

Thomas Jahnke, Potsdam

Die empirische Wünschelrute und ihre Folgen

Das eigenartige an der empirischen Forschung ist, dass sie immer etwas hervorbringt; ob kleinste Examensarbeit oder International Survey, sie kann nicht leer ausgehen. Immer erhält sie Zahlen und damit Skalen, Intervalle, Ranglisten und, was dann der statistische Apparat dienlich zur Verfügung stellt, Mittelwerte, Standardabweichungen, Korrelationen und so fort. Man wünscht sich mit der empirischen Rute gleichsam Ergebnisse herbei. Durch das Messen erhält man Daten – eigentlich sogar nur Zahlen, deren Realitätsgehalt (oder genauer Realitätsbeschreibungsvermögen) außer Frage steht, ja dem ganzen Prozess eine Art quasi-naturwissenschaftlicher Dignität verleiht. Die Frage, ob man da tatsächlich etwas misst, was die Realität beschreibt und sinnvoll begreift, stellt sich gar nicht, weil der Gegenstand durch den Messprozess hervorgebracht und konstituiert wird. Solche Forschung produziert ihre Begriffe (heute sagt man dann auch *Konzepte*) und Ergebnisse parthenogenetisch, notfalls auch ohne jede Anleihen bei der bemessenen Realität oder Bezüge zu ihr. Und wenn man die Rotzigkeit (Pardon!) des Positivismus zugrunde legt, gilt das dann alles bis zu seiner Widerlegung, die – dessen sind sich die Pisaner und die involvierten *educational assessment*-Firmen, die PISA entwickelt und an bisher achtundfünfzig Länder verkauft haben, sicher – faktisch wenn nicht sogar prinzipiell ausgeschlossen ist. Die Sache ist versiegelt. Wo Kritik nicht abprallt, freut sich die Forschungsindustrie über Folgeaufträge, die dann in der beschriebenen Art bearbeitet werden.

Eine erste Folge einer testorientierten Schulbildung und einer testresultat-gesteuerten Bildungspolitik ist, dass das, was in der Philosophie und der Soziologie als Entfremdung bezeichnet, in einer neuen Qualität den Lehrerberuf und das Lehren und Lernen an der Schule durchdringt. Es geht nicht mehr um die Sache selbst, um die Schüler, um die Lehrer, um ein gemeinsames Bildungserlebnis, es geht letztlich um Testresultate, anhand derer der Erfolg des Lehrens und Lernens gemessen und beurteilt wird. Verächtlicher und desinteressierter kann man mit einem Schüler kaum umgehen, als dass man sich z.B. in Multiple-Choice-Tests nur noch dafür interessiert, ob er an der richtigen Stelle ein Kreuz macht, und ihm dafür keinen oder einen Punkt gibt. Auch das hässliche Kompositum *Bildungsstandards* (die überdies auch noch *normiert* werden) bekräftigt aus sich selbst heraus, dass es nicht mehr um individuelle Lernprozesse oder kulturelle Teilhabe geht.

Beispiele für die Folgen solcher Entfremdung kann man (nicht nur den amerikanischen) Zeitungen entnehmen.

New York: Schüler bekommen für gute Leistungen bis zu 50 Dollar

Wie die "New York Times" berichtet, will die Stadt New York Schüler zu besseren Leistungen anspornen. Für gute Noten können die Schüler bis zu 50 Dollar erhalten. Insgesamt wurde bereits eine halbe Million US-Dollar ausbezahlt. 200 Schulen beteiligen sich an dem Programm, bei dem auch den Lehrern eine Sonderzahlung am Ende des Jahres winkt, falls Schüler ihre Noten verbessert haben. (...) Ende Februar wurde zudem damit begonnen, Schüler für gute Leistungen und vorbildliches Verhalten mit einem kostenlosen Handy zu belohnen. (Quelle: <http://www.shortnews.de/start.cfm?id=700812>)

Dieses Beispiel erscheint noch recht oberflächlich, eher als solches für problematische gesellschaftliche Verkehrsformen, fragliche Schulverhältnisse und für eine Verrohung pädagogischer Sitten oder Normen. Es geht viel grundsätzlicher darum, ob die in schulischen Bildungsprozessen angelegte, unmittelbare und authentische Begegnung der Beteiligten miteinander und mit der Sache nun durch eine äußere Zielvorgabe und ein äußeres Maß in ein mittelbares Verhältnis transformiert wird, dessen Zwecke deutlich und sichtbar außerhalb des eigentlichen Bildungsprozesses liegen.

