

# The Relationship Between Test-Taking Disengagement and Performance on MAP Growth Retests

By Steven Wise

The Collaborative for Student Growth at NWEA Working Paper series is intended to widely disseminate and make easily accessible the results of researchers' latest findings. The working papers in this series have not undergone peer review or been edited by NWEA. The working papers are widely available, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. Unless otherwise indicated, working papers can be cited without permission of the author so long as the source is clearly referred to as a Collaborative for Student Growth at NWEA working paper.

Wise, S. (2019). *The Relationship Between Test-Taking Disengagement and Performance on MAP Growth Retests*. (The Collaborative for Student Growth at NWEA Working Paper).

## The Relationship Between Test-Taking Disengagement and Performance on MAP Growth Retests

Steven Wise, Senior Research Fellow

April 2019

### Abstract

*Educators sometimes ask: do students rapidly guess because they don't know the answer to a question, or do rapid guesses reflect a lack of engagement with the test? Would a student's scores improve if that student engaged more with the assessment and rapidly guessed on fewer items? Examining MAP® Growth™ test scores and levels of student test engagement for over 100,000 tests for which students retested within one day, the results showed that students' test taking engagement often differed between the initial test and the retest. Students who rapid-guessed less on retest tended to see score gains, sometimes greater than 10 RIT points, with the largest gains seen for students who initially showed the highest levels of rapid guessing. Students who rapid-guessed more on retest, in contrast, generally showed score declines. The findings provide evidence that that rapid guesses reflect a lack of engagement rather than a lack of knowledge, and support NWEA's recommendation that educators should retest disengaged students. The findings also underscore the importance of increasing the student's motivation before retesting to improve engagement and suggest that educators should be cautious if they do not retest students who show high levels of rapid guessing, since these scores pose a threat to the validity of the test score interpretations.*

There has been a growing interest in educational measurement regarding the identification of test-taker disengagement and its effects on the validity of test score interpretations. Researchers at NWEA® have played a leadership role in this by conducting and publishing high-quality research to measure rapid-guessing behavior, understand its impact, and develop methods to reduce its occurrence. Moreover, as an organization NWEA has shown a groundbreaking commitment to addressing this problem by providing information on rapid guessing with MAP® Growth™ score reports, modifying its item selection algorithm to ignore rapid guesses, and by implementing an innovative proctor notification feature to curtail student rapid guessing.

These research-driven activities are all based on a basic assumption that a rapid guess to a MAP Growth item reflects a student's choice to not engage with it (i.e., to not try to solve the item's challenge). This assumption leaves open the question of whether or not the student would have been able to correctly answer the item had they been engaged. In contrast, an alternative explanation sometimes expressed is that students rapid-guess when they encounter items for which they do not

know the answer, and rapidly enter an answer to move on to the next item. That is, students rapid-guess because they don't know how to solve an item's challenge. These two explanations have very different implications. The first implies that the correctness of the student's response is uninformative about their achievement level while the second implies that the response correctness is informative.

Because MAP Growth is an adaptive test, we expect that the items a particular student receives will be answered correctly about 50% of the time. Rapid guesses, however, will typically be correct at a much lower rate (generally around 15 to 30%). Consequently, if rapid guessing reflects disengagement rather than lack of knowledge, rapid guessing will tend to distort RIT scores downward, with the amount of distortion related to the number of rapid guesses that occur. Furthermore, we assume that disengaged students would have received higher RIT scores if they had had been engaged. This is the logical basis for our invalidation criterion—that when students give rapid guesses to 30% or more of test items, their scores are apt to be so distorted that they are too untrustworthy to report.

The idea that disengaged students would have performed better if they had been engaged sounds sensible. However, the sometimes-hypothesized alternative explanation—that students rapid-guess because they quickly recognize they don't know how to solve a problem—implies that their test performance would *not* have been better if rapid guessing had not occurred. That possibility, coupled with the fact that low achievers have been found most likely to rapid-guess, suggests that the invalidation criterion would not be warranted because score distortion had not occurred.

