2020 MAP Growth K-2 Scale Maintenance Research Summary

March 2020

nwea RESEARCH

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

Table of Contents

Contents

Section 1: Introduction

As a part of the 2020 Norms release, NWEA will perform routine scale maintenance on the MAP Growth K-2 test. More specifically, NWEA is updating item calibrations on the MAP Growth K-2 test, which will result in adjusted K-2 scores starting in July 2020. We have also revised our MAP Growth grade-level test transition guidance (K-2 to 2-5) to provide clearer guidelines for our partners to use across this test transition period. Taken together, the scale maintenance effort and updated guidance will result in a significant reduction in score drops when students transition between the MAP Growth K-2 and 2-5 tests. This research summary provides details about the scale maintenance efforts, the subsequent effect on student scores on the MAP Growth K-2 test, and some of the implications for NWEA partners. In the following section, we provide some background on the MAP Growth K-2 and 2-5 tests and highlight the rationale for this scale maintenance effort.

Section 2: Background

To maintain the continuity of the RIT scale across NWEA assessments – and in this case, specifically the MAP Growth K-2 and 2-5 tests -- item calibrations undergo regular reviews, and item calibration adjustments may be necessary from time to time.

By design, the MAP Growth K-2 tests measure standards-aligned content that is appropriate for students in kindergarten to second grade who are pre, emergent, or beginning readers, whereas the MAP Growth 2–5 tests assess content aligned to state standards appropriate for students who are independent readers (or are nearing independence) in second- to fifth-grade. While the K-2 tests include embedded audio supports, the 2-5 tests are designed for students who can read connected text, and by default do not include audio support. The NWEA MAP Growth grade-level test guidance provides suggestions for when students should switch from the K-2 to 2-5 assessments, but the transition point has remained a challenging issue for some partners, specifically as it relates to which test to administer to second grade students.

For example, we are aware of concerns from partners about large drops in RIT scores for students when they transition from the MAP Growth K-2 to the 2-5 test. Notably, these drops in scores are often greater for higher achieving students and for second grade students who test on the K-2 test throughout second grade, irrespective of their reading ability.

There are multiple factors that can contribute to drops in student RIT scores when transitioning from one test to the next. For example, some school and district policies may require that all students within a grade/subject take the same assessment, regardless of the ability of students to read independently. Such policies can result in low-achieving students transitioning to the MAP Growth 2-5 test too early (i.e. when they are unable to read independently) and/or high-achieving students who are independent readers continuing to take the K-2 assessment for too long.[1]

[1] Ideally, districts should make the decision about which test to administer to individual students based on their specific reading ability. We realize this isn't always feasible, but this will ensure that a student's testing experience will be developmentally appropriate and return actionable information about student achievement to educators. Managing this transition based on the needs of individual students will reduce the magnitude of the decline in RIT scores when students transition between assessments.

We also know assessment items may drift, meaning that the calibration of test items (i.e. estimates of difficulty level) become easier or more difficult over time. Thus, potential for drift in item parameters within a pool, or drift in item pools across scales that can lead to scale misalignment, needs constant monitoring. As part of our continual effort to monitor items across the entire item pool to maintain the continuity of the RIT scale across tests, NWEA conducted a systematic review of the item calibrations of the MAP Growth K-2 item pool to ensure that the difficulty of items students receive are well matched to their achievement levels. In the following section, we describe our methodological approach to this systematic review, with particular focus on the transitions between the MAP Growth K-2 and 2-5 assessments.

Section 3: Methodological Overview

Step 1: Predicting MAP Growth K-2 Student Trends

The first step in the scale maintenance process utilized expansive student longitudinal data consisting of student RIT scores for both the MAP Growth K-2 and MAP Growth 2-5 tests. Using this dataset, we estimated students' longitudinal trends between kindergarten to fourth grade, including a sample of students who made the transition to the MAP Growth 2-5 test after first grade and a separate sample of students who made this transition after second grade.

Based on the growth results from MAP Growth 2-5 test-takers, we downward projected scores on the K-2 assessment to be in better alignment with score trends observed in the later grades. An illustration of this downward projection is shown in Figure 1. The solid blue line represents the average trend for students taking MAP Growth 2-5, and the solid green line represents the average trend for students taking MAP Growth K-2. The dashed blue line in this figure represents the predicted downward projection of students' K-2 scores based on the 2-5 scores. This downward projection adjusts the MAP Growth K-2 score trajectories to be largely parallel but below the original MAP Growth K-2 RIT scores.

