

Vorurteil:

Mehr Tests steigern die Leistungen von Schülern, Lehrern, Ländern

Vergleichende Leistungstests sind zentrales Element einer wettbewerbsorientierten Reform des Bildungswesens. Diese sog. »Output«-Steuerung fordert präzise definierte Lernziele (»Bildungsstandards«) und die Überprüfung der erreichten Kompetenzen mithilfe von standardisierten Tests. Damit werden drei Hoffnungen verbunden:

- Die Wirkungen konkreter Reformmaßnahmen und der Erfolg des Bildungssystems insgesamt sollen transparenter und besser kontrolliert werden können.
- Lehrkräfte und Schulen sollen durch den Vergleich und den Wettbewerb untereinander die Qualität ihrer Arbeit verbessern.
- Die Lernentwicklung der Schülerinnen und Schüler soll diagnostisch besser erfasst und dann auch besser gefördert werden können.

Aber lassen sich diese Versprechen dadurch einlösen, dass mehr getestet wird?

>> Keine positiven Auswirkungen bei internationalen Ländervergleichen

Beim internationalen IEA-Lesevergleich 1991 lagen die Viertklässler in West- und Ost-Deutschland im Mittelfeld. Zehn Jahre später, bei der ersten IGLU-Lesestudie 2001, gehörte das vereinigte Deutschland zum obersten Leistungsviertel – obwohl es in diesen zehn Jahren im Grundschulbereich keine innerdeutschen und keine länderübergreifenden Vergleiche mit standardisierten Tests gab. Internationale Vergleiche mit standardisierten Tests und die jährlichen VerA-Tests in dritten Klassen wurden erst danach eingeführt. Wieder zehn Jahre später, bei IGLU 2011, lag die deutsche Grundschule nur noch im oberen Mittelfeld, bei im Wesentlichen unverändertem Punktwert (Bos u. a. 2012, 97). Beim innerdeutschen IQB-Bildungstrend (2011 bis 2016) wurde sogar ein

leichter Rückgang der Leseleistungen von 500 auf 493 Punkte festgestellt (Stanat u. a. 2017, 158).

Zudem haben sich in den testfreudigen Bildungssystemen England, Niederlande, Schweden und USA bei IGLU die Leistungen von 2001 bis 2011 in ganz unterschiedlichen Richtungen entwickelt (Bos u. a. 2012, 101). Ähnlich divergent ist die Entwicklung in verschiedenen Provinzen Kanadas mit systematischen Testprogrammen. Vergleiche von stärker bzw. schwächer testorientierten Bundesstaaten in den USA ergeben sogar eher Nachteile für diejenigen Systeme, die auf »High-Stakes Testing« setzten (Brügelmann 2015, 92, 104).

>> Keine nachweisbaren Verbesserungen durch regelmäßige Tests auf Schul-Ebene

Die Rückmeldung von Lernständen in standardisierter Form wird auch als wesentlicher Impuls für die Verbesserung von Unterricht propagiert und hat in Hamburg zu einem Programm regelmäßiger Leistungstests auf verschiedenen Jahrgangsstufen geführt. Zwar haben sich die Hamburger Schulen im IQB-Grundschultrend von 2011 bis 2016 – anders als etwa im benachbarten Stadtstaat Bremen – verbessert. Abgesehen von der günstigeren Zusammensetzung der Schülerschaft (weniger Armut, Arbeitslosigkeit, Schwächen in der Schulsprache usw.) kann das aber auch daran liegen, dass Hamburg wesentlich mehr in die Arbeit der Grundschulen investiert als Bremen und andere Bundesländer. Außerdem wurde über viele Jahre ein differenziertes System umfangreicher Fördermaßnahmen in den Bereichen Sprache bzw. Lesen und Schreiben aufgebaut. Dafür, dass positive Leistungsentwicklungen nicht von regelmäßiger Testkontrolle abhängig sind, spricht auch, dass die weniger stark testorientierten Bundesländer Berlin, Hessen und Schleswig-Holstein im Lesen ebenfalls deutliche Zugewinne erzielten (Stanat u. a. 2017,

159). Auch die Schul- und Unterrichtsentwicklung profitiert nicht wie erwartet von der Rückmeldung der Testergebnisse an Lehrerinnen und Lehrer (Böttcher 2013; Richter u. a. 2014). Es besteht sogar die Gefahr, dass der Unterricht zunehmend durch Inhalte und Aufgabenformate bestimmt wird, die gut zu testen sind (Herzog 2013, 73 ff).

