

YCSD Guidance for Administering NWEA MAP/MPG Assessments When Results are Used Student Growth Measures

The state of accountability in K–12 education has shifted, moving from a focus on school-level accountability to an increased emphasis on accountability for teachers and students. As a result, some schools are now using results from the NWEA™ Measures of Academic Progress (MAP) and MAP for Primary Grades (MPG) interim assessments for any or all of the following: as a component of their teacher evaluation system; to determine whether a student advances to the next grade; and/or as an indicator of student readiness for certain programs or interventions (such as special education or gifted and talented programs). With assessments increasingly used for high-stakes purposes, guidance is needed about how best to protect the integrity of the testing process and the reliability and validity of student assessment 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.

Early Termination of Test Events and/or Retesting

There are a number of situations when it may be appropriate to terminate a student’s test session early and have that student retest, or to suggest the retesting of a student after a testing session is complete. If a student gets sick during the test, appears to be intentionally trying to finish as quickly as possible, or appears to be simply guessing on the test, these are all situations where pausing or terminating the test prior to completion would be warranted.

In general, preventing invalid tests is preferable to retesting students after a bad testing experience. If students become ill, or have some form of emergency during the test, it is best to terminate the test prior to its completion. If students are rushing through the assessment or seem to simply be guessing, the teacher and/or proctor should intervene with the student; if the student does not respond, the test should be terminated before completion. Specific guidance is provided below for a variety of situations.

Regardless of the reason, teachers, principals, and even students are under significant pressure to perform well on these high-stakes tests. This pressure could result in situations where students are retested simply because student scores were different than what was expected or desired. To avoid this type of retesting practice, NWEA has created the following recommendations.

Early Termination/Retesting Guidelines that Applies at Every Term

In some high-stakes testing circumstances, there is some risk that educators may retest students simply to advance their self-interest in receiving a higher score for a classroom or school. To mitigate this risk, prior to the first round of testing, we will follow the following recommendations and 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 could be defined, as could the process by which retesting should be approved if the situation falls outside the scenarios included in this policy contact Tim Filipovich or John LaPlante.

The general principle is that retesting is justified when situations occur that may impact the accuracy of test results. Some of the situations that may be considered are included below:

- 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 actually reading those items;
- a student shows a “substantial” decline in score, as defined by the school or district, between the current and previous testing period (see Item 2 below).

YCS D Teachers are required to contact principal, assistant principal or testing coordinator to request the ability to retest a student. Building leadership should then request privilege to retest from Tim Filipovich and John LaPlante who reviews all retesting decisions prior to the student retaking the test. We would like to reduce the instances of retesting students who have produced valid and reliable scores without cause. If a YCS D student needs to be retested, or if a test event was terminated, the rationale for why this occurred should be documented—in writing—at the time it occurs by the teacher and test coordinator. The documentation should be collected by the school administrative staff and forwarded to the Office of Accountability and Assessment, Tim Filipovich. Every NWEA testing term will require the same protocols to be applied consistently throughout the year.

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

- Because the reason for the termination/retest was documented, this process will protect the teacher from accusations of test manipulation in the event that a student’s test performance is questioned.
- It will better assure the integrity of the testing process because there is clear transparency and accountability surrounding all retesting decisions.

What a “Substantial” Decline in RIT Score Between Two Test Events Entails?

A large decline in test scores between two administrations can be an indicator of an invalid test. In testing, there are more factors that can produce a deflated score (a score that is less than the student’s real achievement) than an inflated score. We say that unhappy accidents are more likely than happy accidents in testing. Because of this fact, there are circumstances in which schools may consider retesting students if they show an unusually large drop in a test score in relation to the prior term.

By defining “substantial” in the written guidelines, situations can be avoided where a student is retested simply because one of the student’s test scores, likely his or her end-of-year score, seemed atypical and resulted in a decline in test performance.

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

Test Duration

The number of minutes it takes for a student to complete a test can be an indicator of whether a student gave appropriate effort during the testing process. The proctor and teacher should be monitoring testing closely and intervening when they see students progressing too rapidly through a test. Our 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 or MPG test sessions that are shorter than 15-20 minutes in duration can be associated with inaccurate estimates of performance, though this may not be the case for every student who completes a test quickly.

Conversely, students can also take a long time to take a MAP test. The difference in measured RIT scores due to extra sustained effort will often be within the standard error of measure. When this happens, there is little value obtained from the extra time spent. Typical test durations vary based on the grade and season. In the fall, early elementary students generally average near 30 minutes, whereas middle into 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 these averages, you should reflect on your current practices and consider whether to guide the test durations to more reasonable levels. For example, you may want to reinforce to students that MAP is seeking their instructional level, so it will ask questions to which they do and do not know the correct answer. Coach students to give their best effort, but move on if it is fairly clear that they don’t know the answer to the presented item.

