

Bartnitzky, Horst

## Wie VERA und Verwandtes die Bildungsqualität beschädigen. Die potemkinschen Dörfer der gegenwärtigen Schulpolitik

*Die Deutsche Schule 98 (2006) 2, S. 201-213*



Quellenangabe/ Reference:

Bartnitzky, Horst: Wie VERA und Verwandtes die Bildungsqualität beschädigen. Die potemkinschen Dörfer der gegenwärtigen Schulpolitik - In: Die Deutsche Schule 98 (2006) 2, S. 201-213 - URN: urn:nbn:de:0111-pedocs-273324 - DOI: 10.25656/01:27332

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-273324>

<https://doi.org/10.25656/01:27332>

### 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

---

Horst Bartnitzky

## Wie VERA und Verwandtes die Bildungsqualität beschädigen

Die potemkinschen Dörfer der gegenwärtigen Schulpolitik

---

Die Vertreter der flächendeckenden Vergleichs- oder Orientierungsarbeiten geben sich gern als Pioniere der empirischen Wende im Schulwesen und behaupten, nur mit den Leistungstests sei Schulentwicklung möglich, so wenn die Landauer VERA-Gruppe in dieser Zeitschrift den Skeptikern entgegensetzt: „Es stellt sich die Frage, mit welchen Verfahren in solchen Fällen überhaupt erfolgreich Schulentwicklung stattfinden könnte.“ (Isaac u.a. 2006)

Die Fragestellung verblüfft durch ihre Naivität, denn die Antwort ist nur im Plural zu geben: Es sind viele Verfahren – interner wie externer, individueller wie struktureller, subjektiver wie objektiver Art, die gerade in ihrer Kombination der Komplexität von Bildungsprozessen zu entsprechen suchen. Sie wurden teilweise in der Zeit vor PISA entwickelt. Bleibt zu fragen: Bieten die Leistungstests ein Mehr an Chancen, Schulqualität und damit Bildungsqualität zu sichern und zu befördern, als die bereits vorhandenen Möglichkeiten? Die Antwort vorweg: Sie steht in der Überschrift dieses Beitrags. Dies ist im Folgenden aus schulpädagogischer Sicht zu begründen. Am Ende sollen dann einige der Alternativen genannt werden.

### 1. Die neue „Output-Steuerung“

Am Anfang stand eine politische Entscheidung: Nach den enttäuschenden Ergebnissen der PISA-Studie und den öffentlichen Reaktionen dazu beschloss die Kultusministerkonferenz unmittelbar nach Veröffentlichung der PISA-Ergebnisse im Dezember 2001 sieben Handlungsfelder. Im Handlungsfeld 5 wurde vereinbart: „Maßnahmen zur konsequenten Weiterentwicklung und Sicherung der Qualität von Unterricht und Schule auf der Grundlage von verbindlichen Standards sowie eine ergebnisorientierte Evaluation.“ (KMK 2001) Damit wollte die KMK für die Schulpolitik die „empirische Wende“ einleiten – verkürzt allerdings auf die Messung von Schülerleistungen, soweit sie denn messbar sind. Denn: „Um gezielt den Unterricht zu entwickeln und um gezielt die Kompetenzen der Schülerinnen und Schüler zu fördern, benötigt man unbedingt eine empirische Basis“, so die für die Vergleichstests VERA verantwortlichen Wissenschaftler (Isaac u.a.2006).

In diesem Zusammenhang wurden dann insbesondere drei Maßnahmen schulpolitisch durch die KMK und in den Ländern festgelegt und den Medien als Ausweis tatkräftiger Politik vorgestellt:

- *Erstens* ein Paradigmenwechsel: von der Input-Steuerung zur Output-Steuerung. Waren bisher die Maßnahmen zur Steuerung der Schulbildung insbe-

sondere auf Richtlinien und Lehrpläne, Aus- und Fortbildung bezogen, so sollten sie sich nun von den Ergebnissen her leiten lassen. Der Output, also in der Sprache der KMK: die ergebnisorientierte Evaluation, solle nun zeigen, was die Schule zu leisten imstande sei, wo ihre Stärken und ihre Schwächen liegen.

- Dazu bedürfe es *zweitens* der Festlegung der erwarteten Ergebnisse schulischen Lernens, eben des gewünschten Outputs: verbindliche nationale „Bildungsstandards“, die dann von der KMK 2004 für den Abschluss der vierjährigen Grundschule in Deutsch und Mathematik und für den mittleren Bildungsanschluss in Deutsch, Mathematik und Englisch veröffentlicht wurden (KMK 2004a). Diese Standards sollten von nun an die Unterrichtsarbeit in den sog. Kernfächern leiten. (Siehe auch: Schlömerkemper 2004)

- *Drittens* solle überprüft werden, inwieweit die Standards erreicht wurden: landesweite Leistungstests sollten dies richten. Zusätzlich wurde in Berlin ein zentrales „Institut für Qualitätsentwicklung im Bildungswesen“ (IQB) gegründet, das beispielhaft Testaufgaben entwickeln und damit die Arbeit der Länder unterstützen soll.

So also passt es zusammen: Output-Steuerung, dazu nationale Bildungsstandards für das Ende von Bildungsgängen und die regelmäßige Überprüfung durch Leistungstests.

Hieraus könnten sich Ansätze für eine Qualitätsentwicklung ergeben: Aus Sicht der einzelnen Schule können bundesweite Bildungsstandards als Bezugspunkte für die schuleigenen Lehrpläne und die Schulprogramme fungieren; die Außenevaluierung kann die eigenen vielfältigen Evaluierungen in den Unterrichtsprozessen und -ergebnissen ergänzen. Aus Sicht der staatlichen Verantwortung kann durch die Außenevaluierung ermittelt werden, welche Schulen wirksame und nachhaltige Unterstützungen brauchen, zum Beispiel Schulen in sozialen Brennpunkten, die aus Gründen der Bildungsgerechtigkeit eine andere Ausstattung an Personal und Material, auch an Bildungszeit brauchen, als andere Schulen.