Eine zweite Folge der Orientierung des Schulunterrichts auf Testbildung ist die Abwertung der Fächer, die nicht abgetestet werden. Ihr Wert sinkt herab zu dem einer ‚sozialen Veranstaltung‘, wie man zuweilen in leistungsorientierten Kreisen formuliert. Ästhetische Bildung und moralische Urteilsfähigkeit – um Termini des Erziehungswissenschaftlers Volker Ladenthin¹ aufzunehmen – sind in dem, was in der Testideologie als Grundbildung bezeichnet wird, offensichtlich verzichtbar. Dieser Verzicht wird sich gerade auch auf den Unterricht in den Grundschulen auswirken, die zunehmend nur als Orte der Präparation für die so genannten weiterführenden Schulen gesehen werden.

Eine dritte Folge der Testorientierung ist die Zunahme des Drucks auf die ohnehin überlasteten Lehrerinnen und Lehrer. Unter der Hand wird gerade von sich fortschrittlich dünkenden Fachdidaktikern – zum Teil wie mir scheint sogar frohlockend – festgestellt, dass man nun endlich ein taugliches Mittel gefunden habe, die Unterrichtenden an der Kandare zunehmen und in die Richtung einer als modern bezeichneter Didaktik zu zwingen. Dieser Zwang zum Fortschritt macht deutlich, dass es sich bei der Wünschelrute tatsächlich auch um eine Rute handelt.

¹ Vgl. Volker Ladenthin: PISA - Recht und Grenzen einer globalen empirischen Studie. Eine bildungstheoretische Betrachtung. Vierteljahrsschrift für wissenschaftliche Pädagogik 84 (2008)

Teaching to the Test?

Eine vierte Folge von Pisa&Co betrifft die vermeintlich qualitätsfördernde Wirkung eines testorientierten Unterrichts. Die internationale Kritik an den Testverfahren und an dem Gedanken, die ‚Erträge‘ schulischer Bildung könnten über periodisierte Tests in sinnvoller Weise gemessen und gesteigert werden, hat Deutschland nicht erreicht, und es ist bisher auch kaum zu einer sorgfältigen und redlichen Diskussion dieses selbstverständlich erscheinenden Gedankens hierzulande gekommen. Das ist auch nicht weiter erstaunlich, weil von den involvierten Testinstituten und den mit ihnen kooperierenden Wissenschaftler kaum zu erwarten ist, dass sie mit ihrem Test-Knowhow auch die Test-Kritik auf den Markt bringen.

Es liegt nahe, wenn man die ‚Weiterentwicklung des Unterrichts und vor allem die individuelle Förderungen der Schülerinnen und Schüler‘ durch Bildungsstandards² befördern oder anordnen will, deren Erreichen im Wesentlichen durch Tests überprüft wird, die Erfahrungen von Ländern und darunter insbesondere der USA zu rezipieren, die seit Jahren oder Jahrzehnten eine solche Politik verfolgen. Dass dies immer wieder auch zu unangemessenen bzw. unangebrachten Formen der Vorbereitung von Schülerinnen und Schülern auf diese Tests und damit zur Verhinderung bzw. Verminderung ihrer Lernfortschritte führt, ist bekannt.

