

Zeitschrift: Schulverwaltung NRW**Autor:** Martin Wellenreuther**Rubrik:** Schul- und Unterrichtsentwicklung /
Bildungsmonitoring**Referenz:** SchVw NRW 2016, 251 - 254 (Ausgabe
9)

15 Jahre nach TIMSS und PISA

Warum Bildungsmonitoring und Lehrerfortbildung nicht ausreichen dürften, unsere Schulbildung zu reformieren.

Mehr als 15 Jahre nach TIMSS und PISA stellt sich die Frage, ob die eingeleiteten Reformen die Schulbildung in Deutschland nachhaltig verbessert haben. Es ist fraglich, ob durch Bildungsmonitoring und durch Lehrerfortbildungen Leistungssteigerungen möglich sind, wenn die erforderlichen Unterstützungsstrukturen für Lehrer fehlen.

Martin Wellenreuther

Wie alles begann

Vor gut 15 Jahren wurde eine breite Öffentlichkeit über die unbefriedigenden Leistungen deutscher Schüler informiert (vgl. Baumert & Lehmann, 1997; PISA 2000). Nach TIMSS erzielten deutsche Schüler in Mathematik im Durchschnitt 509 Punkte und US-Schüler 500 Punkte. Schüler der deutschsprachigen Schweiz erreichten dagegen 590 und asiatische Länder über 600 Punkte (Singapur 643 Punkte, Korea 607, Japan 605). In PISA 2000 lagen die Anteile der Risikogruppe (Kompetenzstufe 1 und darunter) in Mathematik bei fast einem Viertel, beim Lesen betrug der entsprechende Anteil 22%. Gleichzeitig war in Deutschland die Gruppe der Schüler, die ungern lasen, besonders groß.

Den unbefriedigenden Schülerleistungen entsprachen auch Kompetenzdefizite bei Grundschullehrern. Während 70% der chinesischen Grundschullehrer Inhalte der Grundschulmathematik prozess- und verständnisorientiert erklären konnten, betrug der entsprechende Anteil bei den US-Grundschullehrern 20% (vgl. Ma, 1999). Ähnlich problematisch war die Erklärkompetenz deutscher Grundschullehrer (vgl. Seyd, 2005). Wie konnte dies im Land der Dichter und Denker passieren?

Reaktionen auf den Schock

Die schlechten Ergebnisse alarmierten damals die Politiker. Sie beauftragten anerkannte Bildungsforscher, ein Gutachten über Wege aus dem Bildungsjammertal zu erstellen. Im 2003 veröffentlichten Klieme-Gutachten stand dann: »Die internationalen Ergebnisse der PISA-Studie, insbesondere die sehr guten Ergebnisse der skandinavischen Staaten und einiger anglo-amerikanischer Staaten, lassen vermuten, dass Länder, die systematische Qualitätssicherung betreiben – sei es durch regelmäßige Schulleistungsstudien oder durch ein dichtes Netz von Schulevaluationen – insgesamt höhere Leistungen erreichen.« (Klieme, Avenarius, Blum, Döbrich, Gruber, Prenzel, Reiss, Riquarts, Rost, Tenorth & Vollmer 2003, S. 13).

Was hat sich seither verbessert?

15 Jahre nach PISA scheinen die erzielten Leistungsverbesserungen die Gültigkeit eines solchen Bildungsmonitorings zu bestätigen:

- In der Lesekompetenz verbesserten sich die deutschen Schüler von 2000 bis 2009 um 13 Punkte, in Mathematik von 2003 bis 2012 um 11 Punkte.
- Verglichen damit verschlechterte sich die Lesekompetenz der schwedischen Schüler um 19 Punkte und in Mathematik um 31 Punkte. Die Lesekompetenz der finnischen Schüler verschlechterte sich in

diesem Zeitraum um 11 Punkte, die Kompetenz in Mathematik um 25 Punkte.