Wie gesagt: Es könnten sich daraus Ansätze für eine Qualitätsentwicklung ergeben. *Leider aber lassen die real praktizierten Tests, ihre politische Vermarktung und die insgesamt einseitige und schmalspurige Effektivitätssicht dies nicht zu – zum Schaden der schulischen Bildung.*

*Zum aktuellen Bildungsbegriff:* Um einen Bezugspunkt für die nachfolgende kritische Betrachtung zu haben, will ich einige wichtige Bestimmungsstücke eines aktuellen schulbezogenen Bildungsbegriffs in Erinnerung rufen, wie sie allen aktuellen Bildungs- und Lehrplänen zu Grunde liegen und auch die fachbezogenen Bildungsstandards der Kultusministerkonferenz mitbestimmen. Sie sind mithin unabhängig von der Parteienkonstellation der jeweiligen Landesregierungen das gemeinsame didaktische Verständnis der deutschen Schulpolitik (siehe auch Bildungskommission NRW 1995, 30 ff.):

- Bildung umfasst alle Aspekte der Lebenswelt: Sprache und Kommunikation, Mathematik, Natur und Technik, Zeit und Raum, Sozialität und Politik, ästhetische Bereiche sowie Religion und Wertorientierung.

- Bildung als Prozess stärkt die Fähigkeiten, die Lebenswelt in ihren Phänomenen wahrzunehmen, Strukturen und Funktionen zu durchschauen und in ihr selber aktiv mitzuwirken, Mitverantwortung zu übernehmen und zu tragen. Des-

halb ist Unterricht auf Sinnzusammenhänge hin angelegt, die auch Fächergrenzen überschreiten, sowie auf aktive Handlungsmöglichkeiten.

- Bildung ist individuelle Bildung: Die pädagogische Ethik wie neuere Erkenntnisse der Lernforschung verweisen auf das eigenaktive Denken und Handeln, auf Sinnkonstruktion und Entwicklung eigener Lernstrategien und Lernwege.

An fünf Punkten soll im Folgenden ausgewiesen werden, wie die gegenwärtigen Leistungstests sowie der Umgang mit ihnen diesen Bestimmungstücken schulbezogener Bildung abträglich sind.

## 2. Die Verkürzung der Bildungsinhalte

Schon die bildungspolitische Debatte verkürzte sich auf die schulfachlichen Felder, in denen die internationalen Untersuchungen wie TIMSS, PISA, IGLU schwerpunktmäßig getestet haben: auf Lesekompetenz, Mathematik, Naturwissenschaften. Sicher, Lesekompetenz wird zu Recht als eine Schlüsselkompetenz angesehen, mathematisches und naturwissenschaftliches Denken als zukunftsbedeutsam. Dennoch: Ganze Lernbereiche und Lernfelder, die kürzlich noch hoch gewichtet wurden, sind quasi über Nacht in die Bedeutungslosigkeit gefallen:

Der ästhetische Bildungsbereich mit den Fächern Musik, Kunst, Gestaltung, Sport und Bewegung sind der schulpolitischen und der öffentlichen Aufmerksamkeit entzogen. Hier droht Brachland zu entstehen, in der Folge werden nachwachsende Generationen Events für Kultur halten.

Demokratisches Lernen, Friedenserziehung, Umwelterziehung, selbstständiges Lernen und anderes mehr waren vor Jahr und Tag noch wichtige überfachliche Felder, zu denen in den Ländern Fortbildungen angeboten, Materialien entwickelt und Projekte angestoßen wurden. Als seien sie unwichtig geworden, spielen sie höchstens noch eine Randrolle.

Aber auch innerhalb der durch die Standards hoch gewichteten Fächer werden durch die Tests einseitige Gewichtungen vorgenommen. Lesekompetenz ist für das Fach Deutsch der bevorzugte Überprüfungsbereich. Miteinander sprechen; szenisch spielen, Texte planen, entwerfen, überarbeiten oder Sprache untersuchen, allesamt wichtige Kompetenzbereiche des Faches Deutsch nach den KMK-Standards geraten ins Hintertreffen, weil sie nicht zum aktuellen Aufmerksamkeitsbereich Lesen gehören und wohl auch: weil es erheblich schwieriger ist, hierfür geeignete Testaufgaben zu entwickeln.

Und selbst im bevorzugten Testbereich verkürzen sich durch didaktische Vorentscheidungen und die Testbedingungen die Bildungsinhalte. *Was hier Lesekompetenz genannt wird, ist viel weniger als das, was die gegenwärtige Lesedidaktik darunter versteht.* Die Tests folgen einem eher technischen Literacy-Konzept, wie es den internationalen Studien zu Grunde liegt. Es orientiert sich pragmatisch am Gebrauch der Basiskompetenz Lesen in Situationen des Lebensalltags mit dem Ziel der Informationsverarbeitung. Deshalb spielen Tabellen, Schaubilder, Zugfahrkarten als nicht-lineare Alltagstexte sowie Sachtexte eine große Rolle. Deshalb werden auch für die Grundschule keine Texte für die Tests erarbeitet oder bearbeitet, sondern authentische Texte aus der Lebenswelt gewählt, die sich in Sprache und Inhalt für viele Kinder oft nur schwer

erschließen, was kein Problem für die unterrichtliche Auseinandersetzung mit dem Text wäre, aber ein Problem bei der isolierten Testbearbeitung ist. Hauptsache: authentisch (zur Auseinandersetzung mit den Textqualitäten siehe Bartnitzky 2005a, 2005b; Benholz u.a. 2005). Was hier getestet wird, ist also nur ein kleiner Ausschnitt aus dem, was z.B. in den KMK-Standards als Ziele des Bereichs Lesen formuliert wird: „Die Grundschule führt zum genießenden, informierenden, selektiven, interpretierenden und kritischen Lesen und legt damit eine tragfähige Grundlage für weiteres Lernen. Die Kinder erfahren, dass Lesen eine Auseinandersetzung mit der Welt ermöglicht und Vergnügen bereiten kann.“ (KMK 2004b, 10) Fünf Dimensionen machen eine so verstandene Lesekompetenz aus, wie sie führende Deutschdidaktiker wie Kaspar H. Spinner und Bettina Hurrelmann im Einzelnen entwickelten: die motivationale, die kognitive, die emotionale und die kreative Dimension, ergänzt durch die übergreifende Dimension der Lese-Kommunikation (Bartnitzky 2003). Die Tests begrenzen sich aber ausnahmslos auf die kognitive und hier auch nur auf einen Ausschnitt.

