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

ISSUES IN MEASURING STUDENT GROWTH AND CONDUCTING PRODUCTIVITY ANALYSES

This policy brief is based on a presentation by Henry Braun (Boston College) 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. Braun, as well as the other papers presented at the seminar, at http://www.k12center.org/publications.html.

The Common Core Standards and advances in research on student learning progressions help make a new generation of assessments possible. It will be a challenge, however, to get more sophisticated thinking around using student growth as a measure of productivity other than percent proficient.

The financial crisis in state budgets at the same time as the federal level is offering huge investments through Race to the Top funding creates a real opportunity to make critical changes. Most states realize this is the time to improve instruction and school leadership. They will have common core standards that are comprehensive, focused, and multigrade. They will have money for the next generation of assessments that are to inform both instruction and accountability and be developed by state consortia. The strategy should be implemented in stages but also should be radical, exemplifying new policies of collaboration across states and among the different disciplines, new patterns of assessment design, superior measurement properties, and use of new technology platforms.

Assessment Design

Before dealing with growth and productivity issues in new assessment designs, there are some prerequisites to consider. We need a comprehensive model in each domain and models of student learning that lead to expertise in the domain. Learning progressions may provide these qualities because they are based on a research-intensive study of what it means to master a domain and the different ways students can represent and communicate their knowledge in that domain. The research is providing an incredibly rich foundation. We also need high-quality content and performance standards that are vertically articulated. We do not have good ones.

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.

Innovations in assessment design would include the following:

- Integration of cognitive and developmental perspectives with traditional psychometric and logistical requirements.
- Explicit targeting of a growth construct. Right now, the test scores and the scales that we produce are very limited in terms of meaningful and substantive interpretations we can give to the scores. Differences between scores are even less amenable to interpretation.
- Change from *drop-from-the-sky* one-time assessments to assessment systems that include interim probes for diagnostic purposes, curriculum-embedded exercises that are on-demand and can be audited and summative on-demand assessments with multiple formats that use the advantages of technology.
- Technology platforms that enable new formats and test structures, improved accuracy, more uniform precision in psychometrics and scoring, and faster turnaround for both formative and summative assessments.

In designing assessments, one of the unsolved issues is how to balance goals and constraints. For example, what are the tradeoffs in having improved cross-grade articulation and better measures of growth? The latter goal assumes that we know the pathways to mastery (some domains are well along on this), but we do not yet know how to accommodate the fact that different people grow in different ways. How do we build that into an assessment that is fair, relevant, and comparable? We also need answers to how to use growth as a summary indicator of student learning and how to validate measures of growth.

Productivity Issues

Education systems need to answer three questions: Where are students in relation to targets? How much learning took place? What was the relative contribution of *X* teacher or school to that learning? The methodology used to evaluate productivity should match both the questions and the type/quality of data available. Despite their technical flaws, simple indicators always will be popular. Using the percent of students exceeding a proficiency standard, for example, can be misinterpreted, but as long as we are in a standards-based world, this measurement will stay around. The antidote is to create standards that cannot be misinterpreted. If grade-level standards are linked explicitly and appropriately to learning progressions (trajectories), then the longitudinal tracking of grade-level results may generate useful data for productivity analyses. It would still be insufficient, however. We need to have interpretations that also tell us if every student who is proficient in Grade 4, for example, should be proficient in Grade 5. We need good cross-grade coherence.

Value-Added Modeling

With all the unanswered questions about measuring student growth and productivity, what are our alternatives? Is value-added modeling a worthwhile advance? What are the key issues in its use?

- The intent of value-added is to extract, from measures of student learning, the average component due to the teacher or unit (school, program, district).
- It is a step beyond tracking growth because it involves adjustments for selection bias.
- Aggregation of individual-level data places greater burdens on test design.
- Results usually are defined as norms.
- For accountability, statistical descriptions are interpreted as causal effects; this is problematic without randomized experiments.
- Of several models of value-added, those that directly measure growth are statistically weak. The Sanders model (multivariate longitudinal mixed effects) has the best statistical properties.

In the next generation of assessment systems, we should be better at supporting growth-related productivity analyses and know how any trade-offs will affect value-added models or whatever indicators we use to judge system performance. Whatever we do should be as transparent as possible for the people using the results.

For More Information

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

Braun, H. (2010). *Issues in measuring student growth and conducting productivity analyses***.** Retrieved from http://www.k12center.org/publications.html.