

Kapitel 9: Bildungsstandards und externe Überprüfung von Schülerkompetenzen: Mögliche Beiträge externer Messungen zur Erreichung der Qualitätsziele der Schule

Web-Dokument 9.7

Unerwünschte Wirkungen von Leistungsvorgaben und externen Leistungsüberprüfungen unter high-stake Bedingungen

Herbert Altrichter und Anna Kanape-Willingshofer

Koretz (2008) nennt als positive Effekte von Leistungsvorgaben und externen Leistungsüberprüfungen auf Seiten der Lehrer/innen die Steigerung der effektiven Lernzeit, mehr Arbeitsanstrengung oder den Einsatz effektiverer Unterrichtsmethoden. Darüber hinaus berichtet er über weitere Strategien zur Verbesserung von Schülerleistungen: Durch *Umschichten* von Unterrichtsressourcen wird beispielsweise mehr Unterrichtszeit für getestete Fächer und Inhalte sowie auf getestete Schülergruppen konzentriert. Durch *Abstimmen* von Unterricht erfolgt auch eine inhaltliche Fokussierung auf Bildungsstandards und Standardtestungen, wodurch nicht getestete Kompetenzen und Fähigkeiten möglicherweise vernachlässigt werden. Schließlich erfolgt gezieltes *Coaching*, einerseits mit Bezug auf den getesteten Inhalt, aber auch in Hinblick auf Strategien zur Testbearbeitung.

Ein in der Zwischenzeit umfangreicher Literaturkorpus thematisiert nicht-intendierte Folgen von high stake-Testungen. Nicht nur der direkte Druck auf Schüler/innen und Schulen, sondern auch der indirekte Einfluss der Thematisierung von Testergebnissen in den Medien, in der Administration der Schulbezirke und der Gemeinde (vgl. Herman et al., 1990), führt zu strategischem Verhalten von Schulen, Lehrer/innen wie Schüler/innen, wodurch die Ergebnisse letztlich an Aussagekraft und Validität verlieren (vgl. Koretz, 2008). Dabei werden folgende zentrale Probleme genannt (vgl. Franklin & Snow-Gerono, 2007; Koretz, 2008; Maag Merki, 2010; Maier & Kuper, 2012, S. 90-91; Morris, 2011; Nichols & Berliner, 2005):

(1) *Konzentration auf testmethodische und -strategische Kompetenzen*: Mehr Unterrichtszeit wird verwendet, um Teststrategien (z.B. Umgang mit multiple-choice Aufgaben) zu vermitteln und typische Aufgabenformate bzw. Aufgabenbeispielen aus vergangenen Tests, zu üben. Lehrgelenkte Unterrichtsmethoden und testähnliche Aufgabenformate werden bevorzugt. Dies führt zu einer Beschränkung der Lerninhalte sowie zu vermehrtem Auswendiglernen und weniger Problemlösen und Verstehen.

(2) Eine *Verengung des Curriculums* findet sowohl durch Umschichtung von Unterrichtszeit zwischen Fächern (z.B. von Musik oder Geschichte zu getesteten Fächern wie Mathematik oder Englisch) als auch

innerhalb von Fächern (verstärkte Konzentration auf testrelevante Fähigkeiten und Inhalte, z.B. mehr Lesen auch im Geschichtsunterricht) statt. 79% der Lehrer/innen in US-Bundesstaaten mit high stake-Bedingungen berichten solche Handlungen (King & Zucker, 2005).

(3) Ein *Ausschluss von Schüler/innen aus Lerngelegenheiten* kann in Form der Zuweisung zu einer nicht getesteten Kategorie (z.B. Schüler/innen mit Beeinträchtigungen) oder zu anderen Schulen bzw. durch Nicht-Teilnahme von schlechteren Schüler/innen oder Abgang von der Schule erfolgen (vgl. auch Eurydice Network, 2009, für die Niederlande). Schüler/innen im Grenzbereich, die die accountability-relevanten Leistungsziele noch erreichen können, erhalten stärkere Aufmerksamkeit bei der Prüfungsvorbereitung als ‚sichere‘ und ‚hoffnungslose‘ Fälle. Die Verwendung von Wachstumsmodellen, bei denen der Leistungszuwachs und nicht die gezeigte Leistung von vorrangiger Bedeutung ist, könnte diesen nicht-intendierten Effekt verringern.