*Die Gründe für diese Begrenzungen sind offensichtlich:* Zum einen orientiert man sich an den internationalen Untersuchungen, weil schulpolitisches Ziel ist, bei Folgeuntersuchungen besser abzuschneiden. Hier ist auch ein Wettlauf der Länder entstanden. Hatte im Deutschlandvergleich bisher Bayern die Nase vorn, soll sich nun nach Verlautbarung seines Ministerpräsidenten Nordrhein-Westfalen an die Spitze setzen. Deshalb wird das gefördert, was international getestet wird. Alles andere ist nachrangig. Natürlich wirkt sich diese Prioritätensetzung auch aus auf die „strategischen Ziele“ (Hessen), die den Schulen für ihre Entwicklung vorgegeben sind, sowie auf die Fortbildungsmittel.

Zum anderen gibt es für die Ländertests nicht die Zeit und die finanziellen Mittel für Testkonstruktion, Pretest, Durchführung und Auswertung, wie dies bei den internationalen Vergleichsstudien der Fall ist. Also sind insbesondere Testaufgaben gefragt, die leicht zu erarbeiten, zu bearbeiten und auszuwerten sind.

### 3. Die falsche Mittel-Zweck-Relation

Schon im Begriff der Output-Steuerung wird ein verhängnisvoller doppelter Irrtum sichtbar. Der erste Irrtum betrifft den sog. Output. Er ist ein Begriff aus technischen Arbeitsabläufen: Man lege das angestrebte Ergebnis fest und konstruiere den Arbeitsprozess so, dass er zu diesem Ergebnis kommt. Übertragen auf Unterricht heißt das: der Unterrichtsprozess wird zum Mittel für den Zweck des Unterrichtsziels. *Dieses Bild technischer Produktion entspricht aber nicht dem Bildungsprozess, ja, es widerspricht ihm.* Die Qualität der Prozesse hat hier einen hoch bedeutsamen Stellenwert, weil diese selber bildungswirksam sind. Schülerinnen und Schüler lassen sich durch die gestaltete Lernumgebung anregen, machen durch die Verwendung von Sozialformen und Lernsituationen bedeutsame Lernerfahrungen, nicht identische, sondern individuelle. Die langfristig wirksamen Effekte auf Selbstbild, Motivation, Einstellung, bereichsspezifische Qualifikationen sind in der Regel nicht punktuell abprüfbar, aber möglicherweise in längeren Zeiträumen beobachtbar.

Ein Beispiel vom Ergebnis her: Kinder können am Ende der Jahrgangsstufe 4 unter Beweis stellen, ob sie zu einem Bild, zu einer angefangenen Geschichte

einen unterhaltsamen Erzähltext schreiben können, bei Einhaltung bestimmter Schreibnormen. Was aber an den Kindertexten nicht sichtbar ist, das sind die Qualitäten, unter denen die Kinder ihre Fähigkeiten zum schriftlichen Erzählen erworben haben: im traditionellen Aufsatzunterricht mit Bilder geschichten, mit Formalia wie Einleitung – Hauptteil – Schluss oder aber mit Hilfe von Schreibkonferenzen, mit Planungs- und Überarbeitungsstrategien. Genau hier aber unterscheidet sich, welche nachhaltig wirksamen Begründungen die Kinder für das Schreiben von Texten erworben haben, auf welche Weise sie sich stilistisch entwickelt haben, am Ende: wie tragfähig verankert die Schreibkompetenzen sind. Kurz: Welchen Wert der Unterricht für die sprachliche Bildung des Einzelnen hat. An dem Aufsatz in einem Test ist das nicht ablesbar. Dieses falsche Bild vom Output ist der eine Irrtum.

Der zweite Irrtum betrifft die Steuerung. Technische Vorgänge sind im Prinzip steuerbar, Handlungen von Menschen sind es nicht. Sie dürfen es auch nicht sein, weil dies unserer Ethik von der Würde des Menschen widerspricht. *Bildungsprozesse sind Entwicklungsprozesse junger Menschen und es sind Förderprozesse professionell Lehrender.* Beide sind nicht steuerbar, bei beiden bedarf es der inneren Zustimmung zu einer Handlung, sonst entwickelt sich kein Bildungsprozess. Dies zusammenzubringen, ist die hohe Kunst aller Lehrenden. Günstige Bedingungen dafür zu schaffen ist im Übrigen die Pflicht der Schulpolitik.

Hier also ist der beschworene Paradigmenwechsel nötig: von der behaupteten Output-Steuerung zur Entwicklungsförderung mit dem Doppelblick auf die Prozesse wie auf die Ergebnisse.

#### **4. Die Verkennung der Schülerleistung**

Mit der falschen Mittel-Zweck-Relation ist bereits das Fehlbild vom Lerner angesprochen. Schülerinnen und Schüler sind nicht Objekte der Belehrung, sie lernen nicht eins zu eins, was gelehrt, ihnen also „beigebracht“ wurde. Schülerinnen und Schüler sind Subjekte ihres Lernens, die nach ihren je individuellen Voraussetzungen, mit ihren individuellen Möglichkeiten und ihrem aktuellen Interesse Angebote zum Lernen wahrnehmen, Neues im Zusammenhang ihres bisherigen Fähigkeits- und Kenntnisstandes konstruieren. Gehirnarbeit ist individuelle Arbeit. Als guter Unterricht gilt deshalb zu Recht ein Unterricht, der genau dies unterstützt: *die Fähigkeiten zum selbstständigen Lernen mit dem Angebot verschiedener Zugriffsmöglichkeiten und Lernmethoden.* Lernkompetenz ist deshalb ein didaktischer Schlüsselbegriff. Verbunden mit der individuellen Eigenleistung jedes Lerners gilt das didaktische Interesse insbesondere dem Lernweg der Schülerinnen und Schüler, ein Weg auch mit eigenständigen Lösungsstrategien, mit scheinbaren Rückschritten, mit Übergeneralisierungen als wichtigen Etappen auf dem Lernweg. An Beispielen des Erwerbs der Schriftsprache und der Rechtschreibung oder der Lösungen mathematischer Probleme ist dies vielfach empirisch belegt.

