

Guidance on the Use of Student Test Results in Teacher Evaluation Systems

October 2017

© 2017 NWEA. All rights reserved. No part of this document may be modified or further distributed without written permission from NWEA.

MAP® Growth™ is a registered trademark of NWEA.

Table of Contents

1. Introduction	1
2. Test Selection	1
3. Proficiency vs. Growth	2
4. Alignment of Content Assessed and Content Taught	2
5. Context in Goal Setting	3
6. Conclusion	4

1. Introduction

This document presents issues that should be considered before student test results are used in evaluations of teacher performance. As the stakes attached to student test results have increased, so has the need for educators to make informed decisions about the types of tests used to generate data for these purposes and the way student test results will be used.

Prior to using test results as an evaluation tool, school and district leaders should consider the following questions:

1. **Test Selection:** Is the assessment used to measure student performance being used in the manner for which it was designed? For example, is a proficiency assessment being used to measure proficiency, or is it being used to measure growth? Further, can the test adequately measure performance for all students, regardless of their achievement level? Is the test sensitive enough to measure growth over time?
2. **Proficiency vs. Growth:** Are teachers evaluated based on a student proficiency benchmark or on the amount of growth students show between two test events? What are the implications for students and teachers with each approach?
3. **Alignment of Content Assessed and Content Taught:** Does the content assessed on the test align with the content teachers are required to teach?
4. **Context in Goal Setting:** What is a district's standard for "effective" student performance? What level of student test performance is necessary to determine that an educator is doing an "effective" job?

2. Test Selection

Two key issues should be considered when selecting a test for evaluating teacher performance:

1. Will the assessment be used to measure student performance in the manner for which it was designed? Can the test reliably measure the performance of all students, regardless of their achievement levels?
2. If student test results lack precision based on the test design, how does that affect the validity of educator evaluations based on these test results?

Some states have implemented laws that require student growth on state-administered proficiency tests to serve as a primary component in educator evaluation systems. The rationale behind this practice is that states want teachers to be held accountable for the academic growth their students show over the course of the school year. However, many of the state-administered tests being used for this purpose were not designed to be used as a cross-grade growth measure. Rather, these tests were designed to identify if students have mastered the content necessary to be considered proficient in a given grade and content area. To do this, many of the questions posed on these state tests are all at a similar difficulty level (near the proficiency threshold), meaning that high-achieving students respond to items that are generally too easy for them, and low-achieving students respond to items that are generally too difficult.

Because of this design, estimates of achievement have lower precision for many students, especially students not performing at or near “grade level.”

When student growth is a factor in educator evaluations, districts should consider whether the test being used was designed to measure growth and if the test can provide reliable achievement estimates for all students, regardless of where they stand relative to grade-level proficiency standards. Furthermore, given that no tests of student achievement or growth will perfectly capture a teacher’s work with his or her students, student testing information should be considered in the context of additional information (e.g., classroom observations, peer feedback, student feedback) to gain a more nuanced view of a teacher’s work with his or her students.

3. Proficiency vs. Growth

It is important to pay attention to both student proficiency and student growth, as each measure gives educators important information about the performance of their students. However, in the context of teacher evaluation, focusing on student proficiency instead of student growth can be potentially problematic for both teachers and students.

From a student perspective, an evaluation system focused on improvements in proficiency rates may lead to only a small fraction of students receiving the majority of classroom attention—specifically, students at or near the proficiency threshold who can potentially have the greatest impact on an educator’s evaluation (i.e., the so-called “bubble students”). Conversely, students well above or well below the proficiency threshold may receive less classroom attention, given that their proficiency status likely will not change from one year to the next. This problem is not an issue in an evaluation system focused on growth, as teachers are responsible for positively impacting growth for all students, even students who are well above or well below “grade level.”

Regarding the implications for teachers, a focus on changes in proficiency rates may not fully capture the impact a teacher has had on his or her students. For example, a teacher may have helped her students improve from well below grade level at the start of the year to just below the proficiency threshold for that particular grade. This could represent a significant amount of learning shown by these students, but because they still did not meet or exceed the proficiency threshold, this teacher would not be viewed as positively as a teacher whose students moved from just below the proficiency threshold to just above it. And yet, in a system based on student growth, the work of this teacher with her students would be recognized and rewarded given the large amount of growth her students showed and would likely better reflect the positive contribution this teacher made to student learning in her classroom.