- Die Risikogruppen sind in den drei Bereichen Lesen, Mathematik und Naturwissenschaften kleiner geworden. So konnte die Risikogruppe im Bereich Lesen von 25% auf etwa 17% reduziert werden.
- Der Zusammenhang zwischen sozialer Herkunft und der Bildungsbeteiligung in Deutschland ist immer noch besonders stark: Der Anteil der Kinder, die auf das Gymnasium gehen, betrug in PISA 2012 bei den un- und angelernten Arbeitern (EPG-Klasse VII) 19%, bei Kindern der oberen Dienstklasse (I) dagegen 58% (Müller und Ehmke 2012, S. 262)

Der Erfolg hat viele Väter, der Misserfolg ist ein Waisenkind

Der PISA-Schock hat viele Maßnahmen angeregt: Ausbau von Ganztagsangeboten, Reduktion von verzögerten Schulkarrieren (flexible Eingangsphase und veränderte Praxis des Sitzenbleibens) und systematische Sprachförderung schon in der Vorschule. In einigen Bundesländern wurden Sprachklassen eingeführt oder Curricula auf der Basis länderübergreifender Bildungsstandards überarbeitet. Dennoch ist es nach Köller (2016, S. 13) eine offene Frage, »[o] b all diese Maßnahmen irgendetwas Ursächliches damit zu tun haben, dass sich in Deutschland die Kompetenzen Jugendlicher zwischen 2000 und 2012 gesteigert haben.«

1. Erklärung: Die Outputkontrollen (Vergleichstests, zentrale Prüfungen, und Schulinspektionen) haben zu den besseren Leistungen geführt.

Zunächst einmal belegt die neuere Lehr-Lernforschung, dass Tests helfen, Inhalte im Gedächtnis besser zu verankern (vgl. Brown, Roediger und McDaniel 2014). Vielleicht ist jedoch ein anderer Effekt bedeutsamer: Tests können spezifische Kompetenzdefizite bei Schülern identifizieren und dadurch Lehrer veranlassen, diese Kompetenzen stärker zu trainieren.

Ein Beispiel ist die Kompetenz »sinnverstehendes Lesen«. Die sehr bescheidenen Ergebnisse bei PISA 2000 haben dazu geführt, verstärkt Übungen zum sinnverstehenden Lesen in Schulbüchern und Unterrichtshilfen anzubieten. Da Lehrer gemeinhin bemüht sind, ihre Klasse zu möglichst guten Lernergebnissen zu führen, haben sie die Angebote zur Förderung sinnverstehenden Lesens aufgegriffen und vermehrt entsprechende Übungen zum sinnverstehenden Lesen durchgeführt. Ähnliches ist in Mathematik bei der flexiblen Verwendung und Reflexion verschiedener Lösungsmethoden festzustellen. Vieles spricht dafür, dass Bildungsmonitoring vor allem die *Gewichtung von Inhalten* beeinflusst. Dagegen scheint der Einfluss des Bildungsmonitorings auf *die Verwendung bestimmter Unterrichtsmethoden* sehr begrenzt (vgl. Diamond 2007; Terhart 2015).

2. Erklärung: Der Anstieg des Gymnasiastenanteils

Seit 2000 ist der Anteil der Gymnasiasten an den 15-jährigen Schülern von 28% auf 36% gestiegen. Da Schüler auf dem Gymnasium einem viel breiteren und stärker herausfordernden Bildungsangebot ausgesetzt sind, lernen sie erheblich mehr als in der Haupt- oder der Realschule. »*Sicherlich hat eine Rolle gespielt, dass das Gymnasium trotz gestiegener Zahl der Schülerinnen und Schüler in den Leistungsständen stabil geblieben ist. Wenn wir uns die Schulformunterschiede bei PISA anschauen, sind sie riesig zugunsten der Schülerinnen und Schüler am Gymnasium.*« (Köller 2016, S. 14)

3. Erklärung: »Teaching to the test« und Kollateralschäden des deutschen Bildungsmonitorings

Die positiven Effekte können weitgehend durch den Testinflationierungseffekt erklärt werden. Dieser »*Testinflationierungseffekt*« (Koretz, Linn, Dunbar & Shepard, 1991) wurde auf folgende Weise nachgewiesen:

- 1986: Nach dem üblichen Test lag die Leistung der Schüler eine halbe Klassenstufe über der zu erwartenden Leistung.
- 1987: Als ein neuer Test verwendet wurde, »sank« die Leistung um eine halbe Klassenstufe.
- 1988, 1989 und 1990 stieg dann das mit dem neuen Test ermittelte Leistungsniveau um eine halbe Klassenstufe an.
- 1990 wurde unerwartet zusätzlich der früher verwendete alte Test eingesetzt. Nach dem alten Test waren die Leistungen der Schüler plötzlich wieder um eine halbe Klassenstufe niedriger!