*In den derzeitigen amtlich beauftragten Leistungstests ist von einem solchen Verständnis von Lernprozessen aber nichts zu erkennen.* Hier gilt für die Testaufgaben das Gebot der Ökonomie, wie dies auf einen entsprechenden Vorbehalt Helmke u.a. klar stellten. Um „viel diagnostisch relevante Informationen pro Zeiteinheit“ zu gewinnen, muss darauf geachtet werden, „dass die Schü-

lerinnen und Schüler möglichst wenig selbst schreiben müssen, stattdessen: Ankreuzen, Unterstreichen etc.“ (Bremerich-Vos u.a. 2005, 4). Durch diese schlichten Reaktionen lassen sich Gedankenwege und Lernstrategien, die dem Ankreuzen zu Grunde liegen, nicht ermitteln. Eine Durchsicht der Aufgaben und Auswertungsmanuale zeigt sogar, dass klug gedachte Lösungen als falsch gekennzeichnet werden, weil sie nicht der Erwartung der Testkonstrukteure entsprechen. Nur ein Beispiel (VERA 2005, 7): Im letzten VERA-Test Deutsch für Klasse 4 wird eine DB-Zugfahrkarte präsentiert, zu der es dann einige Fragen gibt. Darauf finden sich u.a. die Hinweise: von Bochum nach Frankfurt 10.11.03, von Frankfurt nach Bochum 11.11.03. Eine Frage dazu lautet: „Wo beginnt die Reise?“ Fünf Städte sind angegeben, darunter Bochum und Frankfurt. Bochum ist beförderungsmäßig die richtige Antwort, weil hier die Reise angetreten wird. Wenn Kinder bei genauem Lesen und Durchdenken herausfinden, dass mit der Fahrkarte zwei Reisen gebucht sind: zuerst die Reise, die in Bochum beginnt, dann am Tag drauf die Reise zurück, die in Frankfurt beginnt, und wenn sie deshalb Bochum und Frankfurt ankreuzen, dann haben sie zwar genau gelesen und klug gedacht, die Aufgabe aber falsch gelöst. Das Beispiel kennzeichnet ein generelles Dilemma: *Die Tests und ihre Auswertung lassen nicht zu, dass die Denkkraft der Kinder zum Vorschein kommen können.* Die Schülerinnen und Schüler bekommen keine Gelegenheit zu begründen, warum sie eine Lösung gewählt haben, weil eine solche Auswertung in der durch die Politik vorgegebenen Zeit auswertungsmethodisch nicht zu leisten ist.

Damit wird strukturell auf das verzichtet, was die Evaluierung guten Unterrichts eigentlich ausmacht: das Erhellende des selbstständigen Denkens der Lerner. Entscheidende Informationen über Schülerleistungen bleiben damit nicht erfasst, häufiger werden faktische Schülerleistungen sogar als Fehlleistungen moniert.

Kürzlich haben Benholz u.a. (2005) einige Lesetexte aus VERA Klasse 4 daraufhin analysiert, welche Schwierigkeiten sie für Kinder beinhalten, für die Deutsch die Zweitsprache ist. Das Ergebnis belegt, dass Kinder mit schwachen Deutschleistungen gar nicht erst die Chance bekommen, das zu zeigen, was sie schon können. Damit fehlt in der Rückmeldung an die Schulen jeder Hinweis auf bisherige Lernfortschritte und auf mögliche Förderschwerpunkte.

## **5. Die Verfremdungen in den Aufgabenstellungen**

Kinder und Jugendliche entwickeln dann ihre Leistungsfähigkeit besonders, wenn es um Inhalte und Ziele geht, die sie interessieren, die sie angehen, an deren Erarbeitung ihnen gelegen ist. Sachverhalte „in den Interessenhorizont/ den Fragehorizont der Kinder holen“, dies ist eine Formel, die seit der Reformbewegung zu Anfang des vorigen Jahrhunderts als eine wichtige Bedingung für erfolgreiche Lernprozesse gilt und die im Übrigen auch durch die aktuelle Hirnforschung bestätigt wird. Eine zweite kann mit dem Stichwort „Selbstwirksamkeit“ bezeichnet werden: das Ergebnis seines Lernens als Ergebnis seiner eigenen Anstrengung erkennen, seinen eigenen Lernweg mitgestalten, am Ende etwas geschafft oder sogar geschaffen zu haben, das man selber bewirkt oder mitbewirkt hat. Kurz: *Das Lernen sich zu eigen machen, ist ein wesentlicher Motor für weiteres Lernen.* Solche Erfahrungen reichhaltig zu ermöglichen und entsprechende Haltungen zu stärken, verbindet Selbst-, Sach- und Sozialkompetenz miteinander.

Die vorhandenen Leistungstests repräsentieren aber das genaue Gegenteil. Sie stellen Texte und Aufgaben vor, die für Kinder oft fremd sind. Die Aufgabenstellung und die Art der Testdurchführung lassen gar nicht erst zu, dass die Kinder und Jugendlichen sich mit der Situation, um die es geht, vertraut machen, dass sie zu einer Fragehaltung kommen. Bei Texten zum Beispiel sollen sie zu Themen, die sie wenig berühren, irgendwelche Fragen beantworten, die nicht ihre Fragen sind. Ein Beispiel: Die Kinder erhalten ohne jede thematische Einbettung oder themenführende Überschrift einen hektisch erzählten Text: Mutter und Sohn Uli hasten durch den auf dem Bahnhof stehenden Zug, um für den allein reisenden Sohn eine verlässliche Mitreisende zu finden. (VERA 2005, 4) Vielen Kindern ist die Situation aus ihrer bisherigen Lebenserfahrung nicht vertraut. Sie müssen sich in die Situation gedanklich erst einfinden, dazu hilft in der isolierten Testsituation kein unterrichtlicher Kontext. Viele Kinder würden auch zuerst Fragen stellen, z.B. warum der Junge alleine fahren soll und zu wem er nach München fährt. Diese nahe liegenden Fragen beantwortet der Text aber nicht. Ihre Arbeitsergebnisse geben sie dann ab, sie verschwinden in Aktenbergen und Rechnereingaben (siehe auch Rumpf 2004, bes. 75 ff.). Dies alles können ältere Jugendliche leichter bewältigen als jüngere Kinder, zur Kontextdistanz fähige Schülerinnen und Schüler besser als kontextabhängig denkende. Auf diese Weise werden die Ergebnisse bereits vorgeprägt: Ein bestimmter Lerntyp aus bestimmtem sozialen Milieu wird diese Testsituation besser bewältigen als andere.