Imagining how a student who exhibited rapid guessing would have performed had they not rapid-guessed is seemingly a speculation about a counterfactual (i.e., something we cannot observe). But our MAP Growth data can provide useful insights about this speculation. Specifically, students taking MAP Growth sometimes are retested soon after an initial test administration, and through study of the engagement and performance of students who were retested within a short period of time, we can gain

valuable information about whether rapid guesses reflect a lack of engagement or a lack of knowledge. To provide this information, the current study was conducted.

**Data Sets Analyzed**

Test events for the MAP Growth Common Core test were extracted in Reading 2-5, Reading 6+, Math 2-5, and Math 6+ from four testing terms: Fall 2017, Winter 2018, Spring 2018, and Fall 2018.

Table 1 shows the numbers of test events in each test type, along with the numbers of retests.

Table 1. Numbers of Test Events and Retests Analyzed

MAP Growth Test	Total Number of Test Events	Number of Retests	Number of Retests Within 1 Day
Reading 2-5	4,595,719	110,326	36,971
Reading 6+	4,324,638	100,839	31,694
Math 2-5	4,864,398	59,251	16,106
Math 6+	4,528,638	58,898	16,938

The data sets contained, for each test event, information about both the student’s test performance (RIT score plus standard error) and test-taking engagement (percent of item responses that were rapid guesses). The data analyses focused on those students who were retested within one day of the initial test event, which effectively ruled out changes in the student’s true achievement level between the two test events. In addition, four levels of test-taking disengagement were formed: serious (30% or more rapid guesses), moderate (11 to 30% rapid guesses), mild (1 to 10% rapid guesses), and zero disengagement (no rapid guesses).

**Results**

Test-taking disengagement was found to frequently differ between the initial test and retest. For each of the 16 disengagement combinations, mean differences between the two sets of RIT scores were calculated. Figures 1-4 show the mean RIT differences for the different combinations. Each of the

four tests showed a highly similar pattern of results, though the magnitude of the differences varied across test types.

Of particular interest is the top row of each figure (blue boxes). These include the students who were seriously rapid-guessing on the initial test. For each of the figures, less rapid guessing on the retest was associated with a higher positive RIT score difference. For example, when there were zero rapid guesses on the retest, the RIT differences tended to be large, averaging around 6 points for the two Reading tests and about 12 points for the two Math tests. Keeping in mind that the typical size of RIT score standard errors is roughly 3.5 points in Reading and 3.0 points in Math, these results clearly suggest that, for seriously disengaged students, decreased rapid guessing on the retest tended to lead to considerable score gains.

Figures 1-4 also reveal a more general trend regarding initial versus retest rapid guessing: regardless of disengagement on the initial test event, students who exhibited less rapid guessing during the retest tended to see score gains, while those who exhibited increased rapid guessing saw score declines. In several instances, the magnitude of these mean gains or declines exceeded 10 RIT points.

