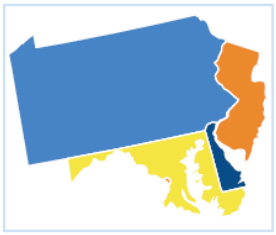
FACTSheet

REL MID-ATLANTIC
Regional Educational Laboratory at Mathematica Policy Research

Serving the education community in Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania

Measuring progress in the classroom: How do different student growth measures compare?



Many educator evaluation systems include growth in student achievement among their measures of performance. Student growth measures aim to describe gains in learning among a group of students, such as those in a teacher's class or a school during a school year, based on how much their test scores changed. These measures can be combined with other educator performance measures, such as scores from classroom observations, to help states, districts, and schools identify the highest- and lowest-performing teachers.[i]

HOW CAN EDUCATION AGENCIES USE STUDENT GROWTH MEASURES?

Education agencies can use the results of student growth measures for several purposes, including helping teachers improve or, in combination with other measures, making consequential personnel decisions about individual teachers. Information gleaned from student growth measures can inform:

Detailed feedback to teachers and principals[ii]

Peer coaching, such as pairing high- and low-performing teachers[iii]

Personnel decisions[iv]

WHAT ARE SOME CONSIDERATIONS WHEN SELECTING A GROWTH MEASURE?

When choosing the best growth measure, education agencies should consider the following factors, which can affect the fairness, accuracy, and usefulness of a growth measure.

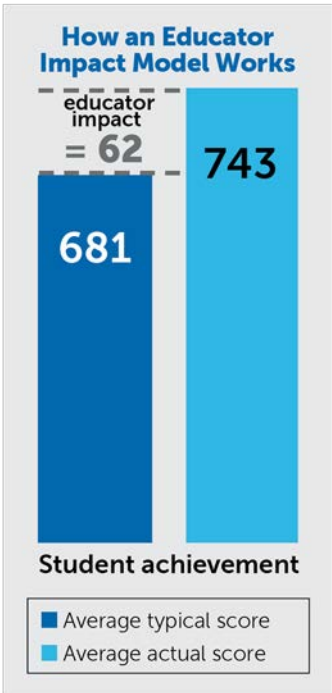
- *Whether the measure has been validated.* A validated student growth measure has been tested and is known to accurately reflect teachers' impacts on student learning and might therefore be a better tool.
- *The approach used to calculate growth and the number of factors outside teachers' control the approach accounts for.* Some approaches use a statistically rigorous process to relate students' prior and current test scores. Some also account for other factors, such as additional student background characteristics, that may produce a more valid and reliable (consistent) measure of the progress students make during the school year. However, statistically rigorous approaches accounting for other factors may be conceptually more challenging for teachers to understand.
- *Whether the measure can be calculated for and used to describe the performance of most teachers.* Growth measures that rely on student test scores from standardized tests can be calculated only for teachers of grades and subjects with the requisite tests.
- *Whether to use multiple growth measures.* Some states and districts use multiple growth measures, such as one for teachers whose students have test scores and another for teachers whose students do not.
- *How the growth measure will be used, such as for developmental or evaluative purposes.* Some growth measures, such as those based on end-of-year test scores, are only available at certain points in the year, potentially limiting their uses.

We want to hear from you! Please contact us!

All growth measures can produce imprecise results when based on few students, which means chance or luck may influence the results. Education agencies should consider combining growth measures with other information, especially for consequential decisions.[i] Additionally, the accuracy and usefulness of a growth measure depends on measuring meaningful outcomes for students, such as a standardized test aligned with the curriculum.

THE MOST COMMONLY USED GROWTH MEASURES AND THEIR TRADE-OFFS

Educator impact models, also called value-added models, use a statistical process to distinguish a teacher's impact on the growth in his or her students' test scores from other factors outside the teacher's control. The models commonly account for prior achievement scores and student background characteristics. Using these factors, the models calculate a typical test score for each student based on the performance of other similar students. The models compare the typical score with the student's actual score and attribute the difference to the educator. For each teacher, the differences are averaged across all of the students in the teacher's classes.



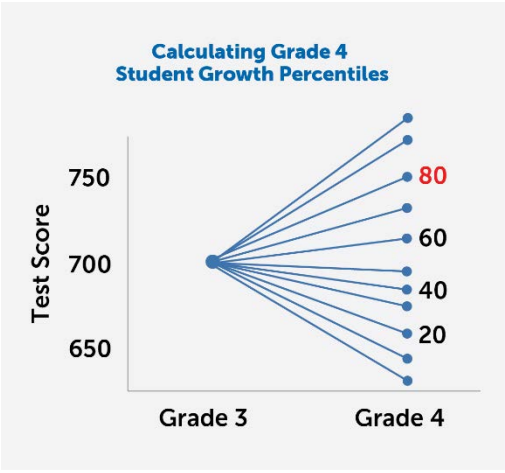
Strengths and limitations

- \+ Have been validated as measures of teachers' impacts on students[v]
- \- Can be calculated only for teachers of grades and subjects with the requisite student test scores
- \- Are statistically complex and can be difficult to report or explain clearly to stakeholders

Student growth percentiles compare the test scores of students with similar prior scores. Using a statistical process, students are ranked according to where they fall in the distribution of students with similar prior scores. For example, a student who scores better than 80 percent of students with similar prior scores is assigned to the 80th percentile. Teachers are assigned a score based on the average or median percentile of their students.

Strengths and limitations

- \+ Typically account for fewer factors outside teachers' control, which may make the approach conceptually easier to understand than educator impact models
- \- Less evidence of validity compared to educator impact models[vi]
- \- Accounting for fewer factors outside of teachers' control could result in less accurate evaluations
- \- Like impact models, student growth percentile models are statistically complex and can be calculated only for teachers of grades and subjects with the requisite student test scores



Student learning objectives measure teachers' progress toward accomplishing goals the teachers set. The goals might be set with input from the teachers' principals or district staff and can be based on a wide range of measures such as state assessments, district assessments, teacher-developed assessments, and non-assessment outcomes like attendance.

Strengths and limitations

- \+ Can be used for all teachers
- \+ Allows teachers to set their own goals, so may be viewed as more connected to instructional improvement
- \- Difficult to implement rigorously and consistently
- \- Difficult to meaningfully compare performance across teachers
- \- Does not use a statistically rigorous process and may not sufficiently account for factors outside of teachers' control[vii]
- \- May not provide a valid or accurate measure of a teacher's contribution to student learning[viii]

REFERENCES

[i] *Kane and Staiger (2012)*
[ii] *Garet et al. (2017)*
[iii] *Papay et al. (2016)*
[iv] *Chiang et al. (2017)*; *Dee and Wyckoff (2015)*; *Glazerman et al. (2013)*; *Walsh and Dotter (2014)*
[v] *Kane and Staiger (2008)*; *Kane et al. (2013)*; *Glazerman et al. (2013)*; *Chetty et al. (2014)*; *Rothstein (2017)*; *Bacher-Hicks et al. (2014)*
[vi] *Walsh and Isenberg (2015)*; *Goldhaber et al. (2014)*
[vii] *McCullough et al. (2015)*; Gill et al. (*2013*, *2014*)
[viii] *Tennessee Department of Education (2012)*; *Proctor et al. (2011)*; *Austin Independent School District (2012)*

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica Policy Research. The content of the fact sheet does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government