## **6. Die widersinnige Überhöhung und die Schuldfrage**

Die Schulpolitik wollte von Anfang an mit den Tests die Leistungen der Schulen überprüfen und dies als Zeichen politischer Handlungsfähigkeit vermarkten. In Pressekonferenzen sind deshalb dies die entscheidenden Fragen: Sind die Leistungen besser geworden? Wo steht unser Land im Ländervergleich? Was ist zu tun? Dabei geht es immer um generalisierende Aussagen, eben um die Leistungen. Entsprechend werden die Ergebnisse in der Öffentlichkeit rezipiert: als Anzeiger des aktuellen Bildungsstands in den Schulen und des Misserfolgs oder Erfolgs der jeweiligen Schulpolitik.

Die beauftragten Institute oder, wie im Fall VERA, die Wissenschaftlergruppe der Universität Landau, sollten für Schülerleistungen entsprechend den geltenden Lehrplänen, jetzt: den vereinbarten sog. „Bildungsstandards“, die entsprechenden Tests entwickeln, jährlich neue für den flächendeckenden Einsatz. Es liegt nahe, dass mit Annahme des Auftrages gar nicht mehr möglich ist, was eigentlich zur wissenschaftlichen Seriosität gehörte: Nämlich angesichts der Beschränkungen die nur begrenzte Aussagekraft der Testergebnisse deutlich zu machen. Entsprechend wurde und wird also auch von Wissenschaftsseite hochgewertet: Die VERA-Ergebnisse, so die Wissenschaftler, zeigten den Lehrkräften „das Leistungsprofil der eigenen Klasse“ (Projekt VERA, 1), sie förderten die diagnostische Kompetenz der Lehrkräfte (Issac u.a. 2006), sie seien „lehrplanvalide“ (Bremerich-Vos u.a. 2005, 3) und anderes mehr. An anderen Stellen wird behauptet, die Testergebnisse seien für Lehrkräfte „der Blick in den Spiegel“. Daraus hätten sie Folgerungen für ihren Unterricht zu ziehen. „Zur Professionalität der Akteure an unseren Schulen muss m. E. gehören, dass sie vor Ort selbstständig Ideen für Maßnahmen und Programme entwickeln können, um die Bildungsqualität der eigenen Schule zu verbessern.“ (Leutner 2005, 12)

*Damit wird eine bestimmte Rollenverteilung festgelegt:* Die Tests externer „Experten“ zeigen die Leistungsprofile. Wo es Defizite gibt, ist es aber Aufgabe der Lehrerinnen und Lehrer, sie in absehbarer Zeit zu beheben, schließlich steht der nächste Test ein Jahr später bereits an. Die Veröffentlichung der Ergebnisse sorgt für zusätzlichen Druck auf die Schulen. Auf die Spitze getrieben wird dies derzeit in Nordrhein-Westfalen, wo regelmäßig die 25 besten vierten VERA-Klassen veröffentlicht werden sollen (25 von rund 7.000!). „Dies soll allen Grundschulen einen Anreiz geben, sich anzustrengen“, so die Schulministerin (WAZ 2005).

Die Verlage sind längst auf diesen Zug gesprungen. Dort werden Aufgabensammlungen mit dem Versprechen angedient: „Die sofort einsetzbaren Unterrichtsbausteine und Kopiervorlagen erleichtern die tägliche Arbeit und bereiten die Schüler optimal auf Prüfungen vor.“ Ein großer Schulbuchverlag bewirbt „Übungshefte für die Lernstandserhebungen, damit die Kinder angstfrei den Vergleichsarbeiten entgegensehen können“.

*Zur Spielfläche für Schulreform wird damit der testorientierte Unterricht.* Andere Bereiche wie selbstständige Schulentwicklung, Verzicht auf dauerhafte Beschämung und Beschädigung der Lerner, Veränderung der Schulstruktur und andere mehr werden oder bleiben geschlossen. Auf entsprechende Investitionen in die Bildung kann verzichtet werden, die Politik rührt z.B. das brisante Thema Auslesestruktur der deutschen Schule gar nicht erst an.

Mit dieser Überhöhung des Aussagewerts der derzeitigen Leistungstests werden sie, schulpolitisch gewollt, wissenschaftlich bestärkt, öffentlich entsprechend rezipiert, nun schulpraktisch zum Kernpunkt schulischer Diagnostik. *Messbare Effektivität wird zum Leitbild der unterrichtlichen Arbeit.* Eine gute Schule ist eine Schule, die gute Testergebnisse vorweisen kann. Auch beim sog. „fairen Vergleich“, bei dem Schulen in ähnlichen sozialen Milieus miteinander verglichen werden, ist dies nicht anders. Es werden dann eben die Schulen in mehrere Landesligen eingeteilt.

Der weitere Verlauf ist absehbar, wie ein Blick auf Erfahrungen in anderen Ländern zeigt.

Hans Brügelmann berichtet von US-amerikanischen Erfahrungen: „US-Bundesstaaten, die in den 90er Jahren die Vergabe von Abschlusszeugnissen, erfolgsabhängige Gehaltszuschüsse für LehrerInnen und die finanzielle Ausstattung der Schulen an die Ergebnisse in ihren jährlichen Tests gebunden haben, berichten immer wieder über einen Anstieg der Leistungen... Eine Studie von Amrein/Berliner (2002) zeigt aber, dass der berichtete Leistungsanstieg in der Regel nur für den engen Bereich der im jeweiligen Bundesstaat etablierten Tests galt. In unabhängigen Tests, z.B. für die Zulassung zu den Hochschulen, zeigte sich für 2/3 der 28 Staaten eine Abnahme der Testleistungen.“ (Brügelmann 2006). An anderer Stelle berichtet Brügelmann: „In einer Befragung in den USA gaben 80% der Lehrer an, dass sie zunehmend mehr Zeit für getestete Fächer und weniger Zeit für nicht getestete Fächer aufwenden.“ (Brügelmann 2005, 280). Und zu englischen Erfahrungen: „In England beklagt selbst der Leiter der Evaluationsagentur OFSTED (Office für Standards in Education) eine starke Einengung des Curriculums auf Mathematik und Sprache und als Folge eines weithin ‚flat an ordinary‘-Lernangebots Niveausenkungen

in den Fächern wie Geschichte, Geographie, Kunst und Musik.“ (Brügelmann 2005, 279).

Dies ist der „teaching-to-the-test“-Effekt: Die Leistungen in den Tests werden zwar besser, weil der Unterricht auf die speziellen Anforderungen der Tests hin ausgerichtet wird, Leistungen in den außertestlichen Bereichen aber nehmen ab, weil sie unterrichtlich nicht mehr wie zuvor gefördert werden. Potemkinische Dörfer werden gebaut.