(4) *Schummeln auf Lehrerseite* erstreckt sich von ‚kleinen‘ Hilfestellungen während der Testdurchführung, wie Aufmerksammachen auf ausgelassene Items, Betonung beim Vorlesen der Instruktion etc. bis zur Manipulation der Testdaten in Form von nachträglichem Ergänzen oder Ändern von Antworten.

„Teaching to the test“ hat keinen guten Ruf, kann aber auch innovative Impulse für den Unterricht bringen. Eine verbesserte Generation von Testaufgaben, die auf höherwertige Lern- und Denkstrategien zielte, zeigte Effekte auf die Unterrichtsgestaltung, die jedoch – wie in den deutschsprachigen Untersuchungen zum Datenfeedback – von begrenzter Reichweite waren: Alternative Aufgabenformate wurden im Unterricht aufgenommen, ohne dass Lehrpersonen deswegen ihre fachdidaktischen Strategien einer grundsätzlichen Reflexion und Revision unterzogen (Stecher, 2002; Faulkner & Cook, 2006).

Franklin und Snow-Gerono (2007) beschreiben eine Reihe von nicht-intendierten Nebeneffekten der Einführung von neuen standardisierten high stakes-Tests auf Lehrer/innen, wie zunehmender Stress und Druck sowie abnehmende Arbeitszufriedenheit (vgl. Herman et al., 1990). Nach Figlio und Loeb (2011) führen high-stake-Messungen auch zu Veränderungen am Arbeitsmarkt: Lehrer/innen meiden schwächere Schulen oder versuchen diese möglichst bald wieder zu verlassen.

Der Rechenschaftsdruck kann sich aber auch direkt auf Schulen richten, in dem bestimmte Testleistungen oder jährlich steigende Anteile von Schüler/innen, die Mindeststandards erreichen (Adequate Yearly Progress), für verschiedene Teilgruppen der Schülerpopulation gefordert werden. Das Verfehlen solcher Leistungsziele zieht verschiedene Sanktionen nach sich, z.B. von der Veröffentlichung der Ergebnisse über den Verlust des bisherigen Rechtsstatus und der Entfernung von Lehrpersonen oder Schulleitung bis zur Schließung.

Referenzen:

- Eurydice Network (2009). *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. Brussels: Education, Audiovisual and Culture Executive Agency. Zugriff am 03.07.2012 unter http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf
- Faulkner, S. A. & Cook, C. M. (2006). Testing versus Teaching: The Perceived Impact of Assessment Demands on Middle Grade Instructional Practices. *Research in Middle Level Education* 29 (7), 1-13.
- Franklin, C. A. & Snow-Gerono, J. L. (2007). Perceptions of teaching in an environment of standardized testing: Voices from the field. *The Researcher*, 21 (1), 2-21.
- Figlio, D. & Loeb, S. (2011). School Accountability. In E. Hanushek, S. Machin & L. Woessman (Eds.), *Handbook of the Economics in Education, Vol. 3* (pp. 383-421), Amsterdam: Elsevier.
- Herman, J., Dreyfus, J. & Golan, S. (1990). The effects of testing on teaching and learning. Center for Research and Evaluation, Standards, and Student Testing, University of California at Los Angeles. Zugriff am 06.07.2012 unter <http://www.cse.ucla.edu/products/reports/TR327.pdf>
- King, K. & Zucker, S. (2005). *Curriculum Narrowing*. Pearson Education. Zugriff am 06.07.2012 unter <http://www.pearsonassessments.com/NR/rdonlyres/D3362EDE-7F34-447E-ADE4-D4CB2518C2B2/0/CurriculumNarrowing.pdf>
- Koretz, D. (2008). Test-based Educational Accountability. Research Evidence and Implications. *Zeitschrift für Pädagogik*, 54, 777-790.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 145-169). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maier, U. & Kuper, H. (2012). Vergleichsarbeiten als Instrumente der Qualitätsentwicklung. *Die Deutsche Schule*, 104 (1), 88-99.
- Morris, A. (2011). *Student standardised testing: Current practices in OECD countries and a literature review*, OECD Education Working Papers, No. 65, OECD Publishing. Zugriff am 30.08.2012 unter <http://dx.doi.org/10.1787/5kg3rp9qbnr6-en>
- Nichols, S. L. & Berliner, D. C. (2007). *Collateral Damage. How High-Stakes Testing Corrupts America's Schools*. Cambridge: Harvard Education Press.
- Stecher, B. M. (2002). Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practice. In L. S. Hamilton, B. M. Stecher & S. P. Klein (Hrsg.), *Making Sense of Test-Based Accountability in Education* (S. 79-100). Santa Monica, CA.: Rand.