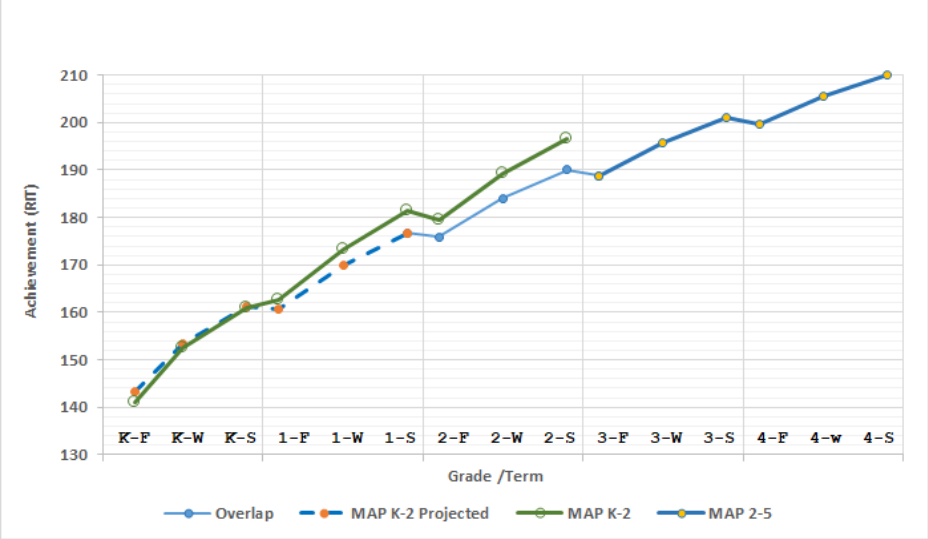

Figure 1. Downward extension of MAP Growth 2-5 growth to growth on MAP Growth K-2

Step 2: Adjusting Calibrations of MAP Growth K-2 Items

After this downward projection was estimated for reading and math, the difficulty estimates of K-2 items were reviewed and adjusted to correspond to the estimated downward trends. As expected, given that the drops observed when transitioning between the K-2 and 2-5 test are greater for higher achieving students, the harder items (i.e. higher RIT) on the MAP Growth K-2 math and reading tests received the largest downward adjustments. In reading, most items were adjusted downwards (with the largest adjustments for more difficult items), though a small number of easier items were adjusted to be more difficult (i.e. have a higher RIT).

Step 3: Review of Subsequent Results

After re-scoring a large sample of MAP Growth K-2 test-takers using the adjusted item RITs, we examined the average impact to student scores by grade, term, and subject. Table 1 presents these results. The score of the average kindergarten student who tested in the fall would be adjusted downwards by 2.81 RIT points in reading but adjusted upwards by 2.83 points in math. By the end of second grade, the average second grader taking the MAP Growth K-2 test in the spring would have their score adjusted downwards by 6.15 RIT points in reading and 6.29 points in math. Because the majority of second grade students should be developmentally ready to take the MAP Growth 2-5 test throughout second grade, only a small percentage of students would be expected to see adjustments of this magnitude if they are assigned to the test most developmentally appropriate for them.

Table 1. Basic descriptive statistics of student RIT scores after re-scoring with adjusted item RIT values

Grade/Term	Math			Reading		
	N	Mean	SD	N	Mean	SD
KF	650,776	2.83	2.37	625,729	-2.81	1.21
KW	644,335	1.15	2.74	628,473	-3.25	1.22
KS	788,983	-0.81	2.65	761,310	-3.86	1.19
1F	916,057	-1.08	2.56	880,687	-4	1.15
1W	804,008	-2.67	2.52	783,351	-4.72	1.5
1S	946,212	-4.17	2.6	913,318	-5.46	1.51
2F	170,408	-3.61	2.63	162,272	-5.14	1.43
2W	134,950	-4.86	2.76	133,043	-5.7	1.54
2S	128,146	-6.29	3.03	126,482	-6.15	1.5

Section 4: Continuity Improvements Resulting from the Scale Maintenance

This scale maintenance approach was validated in multiple ways. For example, a review of the item adjustments by content experts confirmed that the adjusted item RITs maintained an expected ordering relative to the experts' perceived difficulty of the item content. Additionally, we examined common test transition points to understand whether the item RIT adjustments succeeded in improving the continuity of the RIT scale across the MAP Growth K-2 and 2-5 tests.

Figure 2 shows the average change in mathematics RIT scores between spring of first grade to the fall of second grade. Students who remained on MAP Growth K-2 in second grade are shown in teal and students who switched to the MAP Growth 2-5 test in second grade are shown in red. These two groups of students are grouped by their spring achievement level (by decile) at the end of their first-grade year. These results include all students in one large state with testing data across this transition period, which represents a total sample size of over 65,000 students per grade and subject area. In the left panel (prior to score adjustment), students switching from K-2 to 2-5 (red bars) in the fall of second grade showed large drops from spring to fall across the transition, while students who remained on MAP Growth K-2 (teal bars) did not. The large drops are particularly pronounced for higher-achieving students (deciles 8-10). After the adjustment, which is shown in the right panel, the average score drops for the students who transitioned tests between first and second grade (red bars) are greatly reduced.

