



Center for  
K–12 Assessment  
& Performance Management

*An independent catalyst and resource for the improvement of measurement and data systems to enhance student achievement.*

**Exploratory Seminar:**  
Measurement Challenges Within  
the Race to the Top Agenda  
December 2009

## SOME IMPLICATIONS OF CURRENT POLICY FOR EDUCATIONAL MEASUREMENT

This policy brief is based on a presentation by Daniel Koretz (Harvard University) at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, December 2009. Download a copy of the final paper written by Dr. Koretz, as well as the other papers presented at the seminar, at <http://www.k12center.org/publications.html>.

*The Race to the Top competition for federal funding intensifies the trend toward using testing for high-stakes accountability even though there is ample evidence of perverse side effects. The measurement community must make this the core concern.*

For the last half century, large-scale assessment results have been used more and more for monitoring and accountability. No Child Left Behind legislation and now the Race to the Top funding are intensifying that trend.

There is a lack of persuasive evidence of positive effects from test-based accountability. On the other hand, there is persuasive evidence of unwanted side effects, such as degraded instruction (teaching to the test), gaming the system (such as focusing on borderline students), and score inflation. Copious research has shown similar problems in other fields, but the field of education has not learned from this. As a result, there has been a repeated cycle of proclamations of success based on inflated scores, followed by a crisis because of evidence that student performance remains inadequate. The response is to devise a new round of test-based accountability, usually more severe than the previous one. Race to the Top can be seen as the most recent example of this.

Confronting these problems requires improvements in the design of both accountability systems and the tests used in them. This presentation focuses on necessary improvements in testing.

### What the Field Has Done

To date, the field has attempted to make two types of changes in response to test-based accountability:

- *Changes in test design.* This includes many attempts to test higher-order skills better. New designs also encourage certain styles of instruction and classroom assessments.
- *Statistical and psychometric innovations.* Currently these include growth modeling and standards-based reporting. The latter largely is a change for the worse.

*Created by Educational Testing Service (ETS) to forward a larger social mission, the Center for K–12 Assessment & Performance Management has been given the directive to serve as an independent catalyst and resource for the improvement of measurement and data systems to enhance student achievement.*

None of these changes, however, have addressed some of the central problems raised by test-based accountability.

- *The measurement field has not drawn from research in other fields on accountability systems.* Rather, it has proceeded as if it were working in isolation. It also has not conducted sufficient research on the problems being encountered in test-based accountability.
- *It has not addressed adequately the implications of test-based accountability for the field's own activities.* The field needs to consider how test designs should be modified, how validation studies should be augmented, and whether routine operation of testing programs must be changed.

## **Campbell's Law in Education Accountability**

Campbell's Law (Donald Campbell) contends that the more a quantitative indicator is used for social decision-making, the more likely it is to become corrupted and the more likely it will distort and corrupt the social processes it is used to monitor. Applied to high-stakes testing, this means that when test scores become the goal of teaching, they may become misleading as indicators of educational progress and may distort the educational process in undesirable ways.

In the stages of constructing a test, for example, the share of the domain that is tested becomes progressively less complete. There are two aspects to this incomplete sampling. There is generally limited and somewhat predictable sampling of the substantively important parts of the domain (e.g., content standards, or content areas within standards). In addition, there are typically predictable patterns in nonsubstantive aspects of the test, such as item format, other aspects of the style of items, unimportant fine details of content, and so on. Test-based accountability gives teachers an incentive to focus on the tested sample, at the expense of other important aspects of the domain. This focus can take the form of reallocation among content strands or coaching that focuses on unimportant, nonsubstantive aspects of the test. Both forms of test prep can inflate scores.

## **What Can Be Done**

One of the steps needed to lessen Campbell's Law lies outside the field of measurement: Making more than test scores count in the accountability system. For the field of measurement, the needs begin with test design and extend to evaluation:

- *Limit predictable recurrences.* The field needs to reduce both substantive and nonsubstantive recurrences in test designs in order to change the incentives they create and make it difficult for people to game the system.
- *Expand validation and evaluation.* Traditional validation is necessary but it is also incomplete. It cannot evaluate inferences about gains, because it is cross-sectional and is largely completed before inflation occurs. Assessment systems need to routinize audits and ongoing validation and conduct direct monitoring of the systems' effects on schooling.

In addition, the field must scrutinize the entire range of routine activities to determine which other aspects of the measurement enterprise are compromised by the pressures of accountability. One example is linking over time. The assumptions needed for some of the most common linking approaches are not warranted when stakes are high.

### **New Approaches and Campbell's Law**

Growth modeling does nothing to solve the problems of Campbell's Law. It does not lessen bad incentives and score inflation, although the specific forms inflation takes may change (e.g., issue of persistence of coaching effects over grades).

Rather than lessening the effects of Campbell's Law, relying more on complex performance tasks may exacerbate those effects. Performance tasks take more time and therefore reduce the number of tasks tested, which can increase the impact of predictability. Performance tasks are also often memorable, making it easier to focus teaching on the tested sample.

These are not arguments against further development of performance assessments or methods of growth modeling. However, it is important to realize that these innovations do not address the problem of Campbell's Law. The challenge to the measurement community is to make the use of tests for accountability and incentives its core concern.

### **For More Information**

For more information on this subject, please see the paper by Dr. Koretz:

Koretz, D. (2010). *Some implications of current policy for educational measurement*. Retrieved from <http://www.k12center.org/publications.html>.