Dies deutet darauf hin, dass die gefundenen Leistungssteigerungen nicht »echt« sind, sondern durch »teaching to the test« erklärbar sind.

Neben den genannten positiven Wirkungen des Bildungsmonitorings gibt es in der Forschung auch Hinweise für dessen Kollateralschäden:

- Schulen bereiteten ihre Schüler gezielt auf das Beantworten von Testfragen vor.
- »High-stake«-Testinhalte wurden vorrangig unterrichtet, während vor allem musische Fächer an den Rand gedrängt wurden (vgl. Wellenreuther 2011).
- Schüler, die voraussichtlich schlechte Testergebnisse erzielen würden, legte man nahe, am Testtag zu Hause zu bleiben (vgl. Nichols & Berliner 2007).

Die Leistungssteigerungen der »gedopten« Schulen sagen deshalb nichts über ihren tatsächlichen Leistungsstand aus.

Nichols und Berliner (2007) erklären »Kollateralschäden« verschärfter Qualitätssicherung mit Hilfe einer Gesetzmäßigkeit, die D.T. Campbell schon 1976 (S. 49) formuliert hatte: *»Je mehr irgendein sozialer Indikator zur alleinigen Grundlage für Entscheidungsprozesse gemacht wird, umso größer wird die Gefahr eines korrumpierenden Drucks und umso stärker werden soziale Prozesse, die kontrolliert werden sollen, korrumpiert.«* (vgl. Wellenreuther 2015, S. 496 – 498)

Verschärfter sozialer Druck führt eher zu Betrügereien und Manipulationen als zu echten Leistungssteigerungen. Aus der Lernforschung ist bekannt, dass erhöhter Stress Lernen behindert. Diese These hat auch für das Lernen von Lehrern seine Gültigkeit. Nach Butler (1988) kann Feedback dann nicht mehr konstruktiv als Lernangebot aufgenommen und verarbeitet werden, wenn Leistungen durch Noten evaluiert werden.

Besonders problematisch ist: Der soziale Druck, der durch Vergleichsarbeiten, Schulinspektionen und darauf folgende Lehrerfortbildungen ausgeübt wird, kann Schulen betreffen, die »unter den herrschenden Verhältnissen« hervorragende Leistungen erzielen. H.-G. Schönwälder 2016, S. 22) bemerkt dazu: *»Kann eine Schule dafür, in einem Quartier mit bildungsfernen Familien ihren Pflichten nachkommen zu müssen? Gewiss nicht. Interessant bleibt jedoch, wie sie ihre Aufgaben trotz misslicher Umstände meistert. Exzellenz dürfte sie beanspruchen, wenn niemand es besser machen würde als sie. Selbst wenn sie im Vergleich der Schülerleistungen permanent die letzten Rangplätze besetzte, könnte sie das Menschenmögliche getan haben.«*

Schulen, die trotz widriger Umstände Hervorragendes leisten, dürften wenig an solchen Vergleichsmessungen interessiert sein. Wie verhalten sich Lehrer einer Schule, die unter ihren Bedingungen Herausragendes geleistet haben und dennoch das Feedback erhalten, die Ergebnisse ihrer Schüler lägen im unteren Viertel? Werden diese Lehrer motiviert sein, sorgfältig die Ergebnisse zu studieren, um heraus zu bekommen, woran das »schlechte« Ergebnis gelegen hat? Nach Terhart (2015, S. 5) muss hier *»mit »Implementationsuntreue«, das heißt mit Indifferenz, Scheinerfüllung, passivem und aktivem Widerstand oder Obstruktion (beispielsweise durch lächerlich machen oder Verachtung gegenüber der gutwilligen Reform-Fraktion im Lehrerzimmer) gerechnet werden.«*

Erfahrungen in vielen gesellschaftlichen Bereichen zeigen: Kontrollen allein bewirken wenig Positives, wenn sie nicht durch intensive Unterstützungsmaßnahmen unterfüttert werden. Aus der Verhaltensforschung ist seit Langem bekannt, dass Bestrafungen allein keine Verhaltensänderungen bewirken. Zusätzlich müssen Lernangebote gemacht werden, um langsam ein alternatives Verhaltensrepertoire aufzubauen, das zu besseren Lernergebnissen der Schüler führt. Allerdings muss man damit rechnen, dass solche Lernangebote nicht von allen Lehrern akzeptiert werden.

