



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

## **LESSONS LEARNED FROM PAST ASSESSMENT SYSTEM STRATEGIES AND PROMISING TRENDS IN MEASURING STUDENT GROWTH AND INFORMING INSTRUCTION**

**Pascal Forgione, Jr., and Nancy Doorey**

**Center for K – 12 Assessment & Performance Management at Educational Testing Service**

The federal Race to the Top strategy with its emphasis on assessments aligned to the state-sponsored Common Core Standards offers a profound opportunity for designing the next generation of assessment systems. An added benefit is the chance for a reality check. Before states make major changes in their assessment systems, the research community has much to share with the policy community about its best thinking on assessment issues. The Exploratory Seminar, the initial effort of the new Center for K-12 Assessment and Performance Management (K-12 Center) at Educational Testing Service (ETS), captured major contributions of research to two priorities for policymaking on assessments—measuring student growth and informing instruction. Presentations by leading educational measurement experts were divided into four sessions at the seminar:

1. Research on Measuring Student Growth
2. Research on Using Student Growth Data for Productivity Analyses
3. Policy Issues in the Measurement of Student Gains
4. Research on Informing Instruction

### **Session 1: Research on Measuring Student Growth**

**Presentation by Damian W. Betebenner and Robert L. Linn**

Research needs to answer important questions from policymakers including how much growth is achieved, how much growth is typical (magnitude of student growth), and how longitudinal data can be used. Performance standards, vertical scales, and learning progressions are primary approaches to calculating growth magnitudes, but they all need to be embedded in norms in order to have a framework for interpreting student growth. Longitudinal analyses of student growth should address different questions from policymakers, such as whether student growth is on target and/or

*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.*

whether subgroups of students are making as much progress as other students. Longitudinal analyses could use gain scores, cross tabulations (value-added), and regression-based approaches. Whatever techniques are used, they are useful only to the extent that they answer questions people want answered about student growth.

Growth measures are an improvement over the status measures used by the No Child Left Behind Act, but accountability mandates have altered the results from simply being descriptive in nature to use in determining responsibility for student growth. The teacher effect discussion is an example of how value-added measurement has turned into causality measurement, which may be popular but has technical challenges. Policies need to reject the blame game and focus on marshalling a consensus for change, starting with a growth model that delineates up front the questions about student growth that need to be answered.

### **Presentation by James W. Pellegrino**

Research by the learning sciences in recent years has contributed significantly to the development of new approaches to measuring student growth. The research has focused on the most important aspects of knowledge and understanding that should be assessed. Learning progressions are an example of how the research has led to new approaches in measurement. Learning progressions are able to successfully describe more sophisticated ways in which students come to think about the key concepts and practices in each discipline as they move through the grades. The current challenge is that the policy community wants immediate and definitive answers, but the learning sciences community has only begun to explore the answers. The two communities need to meet in the middle and proceed with assessment designs that start with the most defensible approaches to obtaining the evidence we seek about student learning rather than the data we can collect easily.

### **Responses to Session 1 by Wendy M. Yen and Mike Kane**

*Yen:* Revised accountability systems need to use multiple measures and refrain from overpromising.

*Kane:* Simply asking, How much growth? does not account for the fact that growth is multi-dimensional, and a complete enough set of variables is not yet available to support value-added measures of how much growth .

## **Session 2: Research on Using Student Growth Data for Productivity Analyses**

### **Presentation by Edward Haertel**

Getting to the point where growth measures can be used to measure the productivity of groups depends on correctly defining growth for individual students. That definition, however, is influenced by contexts that differentiate between what teachers/schools are responsible for and what lies beyond their control. Whether measuring groups (programs) across multiple districts/schools/classrooms or evaluating large units, comparisons are a useful strategy, but summaries have limited use. Growth

measures can be used to study educational productivity as one way of understanding complex problems, but they do not fundamentally change the nature of the problems.

### **Presentation by Henry Braun**

The prerequisites for new assessment designs should include a research base on what it means to master a domain and different ways students represent their knowledge in the domain, as well as high-quality content and performance standards. The design should move from one-time assessments to a system with different types of assessments resting on a technology platform. The methodology used to evaluate productivity should answer such questions as the relative contribution of teachers/schools to student growth. Simple indicators always are popular but can be misinterpreted. Learning projections can yield useful data in this area, as they can be used to provide cross-grade trajectories and coherence.