### **Why are These Results Important?**

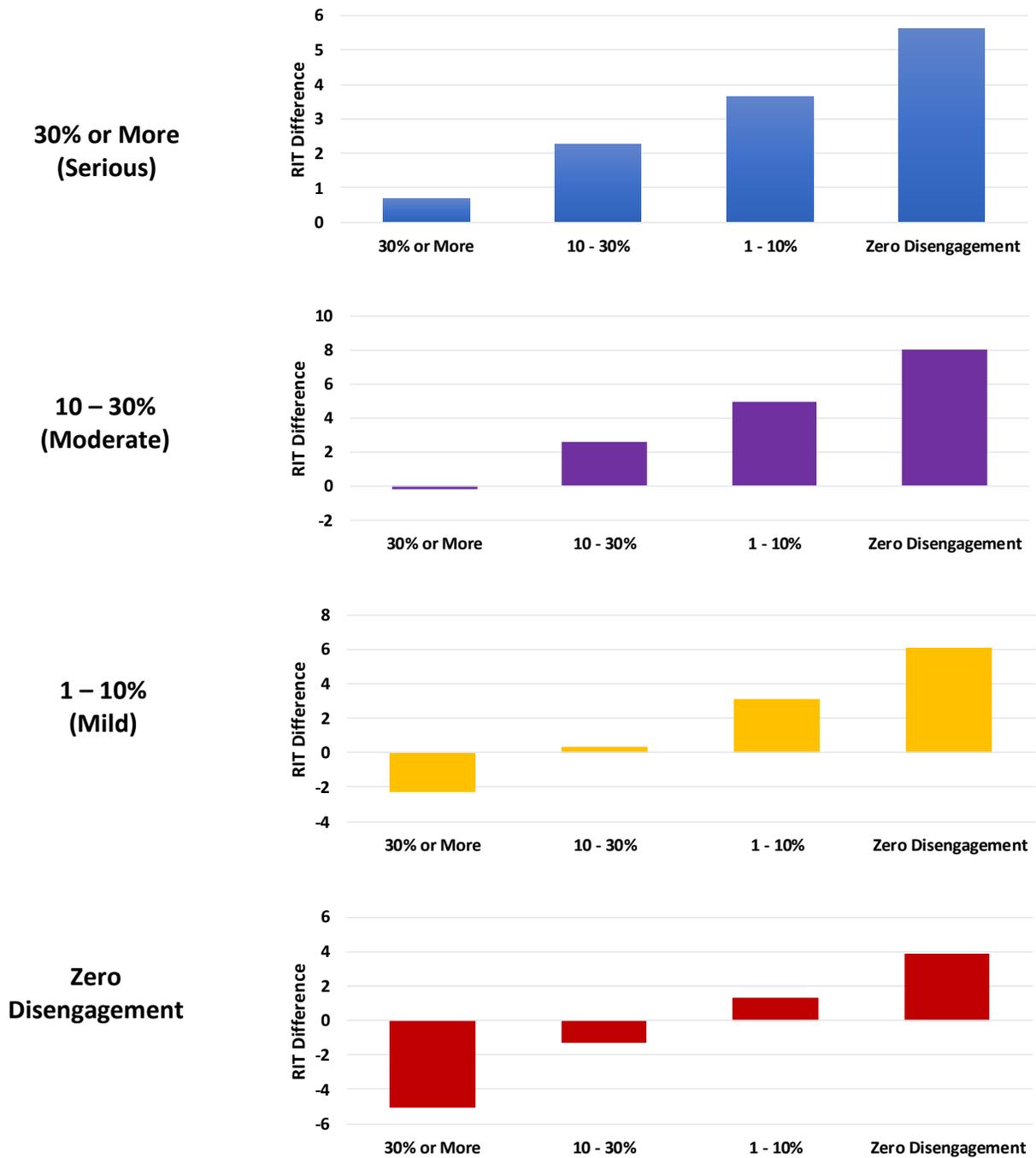
There are two key takeaways from these results. The first is that the score gains associated with lower amounts of retest rapid guessing support our basic assumption that rapid guesses reflect a lack of engagement rather than a lack of knowledge. If they had been due to lack of knowledge, exhibiting less rapid guessing on the retest should not have resulted in improved test performance. Said another way, if you don't know the answer to a set of items, spending more time on them should not yield the magnitude of score improvements observed in this study. These results therefore provide evidence that rapid guessing is an indication of disengagement.

The second takeaway is that if disengaged students can be made more engaged when they retest, they are likely to show sizable gains in test performance. Furthermore, a higher score on a more

engaged retest is logically a more valid indicator of what the student knows and can do. A valid MAP Growth score requires an engaged student, which implies that increased engagement should yield more valid scores. These findings support NWEA's recommendation to educators that they should retest disengaged students. The findings, however, underscore the importance of focusing on increasing the student's motivation before retesting. Simply retesting will accomplish little if engagement does not improve. The research also suggests that educators should be cautious if they do not retest students who show high levels of rapid guessing, since these scores pose a threat to the validity of the test score interpretations. In other words, they likely do not accurately reflect what students may be capable of performing if the students were more engaged in the assessment.