Bleibt festzuhalten: Das Bildungsmonitoring hat vor allem die Gewichtung von Inhalten verändert. Die gefunden Leistungsverbesserungen können allein schon durch intensives Testtraining in bisher vernachlässigten inhaltlichen Bereichen (Leseverständnis, Begründungswissen) und durch den zunehmenden Anteil der Schüler mit gymnasialer Bildung erklärt werden. Dem stehen einige negative Effekte (»Kollateralschäden«, z.B. Vernachlässigung musischer Bildung) gegenüber.

4. Erklärung: Schulinspektion und Lehrerfortbildung

Schulinspektionen und Lehrerfortbildungen sind kostenintensive Bestandteile eines Bildungsmonitorings. Es stellt sich daher die Frage, ob in Anbetracht der Kosten diese Maßnahmen zielführend sind.

Schulinspektionen

Für Schulinspektionen benötigt man Teams, die sich nicht nur aus ehemaligen Lehrern, sondern auch aus Wissenschaftlern mit vertieften Kenntnissen in der empirisch-experimentellen Lehr-Lernforschung zusammensetzen sollten. Diese Voraussetzung ist wichtig, um in den jeweiligen Schulen auch als fachkundige Gutachter akzeptiert zu werden. Dabei kümmert sich die Schulinspektion um die Funktionsweise der Schule (z.B. Informationsaustausch zwischen der Schulleitung, Fachleitern und Lehrern) sowie um die Qualität des Unterrichts. Lehrerfortbildungen im Anschluss an Schulinspektionen können auf festgestellte Defizite reagieren und Informationen bereitstellen, die bei einer Lösung der festgestellten Probleme helfen können.

In der Praxis dürfte es schwierig sein, eine ausreichende Anzahl von Schulinspektionsteams zusammen zu stellen, die den genannten Qualitätsansprüchen genügen, wenn jede Schule wenigstens einmal in fünf Jahren evaluiert werden soll.

In Nordrhein Westfalen werden alle Grundschulen in der Regel in 15 Jahren einmal inspiziert. Entsprechend stellt Böttcher (2013, S. 503) fest, dass bislang »keine oder allenfalls schwache positive Effekte von Schulinspektion auf Unterrichts- und Schulentwicklung empirisch nachgewiesen werden.«

Für die geringen Wirkungen der Schulinspektion gibt es aber noch andere Gründe. Um lernwirksam zu sein, müsste die *Schulinspektion als Lernsituation* konzipiert sein. Die Schulinspektion erzeugt jedoch eine Leistungssituation, keine Lernsituation. Bei vielen Lehrern geht dann quasi die Jalousie herunter: Sachlich zutreffende Bewertungen werden weitgehend abgeblockt und wegrationalisiert. »Schließlich arbeite man schon lange und erfolgreich, und die Theoretiker des Schulinspektionsteams hätten ohnehin keine Ahnung von den wirklichen Problemen dieser Schule.« Solche Reaktanzeffekte treten vor allem auf, wenn erfahrene Lehrer belehrt werden. Um solche Reaktanzeffekte zu vermeiden, wurde in den Niederlanden ein anderes System der Schulinspektion mit darauf folgender Lehrerfortbildung aufgebaut. In den Niederlanden können Schulen ihre Fortbildungen, von deren Teilnahme sich die Lehrer für Ihren Unterricht am meisten versprechen, selbst einkaufen. Van Bruggen (2006, S. 121) schreibt dazu: »Niederländische Schulen haben ein Recht darauf, sich Unterstützungen einzukaufen. Dafür steht im Eintopfbudget Geld zur Verfügung, teilweise ist auch ein »Schulentwicklungsbudget« vorgeschrieben. Bei einem Netzwerk von ca. 50 Schulbegleitungsdiensten mit insgesamt ca. 2 000 Experten kann jede Schule Hilfe, Schulbegleitung, Fortbildung und Coaching, etc. kaufen...Zum Vergleich: Das Inspektorat zählt ca. 200 Inspektoren (plus ca. 230 Mitarbeiter); die Unterstützungseinrichtungen haben ca. 2 500 Mitarbeiter. Das ist ein absolut passendes Verhältnis.«