### **Presentation by Robert H. Meyer**

Using value-added data to measure productivity can support the point that all students can learn at higher levels, but even a very simple model raises a number of questions that need to be answered. For example, can these productivity models produce value-added numbers that increase as districts/states show higher performance, or will the models need to be re-normed around the average? Do some cohorts of students have more variability than others? Data is beginning to be used in powerful ways, but users of data need to push back and figure out how to improve assessments on a continual basis.

### **Response to Session 2 by Stephen Lazer**

Getting better growth models depends as much on being able to articulate growth targets, development, and expectations across grades as it does on the statistics. Even if a system is well-designed for one purpose, it should not be expected to serve all other purposes.

## **Session 3: Policy Issues in the Measurement of Student Gains**

### **Presentation by Daniel Koretz**

The federal Race to the Top funding intensifies the use of testing for high-stakes accountability, even though there is a lack of evidence of positive effects of such testing. And there is persuasive evidence of unwanted side effects. Attempts to change either test designs or to use statistical and psychometric innovations, however, have not addressed central problems with test-based accountability. High-stakes testing eventually becomes a misleading indicator because it corrupts the teaching/learning process. Growth modeling and using more complex performance tasks may exacerbate the problems. Local auditing may help, but the priority for the measurement community is to make these challenges to assessment design and accountability its core concern.

## **Responses to Session 3 by Drew Gitomer and Robert L. Linn**

*Gitomer:* The problems with test-based accountability raise several questions, including how results are produced and the influence of mediating variables such as characteristics of teacher practice and student assignments. Addressing these issues calls for multiple sources of information beyond a single test score.

*Linn:* While there has been some improvement in state test scores, this might be attributable to inflation of those scores. In the future, state assessments should emphasize growth over status, build audit functions into assessments, and improve equating to reduce score inflation.

## **Session 4: Research on Informing Instruction**

### **Presentation by Margaret Heritage**

The current support for redesigning assessment systems is an opportunity to understand how assessments can make teachers more effective in classrooms. Instead of an emphasis on what students have done, new assessments should identify the next phase of a learner’s development—those skills and concepts that are within reach for a student through instruction. Assessments should indicate this zone of nearest development (ZND). The basis of these new assessments would be the determination of the students’ zone of proximal development (ZPD), which characterizes development prospectively. Teachers need to learn what it means for a student to be competent in a domain and how to integrate assessment into instruction in the ZPD.

### **Presentation by Mark Wilson**

Assessment systems should be built first upon improved standards, followed by development of classroom assessments and then large-scale assessments, instead of the current practice of moving from standards directly to summative assessments. In place of the current unmanageable standards, good instructional assessments would encourage a long-term view of student growth that connects standards to meaningful learning progressions. Concrete examples are needed to then make these standards clear to teachers. Learning progressions could use systems like the Berkeley Evaluation and Assessment Research (BEAR) Assessment System that combine what is known about learning with knowledge about how items become more difficult.

### **Presentation by Lauren Resnick**

The window of opportunity to change assessment policies across the country will not be open for long, so it is incumbent upon the research community to recommend new directions that can proceed immediately. The new approaches should link standards, curriculum, and assessment at the classroom level and use end-of-unit and end-of-course assessments to build toward an accountability system. Assessments that support instruction would model desired performance (teachers will always teach to the test), evaluate instructional successes/failures at the end (of unit or course), are embedded inside the instructional units and can diagnose individual student needs. Ultimately, however, the

measurement community cannot rely on longitudinal research to build new assessment systems. Federal policymakers mostly are interested in the best available assessment systems now.

## **Responses to Session 4 by Randy E. Bennett, Cheryl Krehbiel, and Ray Pecheone**

*Bennett:* The research on assessments for instructional use corresponds to developments at ETS that see summative, interim, and formative assessments as components of the same coherent system.