4. Alignment of Content Assessed and Content Taught

If student test results are used in a teacher’s evaluation, the test content should align with the content the teacher is required to teach. For many teachers, this is not an issue. For example, a Grade 4 mathematics teacher will likely teach the content and standards assessed on tests of general mathematics administered to his or her Grade 4 students. However, the same may not be true for a teacher of algebra, geometry, trigonometry, or calculus. These teachers focus their instruction on a specific area of mathematics. As such, the tests used in their evaluations should properly assess student learning in these specific content areas.

This alignment between content taught and content assessed is especially important when districts evaluate all teachers using student test results, even though most teachers in a school or district do not teach in a tested content area. In these situations, teachers of content areas such as music or physical education may have their evaluation based, at least in part, on the performance of their students on end-of-year tests that assess content knowledge in mathematics, reading, or other core content areas. However, music teachers were not hired to teach any of these core content areas—they were hired to teach music. Therefore, establishing an evaluation system where these teachers are evaluated on the reading test results of their students is unreasonable and may cause them to focus on other content areas beyond their primary teaching responsibilities.

Considering these alignment-related issues, it is important to ensure that an assessment used as an evaluation tool appropriately assesses the content a teacher is required to teach. When alignment is not present (i.e., for those teachers in non-tested content areas), district leaders are strongly encouraged to work directly with their teachers to develop student learning objectives that do consider the actual content taught by a teacher, and if appropriate, include the use of grade- or school-level measures of student improvement. If educators can collaborate in determining how test results are used, this should help improve the fairness of evaluations and ensure that student learning remains at the forefront of all educational decisions.

5. Context in Goal Setting

Finally, consider the need for context during the evaluation and goal-setting process. When context is considered, questions like the following examples will be helpful:

- If students grew X points over the course of the year, is X enough?
- Is X above average or below average?
- Is X a reasonable expectation for these students?
- Is X a reasonable expectation for a teacher?

The following contexts should be considered when evaluating student and teacher performance:

- **Historical Context:** What level of performance has been demonstrated by a teacher's students in prior years? Prior performance can be useful in understanding whether the expectations set for teachers and students are realistic and attainable.
- **Similar Student Context:** How do similar students (in terms of prior test performance, race/ethnicity, special education status, free and reduced lunch status, etc.) perform on the tests being used? The performance of similar students can provide context for how a student might improve or grow over the course of the school year.
- **Classroom/School Context:** Do teachers work with a population of students that are low- or high-achieving, or do they work with students for whom it may be more or less difficult to show improvements over the course of the year? These classroom/school differences can inform if and how performance goals for teachers should be adjusted to account for the unique challenges each teacher may face.

- **Goal Context:** Are the performance goals set for students and teachers likely to be attainable, or do the goals represent a goal that would be difficult to attain without making significant improvements? Depending on the purpose of the goal, it is important that teachers understand what is expected of them during the school year, including what changes they may need to make to attain these performance goals.

6. Conclusion

No test of student achievement or growth was designed specifically to be an educator evaluation tool, and student test results alone cannot provide definitive evidence of the impact an educator had on student learning. If the aforementioned issues and questions are not considered, it is likely that educator evaluations based on student test results may misrepresent the quality of a teacher's work with his or her students.

This does not mean that student test results cannot be useful in informing the discussion around a teacher's evaluation. Rather, if test results are used, it is important that they are used cautiously with an understanding of what some of the potential problems may be, along with the understanding that *these results alone do not decide or dictate an educator's final evaluation*. By attending to these issues and collaborating with educators and other stakeholders in the development of evaluation plans, many of the common pitfalls associated with using student test results as a teacher evaluation tool can be mitigated. Furthermore, if the point of these evaluations is to help teachers and students improve, approaching the evaluation process carefully and cautiously should help ensure that summaries of teacher performance accurately capture the contributions teachers make to their students' learning goals.