Entsprechend fordert Diane Ravitch (2010, S. 238) für das amerikanische Schulsystem: »Wenn Schulen zu kämpfen haben, sollten die Autoritäten tun, was notwendig ist, um Schulen zu verbessern.« Denn ihr Job sei die Lösung von Problemen: »Dies kann Fortbildung für Lehrer bedeuten, kleinere Klassen, maßgeschneiderte Programme zum Lesenlernen oder in anderen Bereichen nachschulische Aktivitätsangebote, zusätzliche tutorielle Betreuung für Schüler und ...«

Lehrerfortbildungen

Ein eigenes Budget für Fortbildungen haben auch die nordrhein-westfälischen Schulen. In dieser Hinsicht liegen günstige Voraussetzungen vor. Zu beantworten bleiben dann noch folgende Fragen: Wie gut, wie wirksam sind diese Fortbildungen? Wird ihre Güte zertifiziert – und zwar nach nachvollziehbaren Standards? Gibt es zu den drängenden Fragen (z.B. als zentrales Problem die ungenügende Förderung von Kindern aus bildungsfernen Milieus) in ausreichend Maße gute Fortbildungen? Und wie verhalten sich die in NRW angebotenen Fortbildungen zu den in der Forschung untersuchten Konzepten?

Es gibt Belege, dass Lehrerfortbildungen unter ganz bestimmten Voraussetzungen wirksam sein können, doch werden die dafür notwendigen Rahmenbedingungen in der Praxis nur selten realisiert (vgl. dazu

Lipowsky 2010; Kennedy 2014; Timperley et al. 2007; Yoon et al. 2007). Nach den bisher durchgeführten Forschungen können Lehrerfortbildungen dann am ehesten wirksam sein, wenn sich die Fortbildung über mindestens *30 Stunden* erstreckt und außerdem die Anwendung des Trainings in der Klasse durch Coaches über mehrere Monate begleitet wird. Nicht zuletzt aus Kostengründen dürfte es kaum möglich sein, solche Fortbildungen für einen nennenswerten Teil der Schulen zu realisieren.

Eine offene Frage ist auch, wieweit die Dauer der Fortbildung durch einen verstärkten Einsatz von Unterrichtslektionen und Lehrvideos vermindert werden kann. Die untersuchten Trainingsprogramme wurden bisher nicht unter diesem Aspekt optimiert.

Dass auch vergleichsweise kurze Lehrerfortbildungen deutliche Effekte haben können, zeigen die Studie von Anderson, Evertson und Brophy (1979) sowie eine Studie von Harris, Graham & Adkins (2015).

Fazit: Spitzenergebnisse ohne verschärftes Bildungsmonitoring

Länder wie Finnland (PISA 2000) oder die deutschsprachige Schweiz erzielten auch ohne verschärfte Kontrollen Spitzenergebnisse. Für die positiven Ergebnisse Finnlands scheinen dabei eine strenge Form direkter Instruktion (Heller Sahlgren, 2015) sowie ein System intensiver Förderung (Kivirauma & Ruoho, 2007) eine entscheidende Rolle zu spielen, wobei der deutliche Rückgang der PISA-Werte in Finnland in den letzten Jahren vermutlich durch eine Abkehr von dieser strengen direkten Instruktion zu erklären ist. Auch die Hochleistungsländer in Asien (Japan, China, Korea und Singapur) sind für ihren stark lehrerzentrierten Unterricht bekannt. Etwas anders liegen die Verhältnisse in der Schweiz: Hier orientiert man sich im Mathematikunterricht stark an den Arbeiten Aebli, dessen moderner Unterricht als eine Form direkter Instruktion gedeutet werden kann (vgl. Aebli 1968; Wellenreuther 2015). Vielleicht sind solche Methoden des Unterrichtens für den Unterrichtserfolg bedeutsamer als die Methoden der Qualitätskontrolle.

Literatur

Eine Literaturliste kann über die Redaktion (nicole.pohl@wolterskluwer.com) angefordert werden.



Martin Wellenreuther
Universität Lüneburg