## **7. Der schulpädagogische Befund**

Reinhard Stähling hatte sich in dieser Zeitschrift kritisch mit dem Aussagewert von Leistungstests auseinandergesetzt, festgemacht an den Vergleichsarbeiten VERA (Stähling 2005). Die Landauer VERA-Gruppe hatte in ihrer Replik dagegen den Wert besonders der Vergleichsarbeiten für die „Qualitätsentwicklung und Lehrerprofessionalisierung“ behauptet (Isaac u. a. 2006). Dem muss nach der schulpädagogischen Befundlage widersprochen werden.

Die gegenwärtig praktizierten Tests orientieren sich mit ihrer einseitigen und schmalspurigen Effektivitätssicht nicht an einem komplexen Bildungsbegriff:

- Sie verstärken das Interesse an wenigen ausgewählten Fächern, die im Fokus der internationalen Untersuchungen liegen, und hier testen sie einige Bereiche in der kognitiven Dimension mit eher einfachen Aufgabenstellungen. Überfachliche Bildungsansprüche, andere Lerndimensionen und komplexe Aufgabenstellungen bleiben ausgeklammert.
- Sie nehmen die Lerner in ihren Denkvorgängen nicht wahr, sondern ermitteln Antworten auf der Aussageoberfläche. Damit können sie keine Aussage zur Denkqualität, zu den individuellen Denkstrategien und Lösungswegen der Schülerinnen und Schüler machen. Mithin reduziert sich die Möglichkeit erheblich, Schlüsse für Fördermaßnahmen zu ziehen.
- Sie verwenden verfremdete Aufgabenstellungen. In den gewählten Lebensweltzusammenhängen kann sich immer nur ein Teil der Schüler wiederfinden. Dies wirkt sich bei vielen Schülerinnen und Schülern, insbesondere jüngeren und kontextabhängig denkenden, nachteilig auf Interesse und Leistung aus.
- Sie unterstellen für Lernprozesse eine Mittel-Zweck-Relation und verfolgen damit ein technizistisches Paradigma: Die Unterrichtsprozesse werden auf die Vermittlung von überprüfbaren Zielen reduziert. Damit wird der Blick auf nachhaltig wirksame und qualifizierende Bildungsprozesse verstellt, die Auswirkungen auf Selbst-, Sozial- und Sachkompetenzen außerhalb von Papier-Bleistift-Evaluierungen haben. Dieses Verständnis wird insbesondere in den Kommentaren der VERA-Wissenschaftsgruppe und in der politischen Vermarktung deutlich.

Damit stehen die gegenwärtig praktizierten Leistungstests, ihre Vermarktung und ihre entsprechende öffentliche Rezeption in Widerspruch zum heutigen schulbezogenen Bildungsverständnis, das auch allen Richtlinien und Lehrplänen sowie den KMK-Standards zu Grunde liegt.

Sowohl die Testautoren wie ihre Auftraggeber, die Schulpolitik der Länder, geben generalisierend vor, mit den Testergebnissen die Leistungen der Schülerinnen und Schüler und damit der Schulen in den entsprechenden Lernbereichen zu spiegeln. Damit wird die begrenzte Aussagekraft der Tests zu einzel-

nen fachlichen Leistungen fälschlich in eine Allaussage über Qualität von Unterricht und Schule insgesamt umgemünzt. Die Politik nutzt die Ergebnisse für grundlegende Eingriffe in die Struktur der Schulen, siehe die Absichten zum öffentlichen Ranking, zur Belobigung der „25 Besten“ in Nordrhein-Westfalen, verbunden mit der freien Schulwahl auch im Grundschulbereich. Damit werden neoliberale Vorstellungen aus der Wirtschaft unzulässig auf den Bildungsbereich übertragen. Folge wird sein, dass die Lehrkräfte die Testaufgaben zum Modell für Unterricht nehmen, um die Ergebnisse beim nächsten Test zu steigern. Entsprechende Erfahrungen aus anderen Ländern liegen vor. Damit werden die Begrenztheiten der derzeitigen Leistungstests (siehe oben) zum Maß für Unterricht.

Am Ende definieren die Bedingungen für jährliche Leistungstests, was und wie in der Schule gelernt werden soll, nicht aber der auf Bildung bezogene Diskurs aus Schulpädagogik, Fachdidaktik und Schulpolitik.

## 8. Was sind die Alternativen?

Die Alternative ist *nicht die Abwendung von der „empirischen Wende“* oder gar von wissenschaftlicher kritisch-konstruktiver Entwicklungsbegleitung. Die Alternative ist vielmehr, die Empirie und die Methoden an dem zu orientieren, was eine gute Schule ausmacht. Hierzu gibt es bereits Evaluierungsformen externer wie interner Art in Fülle. Einige Beispiele:

### 8.1 Zur externen Evaluation

Dazu bedarf es anspruchsvollerer Tests, die auch höhere Lernformen in den Blick nehmen. Georg Lind verweist als Beispiel auf den Moralischen Urteils-Test (MUT), der an der Universität Konstanz entwickelt wurde (Lind 2005, 27). Allerdings können solche Tests nicht Jahr für Jahr neu herausgegeben und ausgewertet werden, sie brauchen Entwicklungszeit wie anspruchsvolle Arbeiten zumeist.

Lehrerinnen und Lehrer greifen selbst auf einfache standardisierte Tests zurück, die nicht nur nach der Oberfläche ‚falsch oder richtig?‘ bewerten, sondern die Aufschluss über den Entwicklungsstand des einzelnen Kindes geben und zugleich über den Stand der eigenen Klasse zu anderen Grundschulklassen orientieren wie die Hamburger Schreibprobe (HSP 1-9), die zur Rechtschreibentwicklung für alle Jahrgangsstufen vorliegt (May).

Hans Brügelmann führte eine Leseuntersuchung mit zig-tausend Kindern in Siegen und Umgebung an Hand des Stolperwörter-Tests (LUST) durch, einem einfachen und alltagstauglichen Test übrigens für alle Grundschulklassen, den der Berliner Schulleiter und Fortbildner Wilfried Metze entwickelt hat und von Lehrkräften über das Internet bezogen werden kann. Damit werden individuelle Entwicklungen der Kinder im Laufe der Grundschulzeit deutlich (Brügelmann 2003, Metze 2004).

