



**Guidance for Administering MAP Growth  
Assessments to Improve Measurement  
Accuracy and Maintain Test Integrity**

October 2018



© 2018 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 .....	2
2.2. Percentage of Disengaged Responses (%DR) on a MAP Growth Assessment .....	3
2.3. “Substantial” Decline in RIT Score Between Two Test Events .....	3
2.4. Written Rationale for Terminated Tests or Retests .....	4
Section 3: Monitoring Test Durations and Maintaining Consistency in Testing Conditions .....	4
3.1. Enforce the Test Duration Policy at All Terms.....	5
3.2. Consider Differences in Test Duration Between the Fall and Spring Administrations ....	6
3.3. Document the Need for Consistency in Testing Conditions and Test Duration .....	6
Section 4: Proctoring.....	6
Section 5: Additional Considerations.....	7
5.1. Ensure That All Students Are Testing at All Terms .....	7
5.2. Test Students at a Similar Point in Each Testing Term to Ensure Accurate Growth Comparisons.....	7
5.3. Test Students at an Ideal Time to Encourage High Performance .....	8
5.4. Other Unusual Testing Practices .....	8
Section 6: Summary .....	9
Appendix A: Sample Sections of Policies and Procedures .....	10
Board Policy Excerpt .....	10
Procedure Excerpt .....	11
Retest Request Form .....	12

## **Section 1: Introduction**

Schools use results from the MAP® Growth™ interim assessments from NWEA® in several ways. Some use the results solely for instructional purposes and school improvement. Some include results 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). Whether the purpose is “low-stakes” or “high-stakes”, all students are impacted by decisions made from their assessment data. Educators have a professional and ethical obligation to assure that student results on assessments reflect their achievement as accurately as possible.

In recent years, schools have placed increasing emphasis on the use of MAP data to measure growth. At times that has led educators to confuse MAP Growth’s function, which is measurement of student achievement, with the goal of producing maximum student growth in the classroom. The goal of administering the MAP Growth assessment should be to provide the most accurate measure of student achievement that we can on every occasion in which the assessment is administered. If we do that, then MAP Growth estimates will also be accurate. When the goal of assessing accurately is confused with maximizing growth, educators may feel pressure to engage in testing practices that compromise the integrity of the assessment to show the largest growth between testing periods.

This guidance document is intended to help educators improve the accuracy of their test results and protect the integrity of the testing process since consistent effort and accurate results benefit all. NWEA conducts regular research in this area, and we may refine these guidelines as better information becomes available.

This document offers guidelines in response to questions from educators about how testing policy and practices may improve the accuracy or integrity of test results. It includes guidance related to termination of tests and retesting and procedures to maintain consistency in testing practices over time.

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

It is sometimes appropriate to terminate a student’s test session or to suggest retesting a student after a testing session is complete. In general, if a student is sick, rushing through the test, or showing lack of engagement it is best to quickly correct the situation if possible. If not possible, terminate the test prior to completion rather than allow the student to complete an assessment that will not produce an accurate result. There are cases however, when retesting may be warranted after a test has been completed.

Teachers, principals, and students may also be under significant pressure to perform well on tests, particularly if they are used for high-stakes purposes. 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**

It is better to ensure that students are engaged with their tests rather than have them complete testing and be required to retest. NWEA’s proctor console provides information and notifications to proctors during testing so they can determine whether to intervene with a student. By monitoring these notifications and intervening, engaged testing is more likely to occur, mitigating the need to retest students.

School policy should establish an expectation that educators emphasize the importance of every assessment and encourage all students to do their best every time. In monitoring assessments, educators and proctors should intervene when they see evidence that:

- A student may be ill or distraught 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.

In these circumstances, educators should first intervene with the student to identify the reason they are not engaged with the assessment and, assuming the student is OK, then encourage the student to try their best. If the student still is not trying on the assessment after the educator’s efforts at encouragement, the test should be terminated prior to completion, so no score is generated. The student should be retested at a time when they are better able to demonstrate their learning.

In terms of retesting, in some high-stakes testing circumstances, there is risk that educators may inappropriately retest students to receive a higher score for their classroom or school. To ensure accurate student data and mitigate this risk, prior to the first round of testing, schools or districts should develop a written policy that establishes guidelines governing when a student should be retested.

The general principle is that retesting is only justified when there is evidence that a completed test is not likely to reflect an accurate measure of the student’s achievement. In general, it should be assumed that if the proctor or educator allowed a test to be completed, no evidence was visible during the test that would have compromised its accuracy. Thus, retesting should only occur when objective evidence is produced that would indicate the validity of the test is at risk.