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, teachers should monitor how many minutes on the average a test takes to be considered valid and reliable.

Our advice on the duration would be what teachers agree to be reasonable and enforce that standard for every term. We also suggest that these standards contain enough flexibility to allow for implementation of accommodations or modifications as directed by a student's Individualized Education Plan.

Please be advised, for a growth score to be an accurate measure of student progress, it is essential that the conditions in which MAP was administered be consistent between terms. If a student's fall test is significantly shorter than the spring test, which means that testing conditions were not consistent and that negatively impacts the validity of a student's growth score. It's particularly problematic when conditions are different for groups of students. For example, if an entire classroom of students completes fall tests in significantly shorter time than their spring tests, it calls into question the validity of the growth scores for the entire class. For example, if a student took 30 minutes to test in the fall, but then took 80 minutes to complete the spring test, the amount of growth that student shows is likely going to be greater than if the student had taken approximately the same amount of time on each test. As such, steps should be taken to ensure that students have sufficient time to complete MAP assessments in both the fall and spring, so that observed student growth is an accurate reflection of the amount of learning that occurred over the course of the year.

There is an explicit need for consistency in testing conditions and test duration to validate the integrity of the student testing results. Building leadership should periodically monitor testing conditions and the duration data to ensure the consistency of the testing event.

Proctoring

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 that are conducive to good performance, and actively monitoring testing to prevent problems. We would encourage you to revisit proctor training to ensure that at proctoring practices will help us to maintain the integrity of the testing process for every term.

Proctoring best practices should include the following steps.

- We recommend that both a teacher and an additional proctor monitor student testing.
- There are several reasons why a teacher should serve as the primary proctor during the testing of his or her students. The teacher is the most aware of the learning needs of his or her students, and is likely able to keep his or her students focused on the testing process better than different teachers, aides, or other instructional personnel. Research also suggests that students perform better when their teacher is present during testing.
- When results from the MAP or MPG assessment are used for a high-stakes purpose, it is good practice also to 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 those 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 false accusations of cheating.
- Manipulating the testing process can significantly undermine the accuracy of student test results and negatively influence decisions made based on these results. Having a second proctor in the room protects the integrity of the testing process and all subsequent testing data by reducing the likelihood that test manipulation will occur.
- 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 the performance of their students. 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

Important Considerations

If aggregated student test results (especially student growth) are used for evaluation purposes, we have the responsibility to ensure that all students are tested at all terms.

If some students in a given group (class, grade, school, etc.) do not test in the fall or spring (especially those 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. Because of this, schools should make certain that all students are tested in both the fall and spring terms. If students are not tested, teachers should document the reason why these students did not test.

All students are tested at a similar point in each testing term to ensure accurate growth comparisons.

MAP and MPG measure growth by measuring achievement at two different points in time and calculating the difference. To put the measured growth in context, it is compared to the NWEA growth norms for students who tested in the same grade, subject, starting achievement level, and had the same number of instructional weeks. Because the number of actual weeks of instruction matter in how much a student learns, having a student test early in one term and late in another will cause more or less actual learning between tests to occur. This impacts the comparisons made to the calculated normative growth. When compared to the national growth norms based on the selected weeks, this student's growth percentile declines approximately ten percentile points. 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 term to term. Students who test in reading early in the testing window should continue to test in reading early in subsequent windows.

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

We recommend that schools do not administer tests at times during the day when students have a limited amount of time to complete their tests, or when they may have difficulty concentrating on testing (such as 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 going to be focused and will not have to rush to complete the test. Additionally there should be serious consideration to the testing environment as well quiet with no distractions like an AIR Testing Session. It is good practice to complete assessments on building schedule by grade levels starting with one content area preferably reading.

Summary

We have provided these recommendations and guidelines to give school leaders and test coordinators guidance about key issues that should be considered when using NWEA MAP or MPG test results as a factor in making critical decisions about students, educators, and/or buildings. These recommendations should also be viewed as important testing best practices even if assessment results are not used for significant purposes. These recommendations will, in general, help to improve the overall reliability and validity of student test scores.

The guidance in this document is by no means comprehensive. Rather, our aim is to highlight what is believed to be the major areas that merit consideration by schools or districts using MAP or MPG test results for evaluation purposes. In such instances, increased attention needs to be given to the testing process to ensure that the testing data accurately reflect changes or improvements in student achievement and growth.

In summary, guidance document will help to maintain the integrity of the testing process, and they should provide teachers with protection and support in the event that their student's test results are made publicly available and subjected to additional scrutiny. And perhaps most importantly, the implementation of these adopted protocols should ensure that student achievement and growth data are as accurate as possible so that these data can provide valuable information to educators as they continue to help all students learn.

NWEA has nearly 40 years of experience helping educators accelerate student learning through **computer-based assessment suites, professional development offerings, and research services**. Visit [NWEA.org](https://www.nwea.org) to find out how NWEA can partner with you to help