

Memorandum

To: Jodi Bossio-Smith, Krista Averill MDOE
From: NWEA
Date: February 16, 2024
Subject: MTYA Winter 2024 Math Administration

The purpose of this memorandum is to present our initial findings in review of anomalies that were identified in Maine SAUs in the most recent winter administration, with specific focus on grades 3 & 4 mathematics.

What occurred?

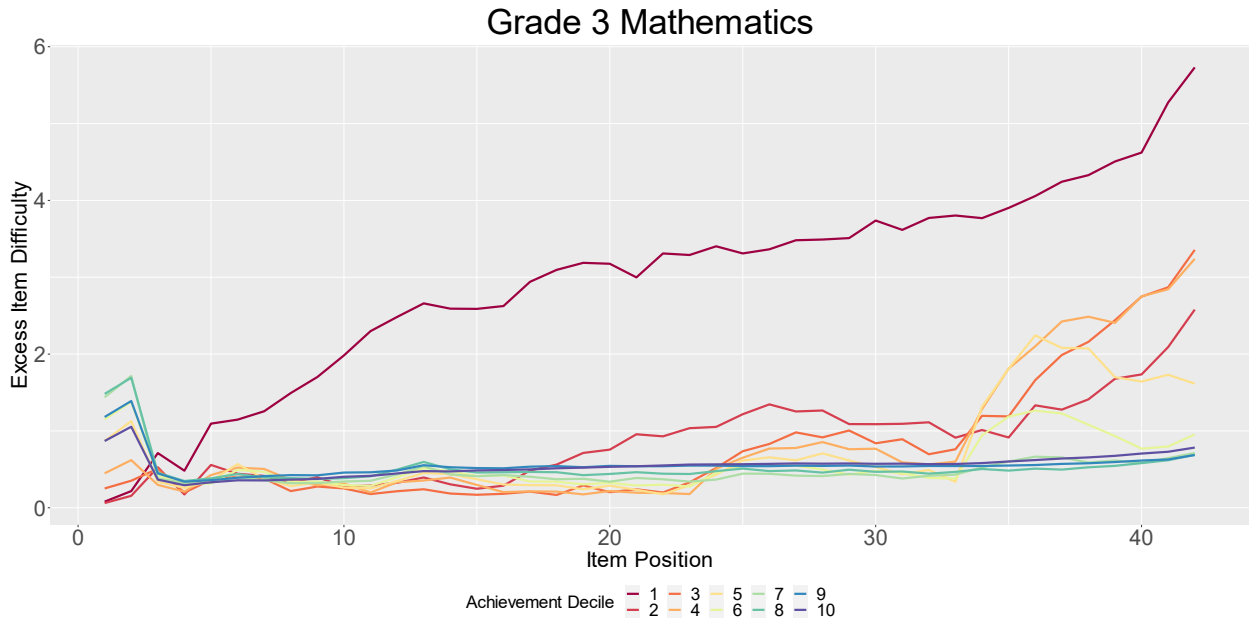
Some educators in Maine questioned the results for specific students after reviewing their RIT scores and the associated student item report from this past winter administration. Most of these queries were related to mathematics, with specific attention to lower grades (e.g., grades 3 & 4).

NWEA conducted multiple analyses to review the test events reported, as well as global analyses across all students. At this time, NWEA has determined that the test engine performed as designed and correctly administered all assessments (i.e., no computer or coding errors). We have determined that the adaptive testing algorithm selected atypical test items for a small group of students – usually in the lowest decile – as they proceeded to the end of the test. These students had unique anomalies that deviated from our expectations. While this system behavior is not unexpected given state-specific item exposure controls and additional blueprint considerations, we recognize the need for deeper explanation.

The constraint-based system chooses items to maximize information and meet other content (e.g., test blueprint) and psychometric requirements (e.g., item exposure control) simultaneously. The maximization of item information is typically accomplished by selecting items with difficulties close to the student's ability estimates at the moment. Toward the end of the test, the item difficulty may not always align with the student's ability as closely. This is because the constraint-based system also aims to fulfill the test blueprint and psychometric requirements after a precise estimate of the student's achievement has already been obtained. The optimal item is one that has content needed by the blueprint or psychometric requirements, not the closest match to student ability. A review of our simulation data revealed that after the first twenty items, item difficulty match began to depart more than expected, especially for students in the lowest decile of achievement. For students in the lowest deciles, the testing engine algorithm begins to select more challenging items than would be desired because some item(s) in the bank are suppressed from administration to this student to control the exposure of the item.

