



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

RESPONSES TO SESSION 3: SOME IMPLICATIONS OF CURRENT POLICY FOR EDUCATIONAL MEASUREMENT

This policy brief is based on reactions of Robert Linn (University of Colorado, Boulder) and Drew Gitomer (Educational Testing Service) to the Session 3 presentation at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, December 2009. Download a copy of the final papers written by Dr. Linn and Dr. Gitomer, as well as the other papers presented at the seminar, at <http://www.k12center.org/publications.html>.

ROBERT LINN, CRESST; UNIVERSITY OF COLORADO, BOULDER

Before No Child Left Behind (NCLB), there were only modest levels of accountability for schools at both state and national levels, and that accountability was mostly carried out through the publication of results. NCLB created sharp increases in the use of sanctions, and it seems that Race to the Top will promote even greater use of sanctions in the form of individual teacher accountability.

Dan Koretz has studied the issue of score inflation more than anyone in the field and has provided convincing evidence that score inflation is a major problem that undermines the validity of inferences from assessment results. He argues that the effects are largely negative, and he distrusts gains.

Gains shown on state assessments are exaggerated due to score inflation, but the magnitude of the score inflation is largely unknown. An analysis by the Center on Education Policy of assessment results from all states with sufficient data over several years, found that increases in results were much more common than decreases, since NCLB. That does not prove, however, that NCLB improved achievement. While there were generally positive gains on state assessments and the National Assessment of Educational Progress (NAEP). Although gains on NAEP were generally smaller than gains on state tests, they were substantial, nonetheless, for many states, especially on the NAEP mathematics assessments at the elementary and middle grades.

NAEP has served an audit function for state assessment results, and while there is substantial inflation of state test scores, there also appears to be substantial improvement. State assessments in the future should put an increased emphasis on growth in comparison to status, build audit functions into ongoing assessments, and improve linking designs in order to reduce inflation.

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.

DREW GITOMER, EDUCATIONAL TESTING SERVICE

Assessment reform efforts since the 1980s have been largely a reaction to the concern that traditional tests skew practice because educators attempt to emulate both the surface form of the items and the impoverished view of learning and understanding that the items represent. Tests designed for more complex reasoning and responses, however, were costly and largely abandoned until NCLB accountability strategies renewed the concerns. With federal funding and policies pushing for assessment reforms and more sophisticated technology for automated scoring, some of the traditional drawbacks may be reduced.

Dan Koretz asserts that we need assessments that more fully represent what we want students to learn, but he forcefully notes that simply changing the nature of assessment tasks will not halt the undesirable behaviors caused by tests that are a limited sample of desired outcomes. In fact, the potential to focus instruction on particular items rather than the underlying construct might be greater when there are fewer assessment tasks that are easier to memorize.

Koretz also argues that while tests belong in an accountability system, they can never be *the* accountability system. He raises the following key points in that regard:

- *The importance of mediating variables.* An accountability system that does not pay attention to how results are produced is ripe for distortion. If scores dramatically increase in a school, district, or state, is there any rationale that would explain the growth? If nothing changed, then it is likely that the increased achievement is merely a chimera. There are a substantial number of low-stakes research studies looking at mediating variables such as characteristics of teacher practice, student assignments, and measures of teacher knowledge. Such mediating variables have the potential to ensure that the accountability system can make better sense of changes in test scores and to give information that supports improvements.
- *The need for multiple sources of information that can be synthesized through processes of disciplined judgment.* Educational goals are complex and multi-dimensional and cannot adequately be reduced to a single test score. Judgments also must be open to inspection. Koretz recommends auditing, which means judgments would be made at the local level. The state or another monitoring entity, however, would need to make checks on the quality of local decision-making.

Taken together, Koretz's presentation provokes us to think about a different kind of accountability system in which (a) tests would be more ambitious and represent a full range of student outcomes; (b) mediating variables would be used to confirm changes in achievement and suggest ways to improve the system; (c) a broader range of measures would be included; and (d) judgmental processes and their external audits would integrate multiple sources of information.

In almost every other aspect of our lives—admission to school, a job evaluation, or judging the quality of a restaurant or the performance of a company—we make complex, multi-faceted judgments subject to mediating factors. Students and schools deserve that same deliberation.

For More Information

For more information on this subject, please see the papers by Dr. Gitomer, Dr. Koretz, and Dr. Linn:

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

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

Linn, R. (2010). *Test score inflation: Comments on “Some implications of current policy for educational measurement” by Daniel Koretz*. Retrieved from <http://www.k12center.org/publications.html>.