Indem Lehrkräfte und Schulkollegien zu solchen Tests greifen, können die Testergebnisse angemessen in die Gesamtevaluierung integriert werden und haben dann ihren immer nur ergänzenden Stellenwert als „Warnlampe“, die genauere Beobachtung anregt, nicht aber ersetzt.

Das Modell der Schulinspektion, das derzeit in einigen Ländern entwickelt und erprobt wird, kann ein für die Evaluierung und Förderung der Bildungsqualität hilfreiches Instrument sein, wenn es auf kritisch-konstruktive Beobachtung und Kommunikation mit den Lehrenden setzt und wenn es das komplexe Faktorengefüge im Blick hat, das eine „gute Schule“ ausmacht. Nur wenn es die Bestandsaufnahme in Bezug auf die Effektivität in die Bestandsaufnahme der Schulqualität integriert, wird es die Bildungsqualität erhellen und angemessene Ansatzpunkte ihrer Weiterentwicklung finden können. (Siehe zum Forschungsstand: Brügelmann 2005, 283 ff., Czerwenka 2005).

Ein Blick nach Schweden zeigt, wie dort der „Blick von außen“ so genutzt wird, dass er sich in die Unterrichtsentwicklungen der jeweiligen Schule integriert: Jeweils am Ende der Klasse 5 werden nationale Fachprüfungen durchgeführt. „Die Prüfungen“, so die Durchführungsanweisung, „sollten nicht als Zusatz zur übrigen Arbeit in der Schule angesehen werden, sondern vielmehr als Material, welche gewisse Teile der täglichen Arbeit ersetzen.“ Vermieden wird die deutsche Klausurmanie. In den schwedischen Anleitungen heißt es zum Beispiel: Das Material enthält „sowohl Aufgaben, welche die Schüler individuell lösen sollen als auch Aufgaben, die sie zusammen lösen sollen ... Es ist vorteilhaft, den Prüfungscharakter herunterzuspielen ... Die Lehrkräfte können ... beschließen, mehr oder weniger Zeit zu veranschlagen. Es ist nicht beabsichtigt, die Schüler unter Zeitdruck arbeiten zu lassen, sondern ihnen die Möglichkeit zu geben, mit den Aufgaben fertig zu werden.“ Zudem enthalten die Materialien immer auch Bögen zur Selbstbeurteilung oder -einschätzung durch die Schülerinnen und Schüler selbst: „Eine derartige Selbstbeurteilung bezweckt, das Bewusstsein des Schülers für das eigene Lernen zu schärfen und seine Verantwortung dafür zu unterstreichen, welches eines der übergreifenden Ziele des Lehrplans ist.“ (Skolverket 2003/2004) Geschadet hat diese Art der externen Evaluierung den Leistungsergebnissen schwedischer Schülerinnen und Schüler nicht, im Gegenteil.

## 8.2 Zur internen Evaluation

Noch einmal das Beispiel Nordrhein-Westfalen: Hier wurden bereits vor VERA für Deutsch und Mathematik Aufgabenbeispiele entwickelt, die zugleich Modelle für qualitativ vollen Unterricht sind (z.B. Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen 2002). Die Lehrkräfte der Parallelklassen wählten hieraus für ihre Klassen aus, erörterten die Unterrichtsprozesse, die Lernentwicklungen und die Lernergebnisse im Lehrerteam, auch mit Vergleich der Parallelklassen. Diese Arbeit war für zwei bis drei Jahre verpflichtend für alle Schulen und trug zur schulinternen Verständigung über Unterricht und Leistung bei. Mit der ersten VERA-Runde 2004 verschwanden die Aufgabenbeispiele, die entsprechenden Konferenzbeschlüsse, Fortbildungen u.a. in der Schublade. Dieses Beispiel zeigt zweierlei: Es gibt Alternativen auf der Schulebene zur Qualitätsentwicklung, viele weitere sind der Literatur zur Schulentwicklung zu entnehmen. Das Beispiel zeigt aber auch, wie solche entwicklungsbezogenen Ansätze mit der politischen Prioritätenentscheidung für die sog. Output-Steuerung abrupt an Bedeutung verloren und nicht mehr unterstützt wurden.

Der Grundschulverband entfaltet in seinem Projekt „Pädagogische Leistungskultur“ die wichtigsten Aspekte einer wirksamen, weil ermutigenden Leis-

tungsförderung, in die auch die Kinder dialogisch einbezogen sind: Leistungen der Kinder wahrnehmen – Leistungen würdigen – Lernwege öffnen – Kinder individuell fördern (Bartnitzky, Speck-Hamdan 2004). Derzeit veröffentlicht der Grundschulverband auf dieser Grundlage Materialien für alle Fächer und Schuljahre der Grundschulzeit, mit deren Hilfe Lernentwicklungen und Lernstände der Kinder festgestellt, Lerngespräche mit Kindern geführt, individuelle weitere Lernwege geöffnet und Lernwege von den Lernern selbst, also den Kindern, reflektiert werden. Im Sommer 2005 erschienen die Materialien für die Klassen 1 und 2 für Deutsch, Mathematik und Sachunterricht, im Sommer 2006 werden sie für die Klassen 3 und 4 erscheinen, Anfang 2007 für die übrigen Fächer und für überfachliche Bildungsziele (Bartnitzky u.a. 2005/2006).

Am Anfang dieses Beitrags wurde die ratlose Frage der VERA-Wissenschaftler zitiert, bezogen auf „Lehrkräfte, die VERA nicht sinnvoll nutzen“: „Mit welchen Verfahren (könnte) in solchen Fällen überhaupt erfolgreich Schulentwicklung stattfinden?“ (Isaac u.a. 2006) Ist es die schiere Unkenntnis über alternative Schulentwicklungs-Strategien sowie über Beispiele wie oben angeführt oder ist es deren Geringschätzung? Das eine wäre so erschreckend wie das andere. Das Vergessenstempo in manchen Schulministerien ebenso.

„Empirische Wende“ ja. Aber die Empirie muss die Komplexität der Faktoren beachten, die Bildungsprozesse und -ergebnisse bestimmen. Seriöse Effektivitätsforschung muss deshalb auch in aller Bescheidenheit den eigenen Stellenwert verdeutlichen und an die Politik vermitteln. Ob dann die Schulpolitik noch so interessiert an ihr ist, steht allerdings auf einem anderen Blatt.