(a) Pre-adjustment *(b) Post adjustment*

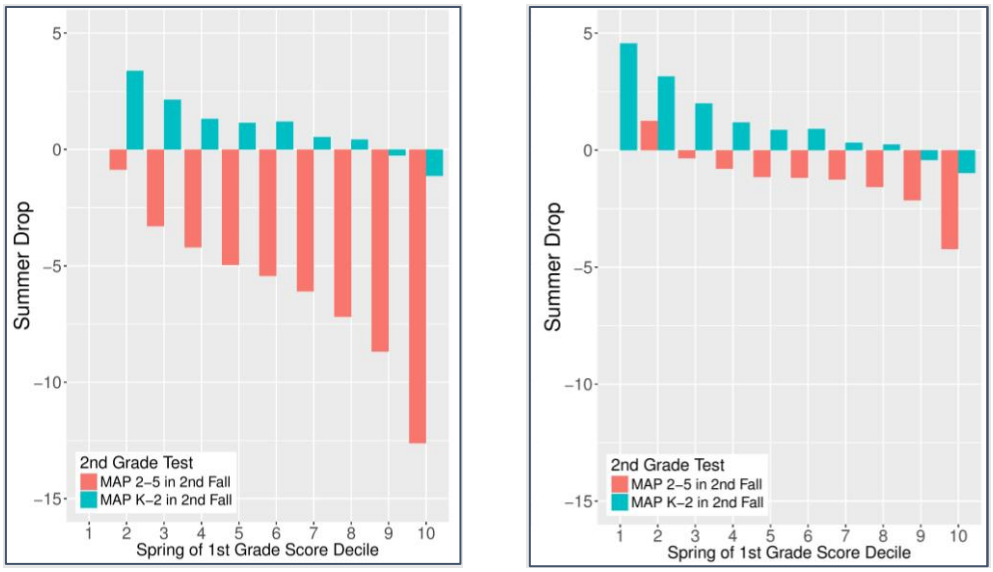

Figure 2. Summer drop (pre- and post-adjustment) for students who tested on math MAP Growth K-2 in 2nd grade (teal boxes) compared to students who switched tests between 1st and 2nd grade (red boxes)

A similar comparison of changes in student RIT scores between spring and fall was made for the other widely used test transition point: between second and third grade. The results represented by the red bars are for students who tested on MAP Growth 2-5 throughout second and third grade, while the teal bars are for students who tested on MAP Growth K-2 in second grade and then switched to MAP Growth 2-5 in third grade. Prior to the adjustment (left panel), the group of students switching from MAP Growth K-2 to MAP Growth 2-5 (teal bars) showed large drops between the spring of second grade and the fall of third grade. After the adjustment (right panel), the test score patterns for these students more closely resemble the patterns for students who remained on the MAP Growth 2-5 test from second grade into third grade (red bars). The observed trends in reading (not shown) are generally consistent with these patterns observed in mathematics.

(a) Pre-adjustment *(b) Post-adjustment*

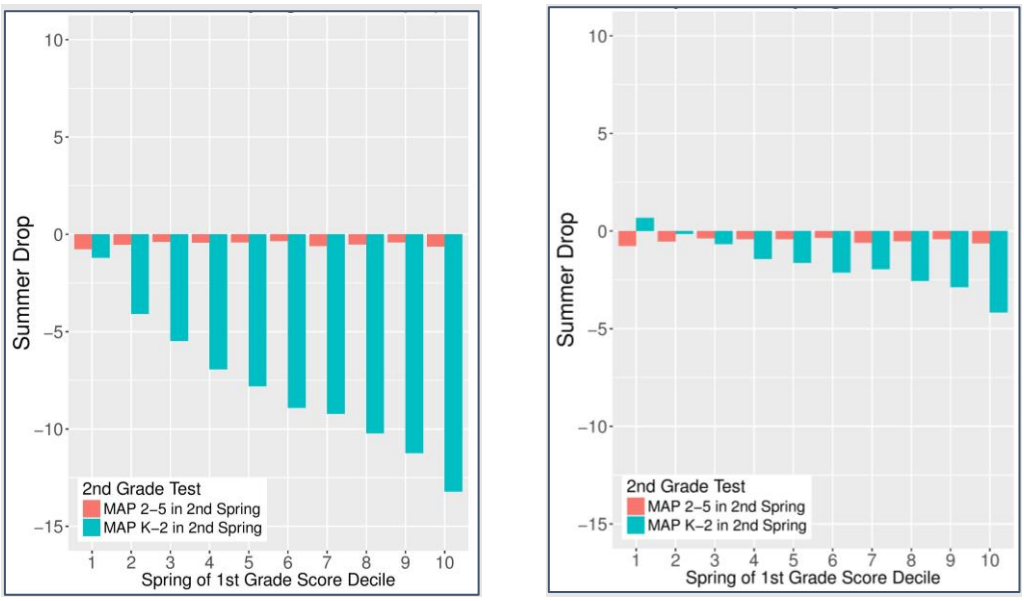

Figure 3. Summer drop (pre- and post-adjustment) for students who tested on the mathematics MAP Growth 2-5 in 2nd and 3rd grade (red bars) compared to students who switched tests between 2nd and 3rd grade (teal bars)

