



## **Guidance for Administering MAP Growth Assessments When Results Are Used for High-Stakes Purposes**

August 2017

© 2017 NWEA. All rights reserved. No part of this document may be modified or further distributed without written permission from NWEA. NWEA and MAP are registered trademarks, and MAP Growth is a trademark of NWEA in the US and in other countries.

Disclaimer: This report is the product of research conducted by NWEA.

NWEA  
121 NW Everett Street  
Portland, OR 97209  
866-654-3246  
<https://www.nwea.org>

## Table of Contents

Section 1: Introduction .....	1
Section 2: Early Termination of Test Events or Retesting.....	1
2.1. Written Policy on Early Termination and Retesting Guidelines.....	1
2.2. Definition of “Substantial” Decline in RIT Score Between Two Test Events .....	2
2.3. Written Rationale for Terminated Tests or Retests .....	2
Section 3: Student Engagement and Test Duration.....	3
3.1. Enforce the Test Duration Policy at All Terms.....	4
3.2. Consider Differences in Test Duration Between the Fall and Spring Administrations ....	4
3.3. Use the Proctor Notifications to Ensure Valid Scores Before a Retest is Required .....	4
3.4. Document the Need for Consistency in Testing Conditions and Test Duration .....	4
Section 4: Proctoring.....	5
Section 5: Additional Considerations.....	5
5.1. Ensuring That All Students Are Testing at All Terms .....	6
5.2. Testing Students at a Similar Point in Each Testing Term to Ensure Accurate Growth Comparisons.....	6
5.3. Testing Students at an Ideal Time to Encourage High Performance .....	6
Section 6: Summary .....	7

## **Section 1: Introduction**

Some schools use results from the MAP® Growth™ interim assessments from NWEA® in a number of high-stakes ways. These include as a component of their teacher evaluation systems; to determine whether a student advances to the next grade; or as an indicator for student readiness for certain programs or interventions (such as special education or gifted and talented programs). Therefore, guidance is needed about how to protect the integrity of the testing process and the test results.

NWEA offers these guidelines based on our current research and experience with school districts using our tests for high-stakes purposes. NWEA conducts regular research in this area, and we may refine or redefine these guidelines as better information becomes available.

## **Section 2: Early Termination of Test Events or Retesting**

It may sometimes be appropriate to terminate a student’s test session or to suggest the retesting of a student after a testing session is complete. A student getting sick, rushing, or guessing during the test are situations where pausing or terminating the test prior to completion may be warranted. In general, taking action during the assessment process to prevent invalid tests is preferable to retesting students after a bad testing experience.

Teachers, principals, and students are under significant pressure to perform well on these high-stakes tests. This pressure could result in situations where students are retested because student scores were lower than what was expected or desired. To avoid this type of retesting practice, NWEA has created the following recommendations, which are further described in the subsequent sections:

1. Establish a written policy on early termination and retesting guidelines that applies at every term.
2. Define what a “substantial” decline in RIT score between two test events entails.
3. Require a written rationale for terminated tests or retests.

### **2.1. Written Policy on Early Termination and Retesting Guidelines**

In some high-stakes testing circumstances, there is risk that educators may retest students to receive a higher score for their classroom or school. To mitigate this risk, prior to the first round of testing, schools or districts should develop a written policy that clearly summarizes guidelines for when a test should be terminated early or a student should be retested. While not every possible scenario could be addressed in this set of guidelines, the general rules and retesting approval process could be defined.

The general principle is that retesting is justified when situations occur that may impact the validity of test results. Some of the situations that may be considered in this written policy include:

- A student becomes ill during the test.
- A student refuses to take or complete the test.
- A student is rushing to complete the test items.
- A student is observed responding without reading the items.
- A student shows a “substantial” decline in score, as defined by the school or district, between the current and previous testing period (see Section 2.2).

If possible, a school or district should consider implementing a system in which a principal, building administrator, or, ideally, an impartial designee reviews all retesting decisions prior to the student retaking the test.

