böttcher, Wolfgang: Bildungsstandards und Evaluation im Paradigma der Outputsteuerung. In: Wolfgang Böttcher, Heinz Günter Holtappels, Michaela Brohm (Hg.): Evaluation im Bildungswesen. Eine Einführung in Grundlagen und Praxisbeispiele. Weinheim und München: Juventa 2006, S. 39-49
- Campbell, Donald T.: Assessing the Impact of Planned Social Change In: Lyons. G., a.a.O., p. 35
- Comer, James P.: Our Mission. It Takes More Than Tests to prepare the Young for Success in Life In: Education Week: Quality counts at 10: A Decade of Standards-Based Education, Vol. 25, Issue 17, Jan. 2006, p. 59-61
- Elias, Norbert: Die Fischer im Mahlstrom. In: ders.: Engagement und Distanzierung. Arbeiten zur Wissenssoziologie I, Frankfurt/Main, 1983, S. 74-120

- Hentig, Hartmut von: Die vermessene Bildung. Die ungewollten Folgen von TIMSS und PISA In: Neue Sammlung, 43, 2003, 2, S. 211-233
- Lersch, Rainer: Unterricht zwischen Standardisierung und individueller Förderung. Überlegungen zu einer neuen Lernkultur angesichts der bevorstehenden Einführung von Bildungsstandards In: Die Deutsche Schule, 98, 2006, 1, S. 28-40
- Lyons, Glenn (Ed.): Social Research and Public Policy, Hanover, NH 1975 (Dartmouth College)
- Meyer, Hilbert: Was ist guter Unterricht? Berlin 2004
- Moll, Sebastian: Lernen nach der Bush-Methode. in: FR, 23. Mai 2006, S. 39
- Mortensen, Poul Otto: Why Danes Detest Test. The Danish Approach to Curriculum Thinking – a Remnant from the Past or a Model for the Future?, The Danish Educational System  
<http://www.hltmag.co.uk/sep05/sart08.htm>  
<http://eng.uvm.dk/education/General/general.htm?menuid=1505>
- Nichols, Sharon L., Berliner, David C.: The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing, Arizona State University (EPSL). Tempe, AZ 2005
- Oser, Fritz: Ethos – die Vermenschlichung des Erfolgs. Zur Psychologie der Berufsmoral von Lehrpersonen, Opladen 1998
- Peek, Rainer, Döbelstein, Peter: Zielsetzung: Ergebnisorientierte Schul- und Unterrichtsentwicklung. Potenziale und Grenzen der nordrhein-westfälischen Lernstandserhebungen, in: Böttcher, W. Holtappels H.G., Brohm, M. (Hg.): Evaluation im Bildungswesen. Eine Einführung in Grundlagen und Praxisbeispiele, Weinheim und München 2006, S. 177-193
- Paige, Roderick R.: No Child Left Behind In: Carnegie Reporter, Vol. 1, No. 2, 2001 The Back Page; [www.carnegie.org/reporter/02/backpage](http://www.carnegie.org/reporter/02/backpage)
- Roeder, Peter Martin: TIMSS und PISA – Chancen eines neuen Anfangs in Bildungspolitik, -planung, -verwaltung und Unterricht In: Zeitschrift für Pädagogik, 49, 2003, 2, S. 180-197
- Ryan, Richard M., Sapp, Aislinn: Zum Einfluss testbasierter Reformen: High Stakes Testing (HST). Motivation und Leistung aus Sicht der Selbstbestimmungstheorie. In: Unterrichtswissenschaft, 33, 2005, 2, S. 143-159
- Wolk, Ronald A.: A Second Front, In: Education week, 25, Issue 17, Jan. 2006, p. 49-50, 52
- Wößmann, Ludger: Zentrale Prüfungen als „Währung“ des Bildungssystems: Zur Komplementarität von Schulautonomie und Zentralprüfungen. In: Vierteljahrshefte zur Wirtschaftsforschung, 72, 2003, 2, S. 220-237

## Anmerkung

Die Übersetzungen der amerikanischen Texte stammen vom Autor.

*Heinz Schirp*, geb. 1942, Prof. Dr., LRSD, Stellvertretender Direktor des Landesinstituts für Schule/QA, Soest; Honorarprofessor an der Universität Bielefeld; Arbeitsschwerpunkte: Curriculumentwicklung, moralkognitive Entwicklung, Lernkonzepte und Unterrichtsentwicklung  
 Anschrift: Hans-Holbein-Str. 37, D-58706 Menden  
 Email: [heinz.schirp@gmx.de](mailto:heinz.schirp@gmx.de)