Section 5: Implications and Actions for Partners

The MAP Growth K-2 scale maintenance approach was adopted to improve continuity on the RIT scale between the MAP Growth K-2 to MAP Growth 2-5 tests. While partners may have previously observed notable changes in student RIT scores at this transition point, the result of this new approach has largely remedied this issue. It is important to note that this new approach will not fully eliminate declines in student RIT scores over the summer ("summer loss"), nor does it mean that schools can administer either the MAP Growth K-2 or the MAP Growth 2-5 assessment to their second grade students irrespective of student reading ability. Rather, the implication of this approach is that if students are transitioned to the MAP Growth 2-5 assessment when they are developmentally ready, declines in RIT scores between the two assessments should be reduced in magnitude or mitigated altogether.

The adjustments to item difficulties in the MAP Growth K-2 item pool, as well as adjustment to subsequent student RIT scores, will be implemented as part of the July 2020 MAP Growth software release. From that point forward, all scores students receive on the MAP Growth K-2 test will reflect this new approach. Further, the NWEA 2020 nationally representative student and school norms are also set to be released at the same time, and they will reflect this MAP Growth K-2 scale maintenance adjustment.
While student MAP Growth K-2 RIT scores from the July 2020 release forward will account for the item difficulty and RIT score adjustment, student MAP Growth K-2 RIT scores attained prior to the July 2020 release will not be changed in NWEA reporting systems. **As such, trends in student longitudinal data—especially around the MAP Growth K-2 to 2-5 transition point— should be interpreted with caution, given that the new approach will <u>not</u> be applied retroactively**. The implication of this change is that for some students there may be a decline in student RIT scores when this new approach is applied, since student scores on the MAP Growth K-2 test from July 2020 forward will be adjusted downward on a magnitude

consistent with the results presented in Table 1. The change in student scores across this scale maintenance period is not a reflection of changes in student achievement; rather, this change reflects the difference between the old and new approach being implemented as part of our scale maintenance process. NWEA will provide support to those partners interested in reviewing their historical MAP Growth K-2 data in the initial years prior to the implementation of these changes (pre-July 2020) by making historical, rescored data files available to partners.

Because historical scores (tests taken prior to July 2020) will not be adjusted in the NWEA reporting system, we also strongly encourage partners to abstain from measuring growth for students in first and second grade on the MAP Growth K-2 test across this scale maintenance period (July 2020), especially if student growth is used for high-stakes decisions for students, teachers, or schools. The reason for this recommendation is straightforward – if student growth is measured across years on the MAP Growth K-2 test across this scale maintenance period, then a student's pre-test score will be based on the old methodology, and his or her post-test score will be based on the new methodology. For example, consider a student who was in kindergarten during the 2019-20 school year (pre-adjustment) and a teacher wanted to measure growth for that student from the spring of his kindergarten year to the spring of his first-grade year. If the student tested on MAP Growth K-2 in both kindergarten and first grade, his kindergarten score would be based on the previous methodology, compared to his first-grade score, which would reflect the updated approach. This could result in misinterpretations of the improvement the student made between test events and may suggest that a student made less progress than he actually did. Further, starting in July 2020, the new NWEA norms will also reflect this updated scale maintenance adjustment, and could therefore lead to differences in interpretations of student growth if applied to historical MAP Growth K-2 test scores. For these reasons, we strongly recommend only measuring growth on the MAP Growth K-2 test from fall 2020 forward during the 2020-21 school year. After the 2020-21 school year, we do not anticipate issues measuring growth across years between the MAP Growth K-2 test (at both terms), or between the MAP Growth K-2 and 2-5 tests.

Section 6: Conclusion

Ultimately, the adoption of these improvements will lead to better continuity between the MAP Growth assessments and minimize transition issues partners experience (i.e. score drops) when moving students from the K-2 to 2-5 test. NWEA is committed to supporting its partners during this transition period and will provide resources to help explain the rationale behind this new approach.

For additional information about how to determine which test to administer to 2nd graders, please reference the following resources:

- MAP Growth grade-level test guidance (K-2 to 2-5)
- Psychometric evidence for the MAP Growth grade-level test guidance (K-2 to 2-5)
- Expanded guidance on MAP Growth test administration decisions for second-grade students