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 2: USE OF STUDENT GROWTH DATA IN PRODUCTIVITY ANALYSES

This policy brief is based on reactions by Steve Lazer (Educational Testing Service) to the Session 2 presentations at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, December 2009. Download copies of papers presented at the seminar at http://www.k12center.org/publications.html.

STEVE LAZER, EDUCATIONAL TESTING SERVICE

In their presentations, Edward Haertel, Henry Braun, and Robert Meyer gave careful consideration to both growth modeling and the use of data from the models in conducting productivity analyses. Their presentations also were a bit humbling. The idea that an index can tell who is doing well and who is doing badly is seductive, but the presenters agreed that reality is far more complex, and there are real dangers in doing the wrong thing. The problem is not that people have little faith in what analysts of data do, but rather that people have too much faith that the numbers can be estimated without error. It is naïve to think we can replace human judgment, but in a system where everyone is awfully busy, it is a good thing to provide tools to help make judgments.

Some Specific Insights From the Presentations

- *Getting better growth measures is not just a statistical problem.* We do need to get the statistics right, but in the end, the statistics will be no better than the underlying thinking about the assessments. If we cannot articulate growth targets, development, and expectations across grades, no amount of statistics and mindless repeating of items across grades will be helpful.

- *We need to tie the data to learning theory*. As Jim Pellegrino pointed out, the cognitive scientists and psychometricians are learning to work together, even though their core principles are not always identical.

- *Units of growth may not be the same at all levels and subject area*s. Because of this at the high school level, the whole end-of-course testing model may simply not work.

- *Other things that we may want to do with assessments may not support growth modeling*. Certain forms of open-ended testing, for example, might be a good way to promote the right kind of learning, but it is questionable whether that would support growth modeling because another source of variability is being added to the mix.

- *Building growth around learning progressions might be a good move if doing so constrained some of the wide open freedom we now have concerning curriculum*. But that might be politically impossible.

- *Getting growth modeling right presumes that the other mechanical things also are done right*. This puts a lot of pressure on components such as equating. We also need to take into consideration that constructs evolve and any indicator system may not be suitable over the long term.

Growth Measures Are Insufficient

Even if the growth measures were perfect, as several of the presentations pointed out they represent only a small part of the challenge. To make productivity judgments, a great deal of data is needed beyond the straightforward information from assessment systems. The decisions about what kind of additional data is needed are tricky and political. Some studies of value-added models, for example, have found that the models are highly sensitive to the control variables used, and do not identify the same high-value teachers in multiple years. This sort of sensitivity to model choices and variability in results is intrinsically worrisome.

We also need to be conscious that the models may have enough noise that we only feel comfortable identifying those at the extremes (either very high or very low performers). We likely know who these are without the new data.

One Size May Not Fit All

As the presentations point out, one size does not fit all. If there is one thing learned from No Child Left Behind, it is that even if we work really hard to optimize a system for some purposes, we should not expect it to serve all other purposes. We have been unsuccessful in persuading people to listen to our statements on the limitations of data. The causal inferences made based on National Assessment of Educational Progress results are a case in point.

A related issue brought up in the presentations is that productivity analyses make little sense without a related theory of action. Not knowing how to use the data to help inform the system is not productive.

How Do We Talk to People About This?

Everyone in the measurement community knows the problems associated with silly uses of the data, and yet we seem collectively nervous to say that out loud to policymakers. How do we explain to people who care about how well our children are doing that these measures are complex and not definitive without sounding like we are standing in the way of progress?

Discussion Among Participants

The participants seemed concerned that common standards were being developed without considering assessment as a parallel issue (one called it “a loony way to go about setting standards”). Several

thought it important that assessments be aligned with curriculum be aligned rather than the mushy standards developed by the states.

The participants were also as critical of teacher capacity to use assessment data as they were of the current limits of assessment systems. However, some noted that teachers have never had a chance, either in preparation or in professional development, to experience what high-quality instruction looks like.

As to the communication issue, participants agreed that the research community needs to point out to policymakers where assessment can make a contribution and where it does not have the answers. Unfortunately, participants believed that the policymaking community wants researchers to have all the answers rather than recognize the need to engage in value-laden decisions.

For More Information

For more information on this subject, please see the papers by Dr. Braun, Dr. Haertel, and Dr. Meyer:[1]

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

Haertel, E. (2010). *Student growth data for productivity indicator systems.* Retrieved from http://www.k12center.org/publications.html.

Meyer, R. H. (in press). *Value-added models and the next generation of assessments.* Princeton, NJ: Educational Testing Service.

[1] When available, Dr. Meyer's paper will be posted at http://www.k12center.org/publications.html.