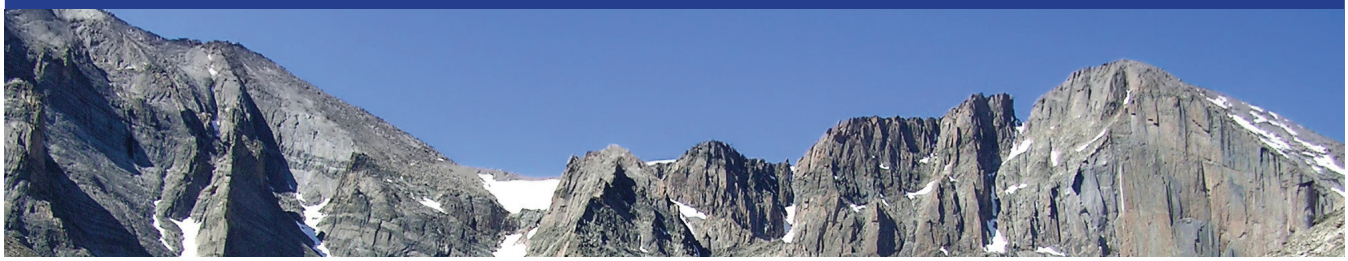
Using Student Growth Percentiles for Educator Evaluations at the Teacher Level:

Key Issues and Technical Considerations for School Districts in Colorado

Elena K. Diaz-Bilello & Derek C. Briggs

JULY 2014



Center for Assessment

CADRE

EXECUTIVE SUMMARY

A common feature of educator evaluation approaches developed over the last five years in states and districts throughout the United States is the combination of teacher practice and student performance measures. In Colorado, Student Growth Percentiles generated by the Colorado Growth Model for mathematics, reading and writing are a key component that can be incorporated into districts' educator evaluation systems. Currently, districts in Colorado may be considering approaches for SGPs for use as part of an educator's evaluation. The mean or median of Student Growth Percentiles, an MGP, is just one of multiple measures that may factor into an educator's overall effectiveness rating. However, there are different ways that MGPs can be used for this purpose, and district decision makers will want to make informed choices to this end.

This brief prepared in collaboration with the National Center for the Improvement of Educational Assessment (NCIEA) and the Center for Assessment, Design, Research and Evaluation (CADRE) at the University of Colorado offers key considerations and insights for school districts in Colorado that are incorporating MGPs into their educator evaluation systems. In particular, two larger topics addressed are: 1) approaches for distinguishing groups of teachers on the basis of their location within a distribution of MGPs; and 2) technical and policy design considerations for using and reporting MGPs.

The mean or median of Student Growth Percentiles, an MGP, is just one of multiple measures that may factor into an educator's overall effectiveness rating. However, there are different ways that MGPs can be used for this purpose, and district decision makers will want to make informed choices to this end.

Categorization of the Teacher MGP Distribution

It is common for "cut-points" to be chosen to establish discrete categories within any distribution of teacher-level MGPs. The purpose of these categories is draw distinctions among teachers whose students have demonstrated qualitatively distinct amounts of growth in achievement, on average. These categories typically range from a minimum of three to a maximum of five levels. States typically use two common approaches to categorize teachers along the MGP distribution:

1. Fixed-cuts approach: Teachers are distinguished by comparing their MGP locations relative to cut-points set in the MGP distribution.

2. Fixed-cuts approach with confidence intervals: Teachers are distinguished by comparing their MGP locations relative to cut-points set in the MGP distribution and confidence intervals.

The selection of a categorization approach includes having to weigh trade-offs that should be considered by school districts. These trade-offs include:

- The first approach offers simplicity of communicating results to teachers, but this approach fails to factor in the influence of measurement error on each teacher's estimated MGP.
- The use of teacher-specific confidence intervals affords more precision in locating a teacher's MGP relative to the first approach, but may pose some communication challenges when teachers with the same observed MGP can have two different ratings on this measure.
- An alternative to forming unique confidence intervals around each teacher's MGP is to apply a common margin of error around the cut-points set in the MGP distribution. Although this may be desirable from a communications standpoint, this approach does not account for differences in classroom size and variability of student performance.