Retesting policies should be applied consistently at every testing term.

## **2.2. Definition of “Substantial” Decline in RIT Score Between Two Test Events**

A large decline in test scores between two administrations can be an indicator of an invalid test. There are circumstances in which schools may consider retesting individual students if they show a substantial drop in test score in relation to the prior term.

NWEA does not define what would be considered a “substantial” decline in RIT score between consecutive test events. However, in general, a decline of greater than 10 RIT points from a prior test event may be indicative of low student effort on the current test, or some other factor that caused the student to score lower than expected. Thus, while it may be reasonable to define “substantial” as any time a student’s RIT score declines by 10 or more RIT points between test events, the definition should be stated by the district at the beginning of the school year and included in the school’s written policy on retesting.

The definition of substantial should be applied at every term. For example, a student whose RIT score dropped by 10 points from the prior spring to the fall should be retested just as a student whose RIT score dropped by 10 points from fall to spring in the same school year would be retested.

## **2.3. Written Rationale for Terminated Tests or Retests**

If a student needs to be retested or if a test event was terminated, the rationale should be documented—in writing—at the time it occurs. The documentation should be collected by the school principal, district-level administrators, or the assessment coordinator of a school or district. This provides school leaders with the ability to track which students were retested and for what reasons.

Documenting instances of early termination and retesting can be useful for two reasons:

1. Protects teachers from accusations of test manipulation if a student’s test performance is questioned
2. Ensures clear transparency and accountability surrounding all retesting decisions

### **Section 3: Student Engagement and Test Duration**

MAP Growth provides a metric on student engagement called total test duration. The time it takes for a student to complete a test can be an overall indicator of whether a student gave appropriate effort during the testing process. Disengaged item responses is another feature being incorporated into MAP Growth. It provides a more nuanced view of student engagement during the testing process, as it shows whether a student was attending to individual items within a test event. The proctor should monitor testing closely and intervene when they see students progressing too rapidly through a test. NWEA's current research indicates that tests completed in less than 10 minutes are unlikely to return an accurate estimate of student performance. The research also suggests that MAP Growth test sessions shorter than 15–20 minutes in duration will likely provide inaccurate estimates of student achievement, although this may not be the case for every student who completes a test quickly.

Conversely, students can take an especially long time to take a MAP Growth test. When this happens, we have found that there is very little additional value obtained from the extra time spent. Typical test durations vary based on the grade and season. Early elementary students in the fall generally average 30 minutes, whereas middle and high school students average a bit over 50 minutes. In the spring, the average times are a few minutes longer.

As a rule of thumb, no more than a few percent of students typically have durations longer than double the averages above. If students take notably longer than this, you should consider how to guide the test durations to more reasonable lengths. For example, you may want to reinforce to students that the MAP Growth assessments are designed to identify their instructional level, so it will ask questions that students may not be able to answer correctly. Coach students to give their best effort, but move on if it is clear that they do not know the answer to the item.

The disengaged item response feature monitors how quickly a student responds to individual items. The test proctor will be notified when a student provides three consecutive disengaged item responses. The proctor should encourage the student to focus and provide their best answer.

Because of this, a district's written policy should include a statement about when students should be retested based on the total amount of time they spend on their test or based on the percentage of disengaged responses. That policy should consider the following guidance, which are further described in the subsequent sections:

1. Enforce the test duration policy at all terms.
2. Consider differences in test duration between the fall and spring test administrations.
3. Use the proctor notifications to ensure valid scores before a retest is required.
4. Document the need for consistency in testing conditions and test duration in a written policy.