MAP Growth reports provide three metrics that can be used to inform decisions about retesting:

1. If a student's Percentage of Disengaged Responses (%DR) statistic is between 10% and 30%, then the Estimated Impact of Disengagement on RIT can be used to understand the impact that student guessing had on a test score. A retest can then occur if you find a serious and deflating impact on the student score (see section 2.2).
2. If the student shows a "substantial" decline in score, as defined by the school or district, between the current and prior testing period (see Section 2.3).
3. If the student's current test duration is unusually short (see section 2.4).

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 should be monitored to ensure that retesting policies are applied consistently at every testing term.

## **2.2. Percentage of Disengaged Responses (%DR) on a MAP Growth Assessment**

Percentage of Disengaged Responses (%DR) is a MAP Growth metric that provides a more nuanced view of student engagement during the testing process. A student response is flagged as a disengaged response when the student answers an item in less than 10% of the normal time it takes to respond to the item. The test proctor is notified when a student provides three consecutive disengaged item responses. NWEA reports provide a %DR statistic that shows the percentage of items on a test that the student answered while being disengaged. In general, a student is considered fully engaged if the %DR is below 10% or N/A.

When the %DR is between 10% and 30%, educators should consider the estimated impact of the student's lack of effort on their RIT score, which is also reported in MAP Growth. If you judge the Estimated Impact of Disengagement on RIT for a student to be a serious and deflating impact on the student score, consider retesting the student.

If the %DR is at or above 30%, we are not confident in the accuracy of the score, so retesting should be considered.

## **2.3. "Substantial" Decline in RIT Score Between Two Test Events**

A large decline in test scores between two administrations can be an indicator of test results that do not reflect a student's actual knowledge. 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 formally 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 the most recent prior test may be indicative of low student effort on the current test, or some other factor that caused the student to score lower than expected. 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, whatever definition the school system chooses to use should be included in the district’s written policy on retesting and 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 winter in the same school year would be retested.

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

If a student test event is terminated, the rationale should be documented at the time it occurs. In general, retesting of students should be triggered when evidence of the need is produced by a high disengaged response metric, unusually short test duration, or a drop in score exceeding the district’s defined threshold. Policy should define who would approve retesting based on other circumstances, and all retests should be documented 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: Monitoring Test Durations and Maintaining Consistency in Testing Conditions**

MAP Growth provides a metric on student engagement called Total Test Duration, which indicates the amount of time the student took to complete the assessment. Test duration can be an indicator of whether a student gave appropriate effort during the testing process. NWEA’s current research indicates MAP Growth tests completed in less than 15 to 20 minutes will likely provide inaccurate estimates of student achievement, although this may not be the case for every student who completes a test quickly. As educators face increasing pressure to produce high growth scores for their students, our research has found that test durations have shown unusual increases in some schools. In these settings, spring tests are not only considerably longer than the prior assessments, but they are so long that the extra time provides no improvement in the accuracy of the assessment. This inconsistency in test administration conditions also leads to questions about the validity of growth scores from these assessments

To put student test durations into context, we published a document titled “[Average MAP Growth Test Durations](#).” This document presents tables depicting the average MAP Growth testing durations by content area and grade based on aggregated test durations from student tests during the 2017–2018 school year. The tables are intended to provide educators with

general ranges that show how long students normally take to complete a MAP Growth assessment, as well as the normal differences in test duration across terms. Typical test durations vary based on the grade and season with elementary students taking less time than secondary students, higher achieving students taking longer than average achieving students, tests taken in later seasons taking a few minutes longer than tests taken in earlier ones, except for spring to fall tests where the fall test duration is usually a few minutes less than the prior spring test duration.

In general, the expectation should be that test durations should not substantially differ from the published test duration ranges, and that test durations should be relatively similar across terms. If classrooms of students take notably longer than times show, steps should be taken to work with educators to keep test durations to more reasonable lengths. For example, educators may want to reinforce with 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 and to move on if they do not know the answer to the item. Some ideas for how to respond to shorter test durations include:

1. ensure students understand how the data is used to improve their education
2. focus on whether students have progressed or meet goals they have set
3. stress equal importance in the effort on assessment each and every season

A school district's written test administration 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.

NWEA researchers collected testing policies from partners and used them to create examples that are included as Appendix A to this document for your consideration and use as you develop your own versions.

### **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. The “[Average MAP Growth Testing Durations](#)” document is likely to provide useful information to inform your decisions. 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. 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:

---

<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).