»» Lerndiagnostisch können Tests das Lehrerurteil nicht ersetzen

Eine Standardisierung von Aufgaben wird auch damit begründet, dass das Lehrerurteil bekanntermaßen fehleranfällig ist (vgl. Faktencheck »Noten«). Für Tests als Alternative werden folgende Vorteile ins Feld geführt:

- Erhebung und Auswertung sind unabhängig von den beteiligten Personen.
- Anforderungen sind für andere berechenbar und die Bewertungen durchsichtig.
- Die Maßstäbe und Ergebnisse werden durch Bezug auf Normstichproben über die Lerngruppe hinaus vergleichbar.

Andererseits folgen aus der Standardisierung der Tests zwangsläufig auch Schwächen im Vergleich zu einer begleitenden Lernbeobachtung (Brügelmann 2015, 51 ff., 120 ff.):

- Durch die punktuelle Erhebung sind die Leistungen stärker form- und situationsabhängig.
- Jeder Test erfasst nur kleine Ausschnitte aus dem Kompetenzspektrum eines Faches, und das für die Standardisierung übliche Format (»Papier und Bleistift«, Multiple-Choice-Verfahren) erlaubt nur bestimmte Aufgabentypen.

- In der Leistung wirken verschiedene Teilkompetenzen zusammen (bei Sachaufgaben z.B. neben der Mathematik Weltwissen, Lesefertigkeit, Konzentrationsfähigkeit), sodass Ergebnisse sich nicht eindeutig einer Teilkompetenz zuordnen lassen.
- Um der Objektivität willen konzentriert sich die Auswertung auf die Leistungsoberfläche (falsch vs. richtig), Lösungen und Fehler sind aber mehrdeutig im Blick auf die beteiligten/fehlenden Kompetenzen.
- Die künstliche Testsituation (zusätzlich unter Zeitdruck) erlaubt keine direkte Übertragung auf Alltagssituationen.
- Meist sind die individuellen Lernvoraussetzungen nicht bekannt, ohne die der Lernfortschritt nicht fair beurteilt werden kann, zumal fachlich wie sozial schwächere Schulkinder in Testsituationen eher noch schwächer abschneiden.

Fazit: Auch Testergebnisse sind interpretationsbedürftig, d.h. auf die diagnostische Kompetenz und das konkrete Hintergrundwissen der Lehrperson angewiesen.

Der Außenblick vermittelt nur eine andere, keine bessere Sicht auf das Lernen der Kinder. So zeigt auch das schon über mehrere Jahrzehnte laufende National Assessment of Educational Performance in den USA, dass sich die Leistungssituation im Lesen trotz ständiger Tests nicht verbessert hat (Rampey u. a. 2009).

Literatur

Böttcher, W. (2013): Das Monitoring-Paradigma – Eine Kritik der deutschen Schulreform. In: Empirische Pädagogik, 27. Jg., H. 4, 5-21.

Bos, W. u. a. (2012): IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Waxmann: Münster u. a.

Brügelmann, H. (2015): Vermessene Schulen – standardisierte Schüler. Zu Risiken und Nebenwirkungen von PISA, Hattie, VerA & Co. Beltz: Weinheim/ Basel.

Herzog, W. (2013): Bildungsstandards. Eine kritische Einführung. Kohlhammer: Stuttgart.

Rampey, B. D. et al. (2009): NAEP 2008 trends in academic progress. Washington D.C.: U. S. Department of Education: Washington, D. C.

Richter, D. u. a. (2014): Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten: Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen von Schülerinnen und Schülern. In: Zeitschrift für Pädagogik, 60. Jg., H. 2, 225-244.

Stanat, P. u. a. (2017): IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich. Waxmann: Münster/ New York.