### **3.1. Enforce the Test Duration Policy at All Terms**

An abnormally short test duration may result in a score that underestimates a student's actual performance, which may impact the amount of growth shown by the student. Students with underestimated fall scores are likely to show inflated growth in the spring. Students with underestimated spring scores would show lower growth between fall and spring than would normally be expected. An abnormally long duration provides little additional measurement or instructional value and can negatively impact testing schedules and instructional time. Thus, to protect the integrity of the testing process and the accuracy of student testing data, schools or districts should include how many minutes on a test is necessary for the test to be considered valid. NWEA recommends setting a minimum and maximum duration that teachers agree is reasonable and enforcing that standard for every term. These standards should contain enough flexibility to allow for the implementation of accommodations or modifications as directed by a student's Individualized Education Plan (IEP).

### **3.2. Consider Differences in Test Duration Between the Fall and Spring Administrations**

For a growth score to be an accurate measure of student progress, the conditions in which MAP Growth is administered must be consistent every term. If a student's fall test is significantly shorter than the spring test, that suggests that conditions under which students tested were not consistent and may negatively impact the validity of a student's growth score. It is particularly problematic when conditions are different for groups of students. For example, if an entire classroom of students completes the fall tests in significantly shorter time than the spring tests, it calls into question the validity of the growth scores for the entire class. In some high-stakes circumstances, this could suggest that there is an effort to game the results.

Even if students take longer than the predetermined short test duration time in both the fall and spring, significant differences in test duration could still have an impact on the amount of growth a student shows over the course of the year. For example, if a student took 30 minutes to test in the fall but then took 80 minutes to complete the test in the spring, the amount of growth that student shows may be greater than if the student had taken approximately the same amount of time on each test. Therefore, steps should be taken to ensure that students have sufficient time to complete MAP Growth assessments in both the fall and spring.

### **3.3. Use the Proctor Notifications to Ensure Valid Scores Before a Retest is Required**

It is better to ensure that students are engaged with their tests rather than have them complete testing and be required to retest. The proctor notification provides an indicator during testing to notify proctors when they need to intervene with a student. By monitoring these notifications and intervening, engaged testing is more likely to occur, mitigating the need to retest students. Should a test be invalidated due to engagement metrics being too low, the student should be retested following the written guidance included in the district's written retesting policy.

### **3.4. Document the Need for Consistency in Testing Conditions and Test Duration**

Periodically monitor testing condition and duration data, and have conversations with appropriate school and district personnel if inconsistencies are identified.

## Section 4: Proctoring

As the stakes around testing have increased, incidents of systematic cheating on assessments have been discovered and have received extensive coverage by the media.<sup>1</sup> Because of this, it is important to implement testing policies and procedures that prevent cheating and, more importantly, protect teachers and students from unwarranted challenges about their results.

The primary responsibility for good testing conditions lies with the proctor and the teacher. Part of that responsibility includes motivating students to do their best, providing testing conditions conducive to good performance, and actively monitoring testing to prevent problems. NWEA encourages districts using MAP Growth to participate regularly in proctor training to ensure that proctoring practices maintain the integrity of the testing process. Proctoring best practices should include the following steps:

1. Both a teacher and an additional proctor should monitor student testing. A teacher should serve as the primary proctor during testing because he or she is the most aware of the learning needs of his or her students and can likely keep students focused on the testing process better than other instructional personnel.
2. When results from the MAP Growth assessment are used for a high-stakes purpose, it is good practice to also have a second proctor in the room to help oversee the testing process. The second proctor should be someone who does not have direct investment in the performance of the students being tested. In many schools, the testing coordinator could serve as the second proctor. The second proctor protects the integrity of testing results and protects teachers from accusations of cheating.
3. Having a second proctor in the room should also help protect the teachers and students being evaluated. Teachers whose students show strong growth will likely have positive end-of-year evaluations as a result of their students' performance. Because a neutral observer was present during the testing process, it is less likely that the performance of the students (and the performance of the teacher) will be challenged or questioned. Even if it is, the teacher can defend the performance of his or her students because they were monitored by an impartial proctor while they tested.