**Percent Disengagement  
on First Test Event**

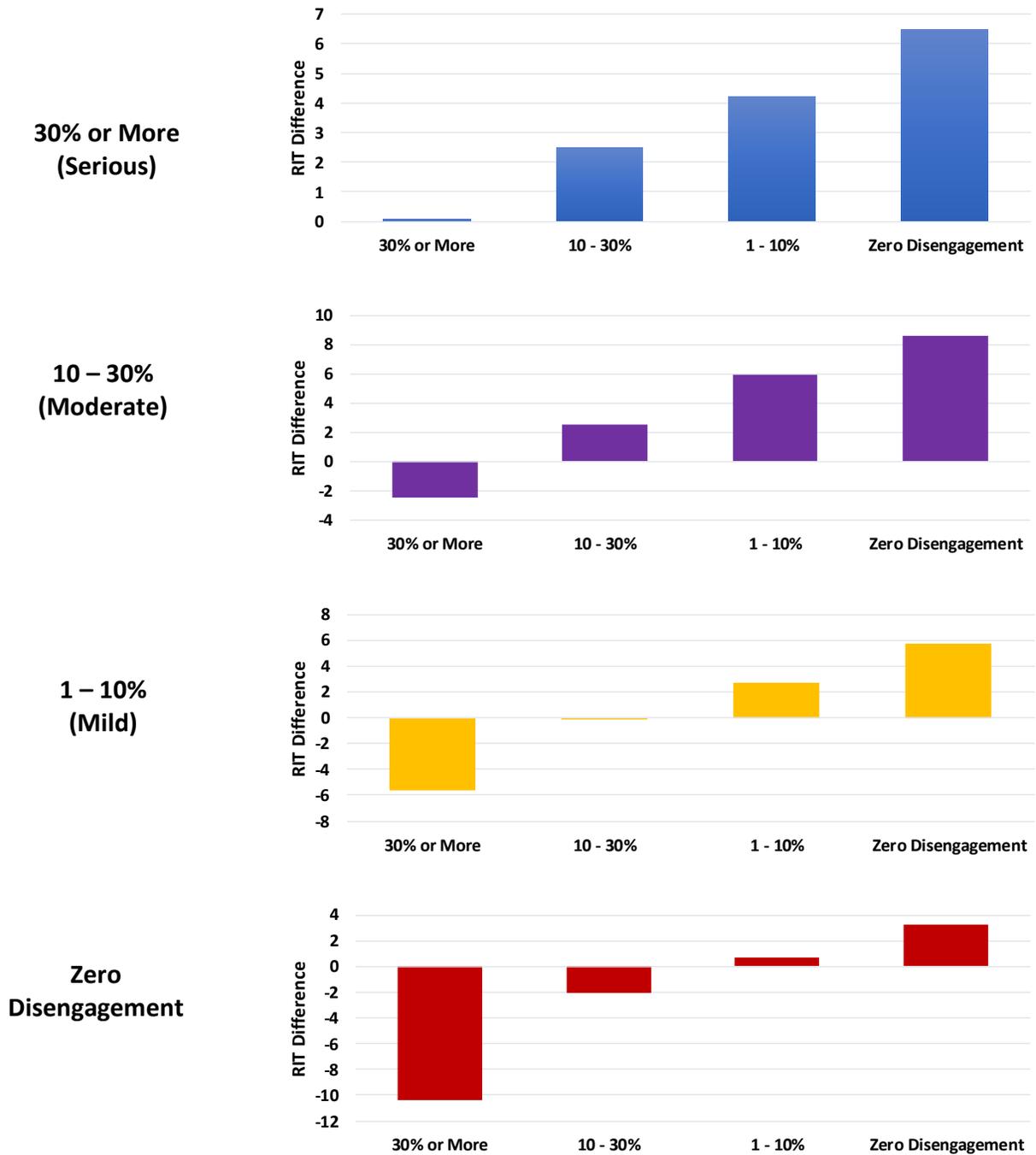
**Percent Disengagement  
on Second Test Event**



*Figure 1. Differences in student RIT scores across two MAP Growth Reading 2-5 test events that occurred no more than one day apart, by levels of test-taking disengagement.*

