



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

## STUDENT GROWTH DATA FOR PRODUCTIVITY INDICATOR SYSTEMS

This policy brief is based on a presentation by Edward Haertel (Stanford 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. Haertel, as well as the other papers presented at the seminar, at <http://www.k12center.org/publications.html>.

*Before using growth measures to evaluate programs or units (i.e., schools, classrooms), we need to define growth for both individuals and groups. No method for any of these steps solves all the problems, but the big psychometric problem is to figure out what policymakers want to capture and then find a way to do it.*

Getting to the point where group growth measures can be used to determine the productivity of programs or units (i.e., districts, schools, classrooms) raises innumerable questions about how to define growth for an individual and to define and compare growth for groups. These questions must be answered before evaluating programs or units.

### Step 1: Defining Growth of an Individual

The challenge of defining growth for an individual is to construct a student's score so that it captures some attributes of the student instead of assuming the problem is solved if we just use the difference in achievement between two points in time, such as comparing post-test scores to pre-test scores.

Defining growth brings up many questions:

- What information is an individual growth score supposed to capture?
- Should the year-end expectation be the same for all students? (If so, we don't need growth measures.)
- Should the expectation be the same for all students with the same prior-year score? (If so, we could use student growth percentiles.) Controlling for prior-year scores, however, is not enough because there always will be measurement errors in prior-year achievement, and it precludes using other variables such as additional test scores and

*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.*

- demographic/out-of-school factors. Moreover, students with identical prior-year scores may be on different growth paths due to their own aptitudes, out-of-school support for learning, or prior classroom experience.

The point is that measurement without context is almost useless, no matter how ideal the measurement scale. There is no bright line between what schools/teachers should be accountable for and what lies beyond their control when measuring whether or not students have met expectations.

## Step 2: Defining Growth for a Group

Consider this metaphor: It is easy to calculate the time of an individual runner, but it is harder to do so for a group of runners or to compare one group of runners to another. The default approach to measuring academic growth is to map individuals' scores and then summarize them. We might set different expectations for different groups, but as other research has pointed out, no one number can be relied upon to do a good job of summarizing differences between distributions. It might be better to just go back to measuring mean growth, but we are stuck right now in standards-based measurements.

Measuring group growth is not so much a technical problem as it is a policy one. This is where policymakers often insist on having technical answers to what are value questions. The best answer will vary from case to case depending on the scope, the data available, the level of aggregation, and the stakes involved. If actual change is what policymakers want, a simple gain score might be enough. Student performance relative to expectations, however, is a different problem and calls for a different way of manipulating the numbers. We can move toward a solution by:

- Defining individual growth scores relative to expectations (best possible outcomes)
- Take an unweighted or weighted average of individual growth scores.
- Create an index of group performance to support intended uses and interpretations.

Some guiding points: keep it simple, use continuous measures (avoid cut points and categories such as *basic*), use a single average (avoid multiple criteria), use weighting to influence the allocation of resources (more weight to underserved groups), and separate creating growth scores from creating descriptive/evaluative categories.

## Step 3: Group Growth Measures as Indicators of Productivity

There are two broad categories for use of group growth measures: one, to evaluate programs in multiple districts, schools, or classrooms; and, two, to evaluate large units such as many individual districts, schools, or perhaps classrooms. In either case, comparisons would be used as a strategy. Summaries of group outcomes, however, have limited usefulness. Simple approaches to evaluating the categories include:

- *For Programs:* A valid design would compare outcomes across treatments. No individual covariates should be needed if individual growth has been properly defined. Group-level covariates would be helpful and required if the assignments are not random.

- *For Units:* There are three choices. The arm's-length good job/bad job accountability choice uses rewards/sanctions for incentives to improve and publicizes relative success of different schools to promote school choice/market pressures. This method however does not increase understanding of what works. A second strategy would be to use group outcomes as one factor in allocating resources, but this creates perverse incentives (rewarding failure). The third strategy is to study the outliers, which resembles the old effective schools research. Making retrospective inferences, however, is not good research. People will tell you what they genuinely believe about why their schools succeed or fail, but they may be quite wrong about what factors account for their schools' success or distinguish their schools from others.

Strong program evaluation models using group growth measures would analyze classroom instruction such as students' prior experiences. They would quantify and collect variables that are relevant. These will differ by school type.

In sum, growth measures can be used to study educational productivity as one way of understanding a complex problem. They do not change fundamentally the nature of the problem, however, despite the potential people hold for them. Nor can an indicator system made for one purpose transfer well to other purposes.

### **For More Information**

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

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