## Section 5: Additional Considerations

Additional considerations when administering the MAP Growth assessments in high-stakes situations include the following, which are further described in the subsequent sections:

1. If aggregated student test results (especially student growth) are used for high-stakes purposes, schools should ensure that all students are tested at all terms.
2. Students should be tested at a similar point in each testing term to ensure accurate growth comparisons.

---

<sup>1</sup> For example, see [http://usatoday30.usatoday.com/news/education/2011-03-06-school-testing\\_N.htm](http://usatoday30.usatoday.com/news/education/2011-03-06-school-testing_N.htm).

3. Students should test at a time of day that allows them to perform at a high level.

### **5.1. Ensuring That All Students Are Testing at All Terms**

If some students in a group (e.g., class, grade, or school) do not test in the fall or spring (especially students who may not show high levels of growth), then end-of-year summaries of student performance would not accurately reflect how student performance changed for all students in the group over the course of the year. Therefore, schools should make sure that all students are tested in both the fall and spring. If students are not tested, teachers should document the reason why these students did not test.

### **5.2. Testing Students at a Similar Point in Each Testing Term to Ensure Accurate Growth Comparisons**

MAP Growth measures growth by measuring achievement at two different points in time and calculating the difference. For context, growth is compared to the NWEA growth norms for students who tested in the same grade, subject, starting achievement level, and who had the same number of instructional weeks. However, MAP Growth does not use actual instructional weeks for each student. Instead, it uses values that a partner selects as best representing their testing windows. Because the number of actual weeks of instruction impacts how much a student learns, having a student test early in one term and late in another will result in a disconnect between the actual number of instructional weeks a student received and the standard (in terms of the growth norm) against which the student's growth will be compared. This impacts the comparisons made to the calculated normative growth.

For example, assume both the fall and spring test windows are five weeks long. The week selected for growth comparisons is the middle week in both windows (Week 4 and 32, which results in 28 weeks of instruction). If this student tests during the first week of each window or the last week of each window, the interpretation of the student's growth will not be affected, assuming he or she gets 28 weeks of instruction between test events. However, if the student has 24 weeks of instruction because the student tested during the last week of the fall window (Week 6) and the first week of the spring window (Week 30), the interpretation of this student's growth may be significantly impacted if the student's growth is still being compared to the 28-week standard. Therefore, it is recommended that once a testing schedule is established within a school for a testing term, a similar schedule should be used consistently at all subsequent terms. If students will receive more or less than 28 weeks of instruction between their fall and spring test events, the school or district should update their reports to reflect the actual number of instructional weeks that the students will receive between tests.

### **5.3. Testing Students at an Ideal Time to Encourage High Performance**

NWEA recommends that schools administer tests at times during the day when students have sufficient time to complete their tests and when they have optimal concentration (e.g., not right before lunch). Schools should be sensitive to the time of day when students test, and should administer tests at a time when the students are focused and will not have to rush. In general, the earlier students test during the day, the better.

## Section 6: Summary

These recommendations provide school leaders and test administrators with guidance about key issues that should be considered when using MAP Growth test results from NWEA as a factor in high-stakes decisions about schools, educators, or students. These recommendations should also be viewed as important testing practices even if test results are not used for high-stakes purposes. These recommendations will help to improve the overall reliability and validity of student test scores.

In summary, NWEA recommends that schools or districts should strongly consider implementing three broad policies:

1. Schools or districts should develop a written policy at the start of the year that clearly outlines expectations for teachers and students throughout the testing process.
2. These policies should be understood by all teachers prior to the first test administration, allowing teachers the opportunity to seek clarifications about the testing policies, which may be different than when the NWEA assessments were used in a low-stakes capacity.
3. Consistency is the key. These policies should be enforced at all test administration periods and should be the same for all teachers in the school.

These recommendations will help to maintain the integrity of the testing process and should provide teachers with protection and support if their student's test results are made publicly available and subjected to additional scrutiny. Perhaps most importantly, the implementation of these recommendations should ensure that student achievement and growth data are as valid and reliable as possible so that these data can provide valuable information to educators as they continue to help all students learn.