*Krehbiel:* Assessments that inform instruction are greatly needed, especially in urban districts, but how can the measurement community inform the process wisely and immediately?

*Pecheone:* The connection of assessment systems to teaching and learning require a different policy structure, including grounding the system in a standards-based curriculum, involving all stakeholders in the development of the assessments, using assessments that present challenging tasks, involving teachers in scoring and relying on multiple measures in multiple formats.

## **Issues Raised by the Exploratory Seminar Presentations and Responses**

The Exploratory Seminar participants were as cautious about promising too much for the next generation of assessment systems as they were optimistic about potentially fruitful research about better assessments. No summary can do justice to the wealth of ideas that were presented and debated, but certain themes seemed to emerge:

- The mistakes of the past need to be avoided, such as sole reliance on a status model of assessment and on a single test to measure student growth. Current assessment systems are inappropriate for use in high-stakes accountability systems; and there was no confidence in and little attention paid to, for example, using these systems to judge teacher responsibility for student growth results.
- As part of the Race to the Top assessment program, planning will be critical for ongoing system improvement, starting with an evaluation plan for the content standards, the performance standards, the assessment system, and the accountability system to ensure they continually reflect college and career readiness. The plan should include monitoring of the impact of the new accountability system, assessment of the system's consequential validity, and identification of what works in the system and what does not.
- Tests measure complex functions and must be specific to the information wanted from them. For each assessment system developed under Race to the Top, the evidence model needs to be clearly articulated, including the purposes of the system and each component, how the data will be accumulated, and how the data will be used. Individual assessments cannot serve multiple purposes well, as designs that serve one purpose well often serve others poorly.
- A stated purpose of the summative assessments in the Race to the Top reform agenda is the determination of the extent to which each student is on track to college and career readiness.

This will require that (a) the standards are defined such that they form an evidence-based progression of student learning, and (b) assessments (perhaps multiple measures) are designed to measure growth along that progression. Student growth and group productivity exist within contexts that need to be understood and accounted for in any analyses of assessment results.

- In order to develop formative or interim assessments that can inform the next steps in instruction, assessments will need to be based on a specific curriculum. Therefore, the U.S. Department of Education should fund the development of a small number of high-quality, detailed curricula that make sure of learning progressions, upon which instructionally useful interim and formative assessments can be built.
- The federal government should lift the current restriction on grade-level items for summative assessments to allow more accurate and useful measurement of current performance and growth.
- For accountability purposes, both normative growth and status data are needed to balance the goals of ensuring that expectations are both achievable and students are not left behind.
- Summative assessment data that is to be used for teacher and school evaluation should be supplemented with other forms of evidence and used in ways that are appropriate to both the assessment design and mode of analysis.
- Assessments should move to computer-based testing as soon as possible because a wider range of skills can be assessed and the constructs that we need to measure are becoming increasingly more technology-based.
- As technology becomes more deeply embedded within the skills and knowledge required for college and career readiness, the content standards and the assessments will have to change. Because this change will impact the ability to produce longitudinal trends, planning is needed.
- Most importantly, test design and use should begin by determining what policymakers and the public want to ascertain from assessment results. The assessment system is only as good as the answers the system can provide to questions about performance of the education system.

In response to this final point, the K-12 Center, in collaboration with the Education Commission of the States and the Council of the Great City Schools, organized an opportunity for policymakers to become informed about the kinds of choices on assessment systems available to them. It is clear that the current Race to the Top assessment agenda would benefit from analyzing different assessment system models developed by leading measurements and education policy reformers. The Invitational National Conference on Next Generation Assessment Systems (March 2010) provided guidance on competing demands of assessment and accountability and how to incorporate these demands into comprehensive and coherent systems.

## **For More Information**

Copies of the December 2009 Exploratory Seminar papers and commentaries, as well as policy briefs<sup>1</sup> summarizing the presentations and participant discussions, are available for download. Please see the K-12 Center website at <http://www.k12center.org/publications.html>.

---

<sup>1</sup> Anne Lewis prepared the policy briefs with the assistance of Martin Orland and Center staff. We thank Andy Latham, Kim Fryer, Jazzme Blackwell, and Sharon Saldaña for their contributions.