Technical and Policy Design Considerations for Using and Reporting MGPs

Following the selection of a method to categorize teachers within an MGP distribution, there are several technical areas that districts planning to use and report teacher-level MGPs should also explore to inform the design of their systems. These areas are:

- ***Developing inclusion or eligibility rules***

  Determine which student scores should count toward the evaluation of a teacher on this measure. Inclusion or eligibility rules are typically developed for both teachers and students to determine whether a score should count or whether an educator should have scores from a specific course included in his/her evaluation.

- ***Using either the median or mean for teacher-level MGPs***

  Use either the median or mean to aggregate the student growth percentiles at the teacher level. The use of the mean is recommended to districts if confidence intervals are selected as the approach for categorizing teachers. If an approach that only uses cut-points without confidence intervals is being taken, the median will generally be the most defensible option.

- ***Pooling the SGP data across years***

  Consider pooling the data used across multiple years for each teacher to strengthen the stability of year-to-year MGP results reported. This approach also allows for the inclusion of more teachers with scores associated with them in the case where teachers do not meet the minimum n size based on just one year of data.

- ***Combining the MGPs by subject to report an overall growth score on state assessments for a given teacher***

  In the case for teachers with more than one MGP associated with them by subject (e.g., elementary school teachers), choose either a decision matrix or composite index approach to combining the MGPs from the two or three content areas to report a final score or rating for teachers on this measure. The choice is typically a policy decision as neither approach has been shown to be better than the other according to technical criteria.

- ***Setting a minimum number of SGP scores for all teachers***

  Ensure that the decision to set the minimum number of SGP scores for all teachers also factors in any business rules used to pool the SGPs across multiple cohorts (i.e., years) and the decision to classify teachers using a confidence interval.

- ***Adjusting for possible sources of bias in MGPs by factoring in teacher and student characteristics***

  Explore the associations between contextual variables of interest (e.g., percentage of free and reduced price lunch students) and MGP results, and determining the extent to which adjustments may need to be made to the MGPs. A key question stakeholders will need to grapple with is the extent to which they would expect to see an association between contextual variables and teacher MGPs.

- ***Weighting the teacher-level MGPs***

  Gather stakeholder input to determine whether it is reasonable for some students to contribute more than others toward a teacher's MGP as a function of differences in attendance and enrollment in the course of instruction. If this is deemed reasonable it might entail the computation of a weighted MGP for each teacher.

Conclusion

Each of the areas listed above should be reviewed and considered by school districts using and reporting teacher-level MGPs in their educator evaluation system and will require that dedicated staff members are available to conduct analyses to help drive design choices. Although there is no specific order for districts to tackle each area addressed in the brief, we recommend that districts should consider all of these areas in order to demonstrate to their stakeholders that a deliberative and informed process has been followed to design the system.

A final point to emphasize is that the MGPs consist of only one of multiple measures being used to evaluate teachers. Districts will need to consider how best to balance the contribution of data from each measure to inform an overall effectiveness rating. MGPs are most likely to be useful as a basis for evaluating teachers when this information is properly balanced against other information about student performance (e.g., evidence from student learning objectives) and direct observations of teaching practice. Further when one measure does not impose an undue influence on the system, this allows for stakeholders to appreciate how all measures are used together to help inform an overall judgment about teachers. Under this scenario, an MGP, like any other measure in the system serves as one of many sources of information that are evaluated using human judgment about teaching quality in order to make informed personnel decisions. This perspective is also consistent with one of the recommendations made to the state by the State Council for Educator Effectiveness advising on the design and implementation of the state's educator evaluation model. As noted by the State Council (2011), "Data should inform decisions, but human judgment will always be an essential component of [educator] evaluations" (p. 6).

Districts will need to consider how best to balance the contribution of data from each measure to inform an overall effectiveness rating. MGPs are most likely to be useful as a basis for evaluating teachers when this information is properly balanced against other information about student performance (e.g., evidence from student learning objectives) and direct observations of teaching practice.