Daniel Koretz, Erziehungswissenschaftler an der Havard-Universität und assoziierter Direktor des Center of Research, Standards, and Student Testing (CRESST) beschreibt in seinem Aufsatz *Alignment, High Stakes, and the Inflation of Test Scores*³ einen Ausgangspunkt, über den eine öffentliche Diskussion in Deutschland bisher kaum hinausgekommen ist:

On common response to this problem has been to seek „tests worth teaching to“. The search for such tests has led reformers in several directions over the years, but currently, many argue that tests well aligned with standards meet this criterion. If tests are aligned with standards, the arguments runs, they test material deemed important, and teaching to the test therefore teaches what is important. If students are being taught what is important, how can the resulting score gains be misleading? (p. 99)

Koretz begründet seinen Widerspruch gegen eine naive Testgläubigkeit theoretisch und empirisch unter anderem eindrücklich mit Sägezahnkurven („sawtooth pattern“) für die gemessenen Leistungen der gleichen oder einer vergleichbaren Population, die sich in verschiedenen Erhebungen je nach den verwendeten Tests in unterschiedlichster Weise ergaben. Eine

² Beschluss der Kultusministerkonferenz zum Bildungsmonitoring vom 02.02.2006

³ Koretz, D.: *Alignment, High Stakes, and the Inflation of Test Scores*. Yearbook of the National Society for the Study of Education (2005) 104 (2), 99–118. (Online unter: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1744-7984.2005.00027.x>)

zusammenfassende Kritik einer testgeleiteten Schulpolitik ist u.a. bei Brian M. Stecher unter „*Consequences of large-scale, high-stakes testing on school and classroom practices*“ in dem von ihm mit herausgegebenen Buch *Making Sense of Test-Based Accountability in Education*⁴ nachzulesen:

The net effect of high-stakes testing on policy and practice is uncertain. Researchers have not documented the desirable consequences of testing – providing more instruction, working harder, and working more effectively – as clearly as the undesirable ones – such as negative reallocation, negative alignment of classroom time to emphasize topics covered by a test, excessive coaching, and cheating. More important, researchers have not generally measured the extent or magnitude of the shifts in practice that the identified as a result of high-stakes testing.

Overall, the evidence suggests that large-scale high-stakes testing has been a relatively potent policy in terms of bringing about changes within schools and classrooms. Many of these changes appear to diminish students' exposure to curriculum, which undermines the meaning of the test scores. (p. 99/100)⁵

Der im letzten Absatz angesprochene Antagonismus scheint der deutschen Kultusministerkonferenz möglicherweise von ihren Beratern vorenthalten worden zu sein: dass sich an den Schulen etwas bewegt, heißt noch nicht, dass diese Bewegung auch sinnvoll ist.

In dem *Position Statement on High Stakes Testing in PreK-12 Education* der *American Evaluation Association* (AEA) wird resümiert:

There is evidence that such testing often leads to educationally unjust consequences and unsound practices, even though it occasionally upgrades teaching and learning conditions in some classrooms and schools. The consequences that concern us most are increased drop out rates, teacher and administrator deprofessionalization, loss of curricular integrity, increased cultural insensitivity, and disproportionate allocation of educational resources into testing programs and not into hiring qualified teachers and providing sound educational programs. (...)

Comparisons of schools and students based on test scores promote teaching to the test, especially in ways that do not constitute an improvement in teaching and learning. Although used for more than two decades, state mandated high stakes testing has not improved the quality of schools; nor diminished disparities in academic achievement along gender, race or class lines; nor moved the country forward in moral, social, or economic terms.

(Siehe <http://www.eval.org/hst3.htm!>)

⁴ Stecher, B. M.: Consequences of large-scale, high-stakes testing on school and classroom practices. In L. S. Hamilton, B. M. Stecher, and S. P. Klein (Eds.): *Making Sense of Test-Based Accountability in Education*. RAND. Santa Monica 2002. P 79 -100 (Online unter: http://www.rand.org/pubs/monograph_reports/MR1554/index.html)

⁵ Hervorhebungen hier und im nächsten Zitat durch Th. J.