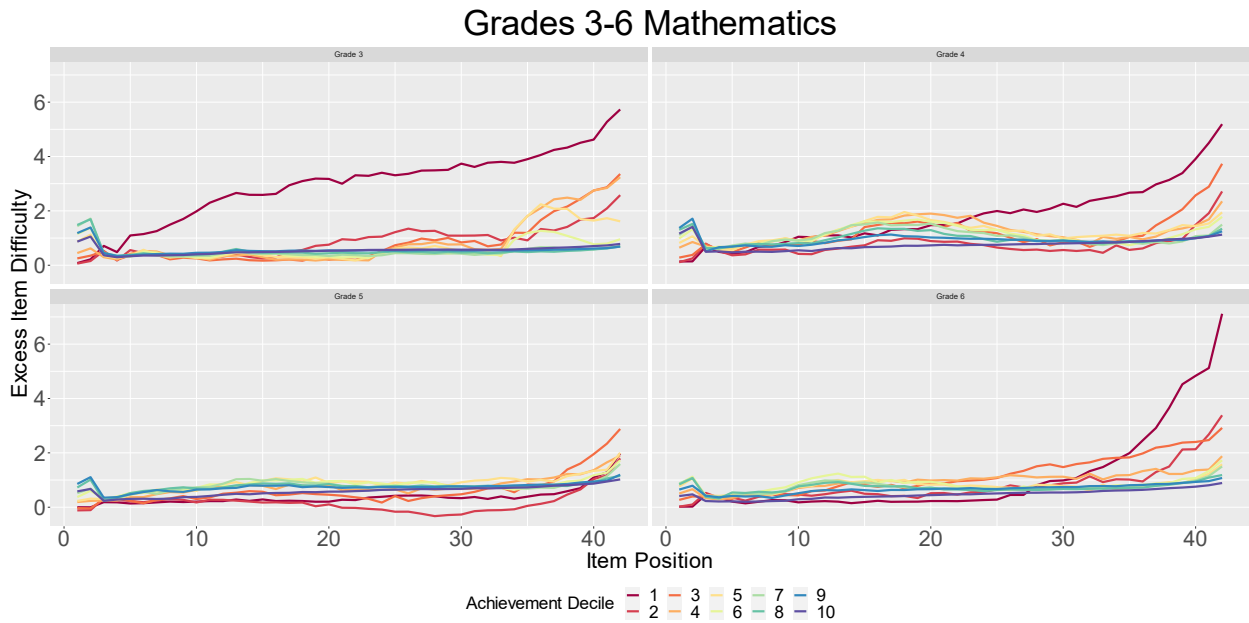
A similar pattern described in the simulation data is also observed in the grades 3 and 4 mathematics operational data for the winter administration in Maine. The grade 3 result is shown in Figure 1. Students in the first decile of student achievement were presented with items that were more challenging than desired throughout the testing event due to item bank suppression.

Figure 1: Grade 3 Mathematics Results



Note. Excess item difficulty is the difference between the difficulty of the item presented to the student and the student's ability estimate at each item position.

Figure 2: Grade 3-6 Mathematics Results



Why did it occur?

Our investigation found that a primary cause of this occurrence is an interaction between exposure rules of the items and the depth of the item pool in mathematics for students in the lowest decile of performance. This item selection behavior does NOT affect the estimation of a student's ability. It may increase the standard error of measurement more than would be expected of a test of the same length.

As much as is possible, the MTYA is designed to emulate the design of the MAP Growth test. In MAP Growth, longitudinal exposure rules govern how often an item can be presented to the same student within a period of time such as 18 months. The MTYA also uses longitudinal exposure rules. A point of difference is how the two systems implement population exposure control. In MAP Growth, a "randomesque" procedure is used, but the MTYA uses a different approach that is configurable. The current configuration of population exposure in the MTYA narrowed the item pool to a point where the content needed to fulfill test blueprints was higher in difficulty for students performing at the lowest decile approaching the end of the test. With this exposure rule in place, the test engine had to select the next best item that met all other requirements but may have been more challenging for the student than would be optimally designed. Accordingly, this guideline led the testing engine to select more challenging items as the student proceeded to the end of the test, as illustrated in the graph above. Please note that this system behavior also occurs on a MAP Growth test where the item pool is shallow, albeit to a lesser extent.

What is NWEA doing regarding this situation?

As the winter test window has completed, NWEA will take the following actions to support winter score interpretations:

1. **Additional score interpretation support** – NWEA can provide additional information to support in the interpretation of the winter scores for this subset of students.
2. **Individual score reviews** – As requested, we have investigated the specific use-cases provided by the Maine DOE, as reported by SAUs. As noted above, we found no effect on the estimation of a student's ability. It may increase the standard error of measurement more than would be expected of a test of the same length. As a result, the student score may be invalidated if it exceeds this standard error of measurement threshold.

How will NWEA mitigate the issue in the future?

NWEA is addressing this situation in two ways for the future. First, we are taking steps to mitigate this issue for the coming spring 2024 administration. These include the following:

- Removing the population item-exposure guidelines referenced above
- Revising our simulation procedures to more closely evaluate test performance among the lowest achievers.

Second, we will conduct a comprehensive review of this condition and propose additional design decision considerations for the 2024-25 school year that will further strengthen the design and item pool to better mirror the design expectations of MAP Growth for these specific student performance levels. These may include considerations such as implementing our Enhanced Item Selection algorithm in the 2024-25 school year.

NWEA is committed to providing actionable data for educators from the assessments we administer. The design goal of the MTYA is to replicate – as closely as possible – the MAP Growth student experience and its scores for students of all ability levels. We remain focused on that goal and are taking steps to address this design limitation in the next and future administrations.