**Percent Disengagement  
on First Test Event**

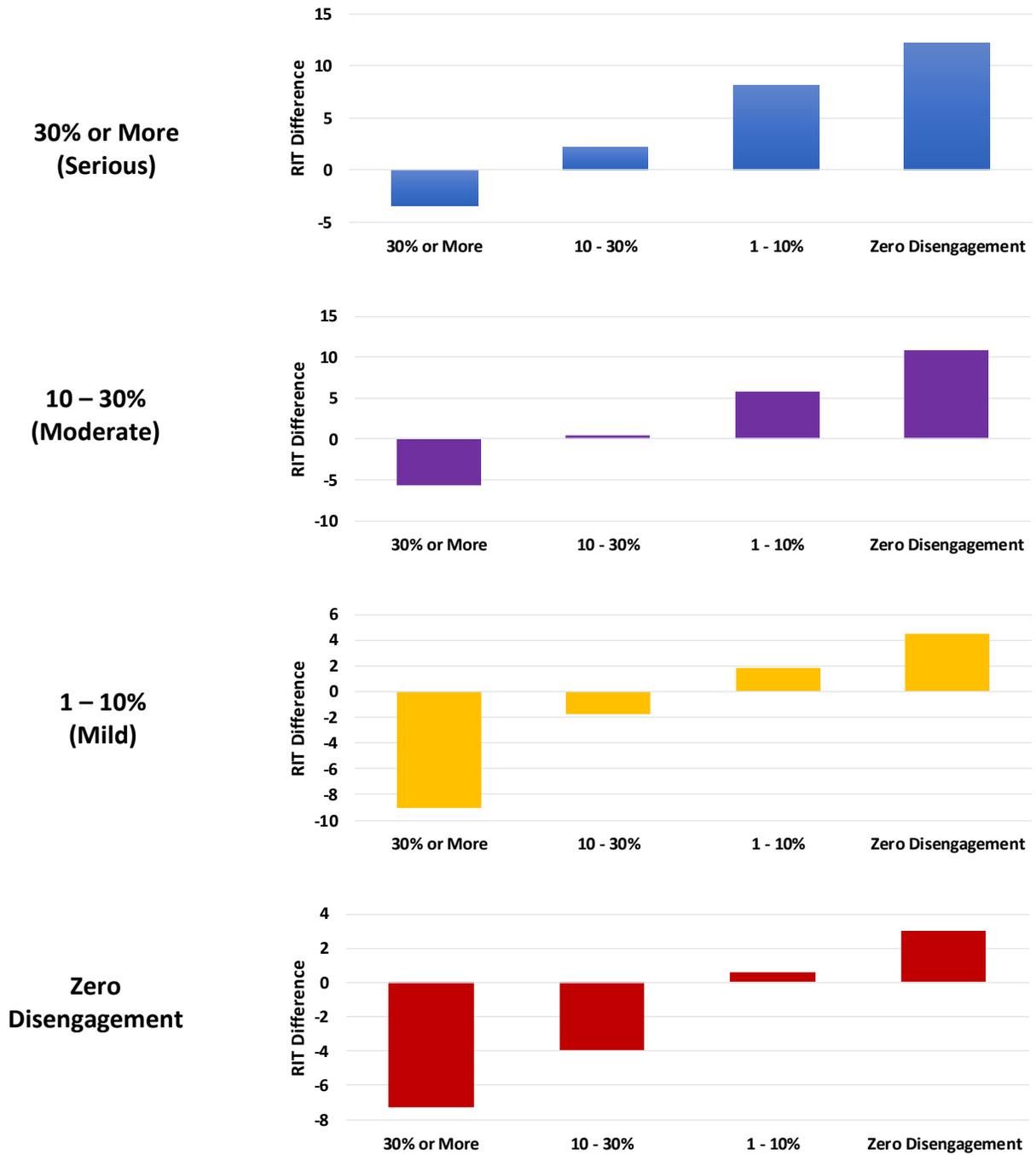
**Percent Disengagement  
on Second Test Event**



*Figure 2. Differences in student RIT scores across two MAP Growth Reading 6+ test events that occurred no more than one day apart, by levels of test-taking disengagement.*