1. 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 (someone in addition to the teacher) 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.

## **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.
3. Students should test at a time of day that allows them to perform at a high level.
4. Be wary of other unusual testing practices.

### **5.1. Ensure 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 enrolled students are tested each season. If students are not tested, teachers should document the reason why these students did not test.

### **5.2. Test 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 tend to produce inflated growth scores, because the actual instructional time between student tests would be greater than the number of weeks used as the norm.

For example, assume both the fall and spring test windows are four 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 a student tests during the first week of each window the interpretation of the student's growth will not be affected, the student would receive 28 weeks of instruction between test events. However, if the student tested during the first week of the fall window (week 1) and the last week of the spring window (week 36), the student's growth score may be inflated because the number of weeks of instruction delivered between tests (35), is considerably greater than the number of weeks used for the normative comparison (28). 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.

### **5.3. Test 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.

### **5.4. Other Unusual Testing Practices**

At times, surprising testing practices are identified because they are outside the bounds of what is expected with routine MAP Growth testing. In general, the more the testing practices align with what is "routine", the more accurate the student results. Some unusual practices found include a high number of pauses during testing, testing occurring on days and times outside normal school hours, and multiple MAP Growth Screening tests administered in close proximity to Growth tests.

## 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 the following:

1. Develop a written policy at the start of the year that clearly outlines expectations for teachers and students throughout the testing process.
2. Socialize all policies with 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. Be consistent – enforce policies 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.

## **Appendix A: Sample Sections of Policies and Procedures**

### **Board Policy Excerpt**

#### **Student Testing and Assessment Program**

The District student assessment program provides information for determining individual student achievement and instructional needs; curriculum and instruction effectiveness; and school performance measured against District, State, and National norms.

The Superintendent or designee shall manage the student assessment program that, at a minimum:

1. Administers the State assessments to all students and/or any other appropriate assessment methods and instruments, including norm and criterion-referenced achievement tests, aptitude tests, proficiency tests, and teacher-developed tests.
2. Informs students of the timelines and procedures applicable to their participation in every state and local assessment.
3. Provides each student's parents/guardians with the results or scores of each State and local assessments and an evaluation of the student's progress.
4. Ensures staff use professional testing practices.

Overall student assessment data on tests required by State law will be aggregated by the District and reported, along with other assessment information, on the District's annual report card.

## Procedure Excerpt

### Test Administration/Re-Testing of Students

Repeatedly retesting students can have many negative impacts. This is in addition to the ongoing dialog around over-testing. We must do our best to protect students as well as the integrity of the testing process. All proctors should try to prevent the need to retest by providing the students with the appropriate information and a stable test environment before testing begins. The following are justification for retesting a student:

- the student's RIT score dropped 10+ RIT points from their last testing event
- the student's NWEA MAP profile reveals that 30%+ assessment items were marked as disengaged
- there was a decrease of 10+ minutes from the student's last testing event
- the student took 20 minutes or less to complete the assessment
- if the student exhibits disabling test anxiety;
- if the student becomes ill during the test; or
- if there is a significant disruption or interruption (e.g. fire alarm, etc.).

Please note that this list is not exhaustive. If possible (i.e. if any of the above behaviors are observed while a student test is in progress), it is preferable to invalidate the student's test through the proctor menu rather than waiting for the student to complete the test and asking them to retest.

**A student cannot be retested without prior approval of an administrator.** In addition to administrator approval, schools are required to collect documentation of the reason for retesting. In such a case, the following information is collected: school, test name, student name and ID, the date and RIT score of the test deemed invalid, and the reason for invalidation. This information should also be sent via email to the District MAP Coordinator.

Retest Request Form

# NWEA Retest Request

Student Name: \_\_\_\_\_ Grade: \_\_\_\_\_

Teacher: \_\_\_\_\_

MAP Growth Test to be Retested: \_\_\_\_\_

- student showed a “substantial” decline in score between the current and previous testing period (more than 10 points);
- student rapidly guessing on more than 30% of test items;
- students testing duration decreased 10+ minutes from prior test event
- student rushed to complete the test items or was observed responding without actually reading the items;
- student refused to take or complete the test or became overly anxious;
- student became ill during the test;
- there was a significant disruption or interruption;
- Other: \_\_\_\_\_

## Required Signatures

Proctor: \_\_\_\_\_

School Administrator: \_\_\_\_\_

Parent: \_\_\_\_\_

*\*Parent signature required if student retests more than once in an academic year\**

Thank you for helping us maintain the integrity of our testing process!