## Literatur

- Bartnitzky, Horst 2003: Lesekompetenz – was ist das und wie fördert man sie? In: Grundschulverband aktuell, 84, 3 ff. (wieder abgedruckt in: Grundschulverband 2006)
- Bartnitzky, Horst 2005a: VERA Deutsch 2004: Ungeeignet und bildungsfern. In: Grundschule aktuell, 89, 10ff.
- Bartnitzky, Horst 2005b: „Schimpansenkinder müssen laufen lernen“ – Lesetest in Bayern. In: Grundschule aktuell 92, 25 ff.
- Bartnitzky, Horst; Angelika Speck-Hamdan (Hg.) 2004: Leistungen der Kinder wahrnehmen – würdigen – fördern. Frankfurt a.M.: Grundschulverband
- Bartnitzky, Horst; Hans Brügelmann, Ulrich Hecker, Gudrun Schönknecht, 2005: Pädagogische Leistungskultur: Materialien für Klasse 1 und 2. Frankfurt: Grundschulverband
- Bartnitzky, Horst; Hans Brügelmann, Ulrich Hecker, Gudrun Schönknecht 2006: Pädagogische Leistungskultur: Materialien für Klasse 3 und 4. Frankfurt: Grundschulverband (in Vorbereitung)
- Benholz, Claudia; Eva Lipkowski, Charitini Iordanidou 2005: Bedingungen des Textverstehens – Stolpersteine und Fördermöglichkeiten. In: Bartnitzky, Horst; Angelika Speck-Hamdan: Deutsch als Zweitsprache lernen. Frankfurt: Grundschulverband, 242 ff.
- Bildungskommission NRW 1995: Zukunft der Bildung – Schule der Zukunft. Neuwied: Luchterhand
- Bremerich-Vos, Albert; Jana Gross Ophoff, Andreas Helmke, Ingmar Hosenfeld, Sonja Wagner 2005: Stellungnahme zur Kritik an VERA in ‚Grundschule aktuell‘ 89. In: Grundschule aktuell, 90, 3ff.
- Brügelmann, Hans 2003: Leseleistungen und der „Karawanen-Effekt“. In: Grundschulverband aktuell, 84, 19 ff. (wieder abgedruckt in: Grundschulverband 2006)

- Brügelmann, Hans 2005: Schule verstehen und gestalten. Konstanz: Libelle
- Brügelmann, Hans 2006: Bildungsstandards und zentrale Kompetenztests: Ansprüche, Probleme, Perspektiven. In: Recht und Bildung, 1, im Druck
- Czerwenka, Kurt 2005: Schulentwicklung und Schulprofil. In: Wolfgang Einsiedler, u.a. (Hg.): Handbuch Grundschulpädagogik und Grundschuldidaktik. Bad Heilbrunn: Klinkhardt, S. 102-115
- Grundschulverband 2006: Lesekompetenz. Ein Lese- und Arbeitsbuch. Frankfurt a. M.: Grundschulverband
- Isaac, Kevin; Ann Christin Halt, Ingmar Hosenfeld, Andreas Helmke, Jana Groß Ophoff 2006: VERA: Qualitätsentwicklung und Lehrerprofessionalisierung durch Vergleichsarbeiten. Kommentar zur Diskussion um VERA. In: Die Deutsche Schule, 98, 2006, 1
- KMK 2001: Kultusministerkonferenz beschließt konkrete Maßnahmen zur Verbesserung der schulischen Bildung in Deutschland – erste Konsequenzen aus den Ergebnissen der PISA-Studie. Presseerklärung vom 5./6. Dezember 2001
- KMK 2004a: Bildungsstandards ([www.kultusministerkonferenz.de](http://www.kultusministerkonferenz.de), weiter zu: Schule, Bildungsstandards)
- KMK 2004b: Bildungsstandards im Fach Deutsch Jahrgangsstufe 4 [www.kultusministerkonferenz.de](http://www.kultusministerkonferenz.de), weiter zu: Schule, Bildungsstandards)
- Leutner, Detlev 2005: Fragen zur Lernstandserhebung 2004 in Klasse 9: Fair vergleichen. In: neue deutsche schule, 11, S. 10
- Lind, Georg 2005: Verfehlte Evaluationspolitik? In: Grundschule aktuell, 92, S. 27
- May, Peter 2005: Hamburger Schreibprobe 1 – 9. Verlag für pädagogische Medien: Hamburg
- Metze, Wilfried 2004: Stolperwörter-Lesetest. Handreichung und Ergebnisse der Stichprobenerhebung, ([www.lesetest1-4.de](http://www.lesetest1-4.de))
- Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen 2002: Qualitätsentwicklung und Qualitätssicherung – Aufgabenbeispiele Klasse 3: Sprache
- Projekt VERA: Pädagogische Nutzung der Vergleichsarbeiten ([www.uni-landau.de/vera](http://www.uni-landau.de/vera))
- Rumpf, Horst 2004: Diesseits der Belehrungswut. Weinheim und München: Juventa
- Schlömerkemper, Jörg (Hg.) 2004: Bildung und Standards. Zur Kritik der „Instandsetzung“ des deutschen Bildungswesens. Die Deutsche Schule, 8. Beiheft. Weinheim und München: Juventa
- Skolverket: Prüfungen im Frühlingshalbjahr 2003 und 2004, aus dem Schwedischen übersetzt von Kathrin Faßrainer. Als Download: [www.grundschulverband.de](http://www.grundschulverband.de), weiter: Standards, Schwedische Fachprüfungen
- Stähling, Reinhard 2005: Qualitätsentwicklung statt Vergleichsarbeiten. Zu einem unfruchtbaren Verhältnis von Forschung und Schule. In: Die Deutsche Schule, 97, 2005, 2, S. 211-221
- VERA 4 2005. Deutsch Aufgabenheft Leseverständnis
- WAZ Westdeutsche Allgemeine 2005, Ausgabe vom 20. Dezember 2005. Essen, 1
- Horst Bartnitzky*, geb. 1940; Dipl.-Päd., Leitender Regierungsschuldirektor i.R., Leiter mehrerer Lehrplankommissionen in NRW; Bundesvorsitzender des Grundschulverbandes; Arbeitsschwerpunkte: allgemeine Grundschulfragen, Deutschdidaktik  
 Anschrift: Heinrich-Albrod-Str. 54, 47249 Duisburg;  
 Email: [horst.bartnitzky@gmx.de](mailto:horst.bartnitzky@gmx.de)