**Percent Disengagement on First Test Event**

**Percent Disengagement on Second Test Event**

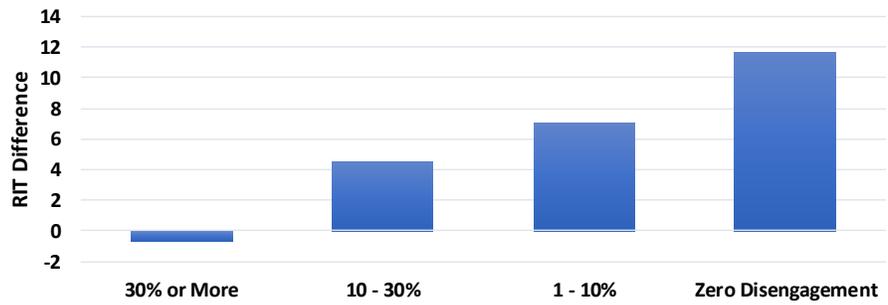


*Figure 3. Differences in student RIT scores across two MAP Growth Math 2-5 test events that occurred no more than one day apart, by levels of test-taking disengagement.*

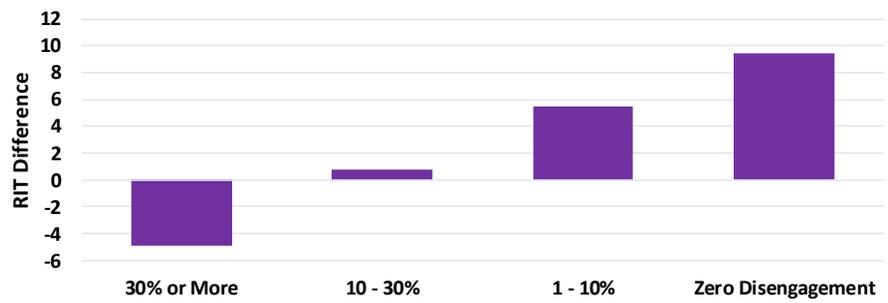
**Percent Disengagement  
on First Test Event**

**Percent Disengagement  
on Second Test Event**

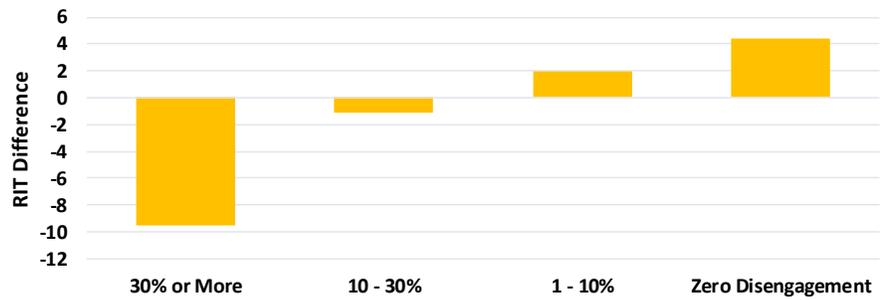
**30% or More  
(Serious)**



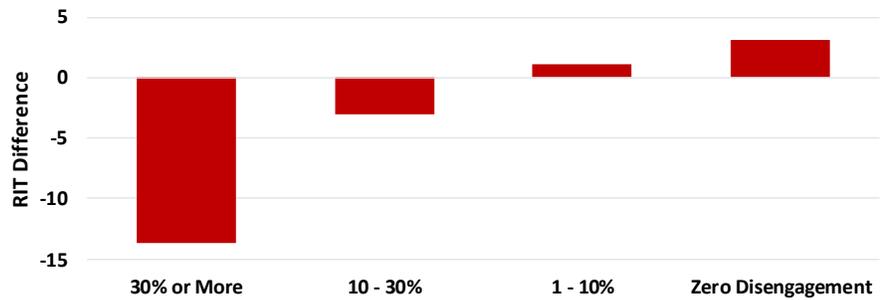
**10 – 30%  
(Moderate)**



**1 – 10%  
(Mild)**



**Zero  
Disengagement**



*Figure 4. Differences in student RIT scores across two MAP Growth Math 6+ test events that occurred no more than one day apart, by levels of test-taking disengagement.*