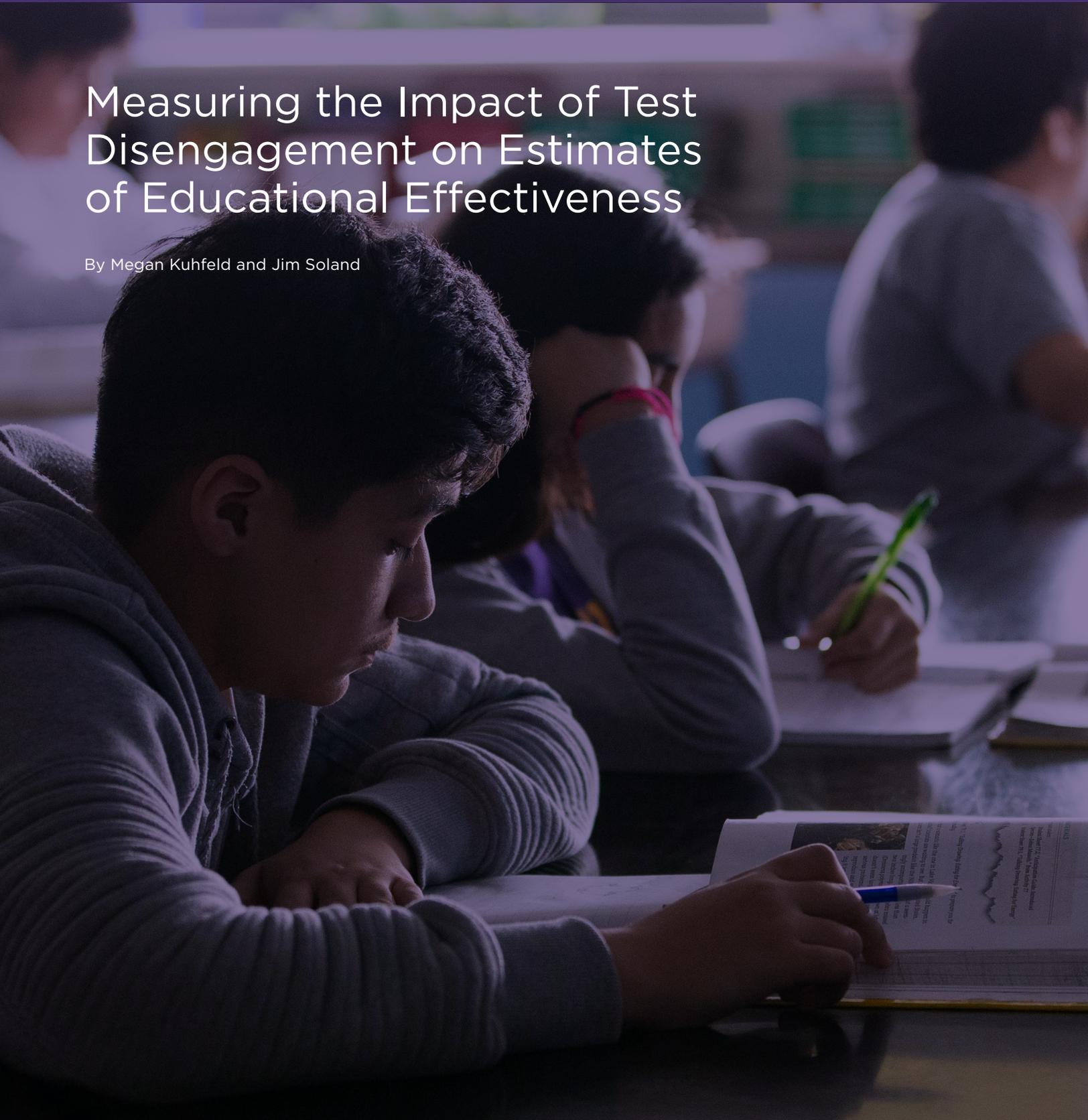


Measuring the Impact of Test Disengagement on Estimates of Educational Effectiveness

By Megan Kuhfeld and Jim Soland



KEY FINDINGS

- **Disengagement by subgroups:** Disengaged test taking rates are higher in middle school than elementary school and vary greatly by subgroup. Black students show the highest rates of rapid guessing, followed by White and Hispanic/Latinx¹ students (whose rapid guessing rates are fairly comparable). Asian students consistently show the lowest rates.
- **Impact of disengagement:** Overall, the two most common approaches used to account for test disengagement (filtering and score adjustments), do not impact estimates of school effectiveness, achievement gaps, or summer learning loss.
- **Filtering bias:** Removing unengaged test takers also removed students who are much lower achieving, even after adjusting scores for rapid-guessing. This may unintentionally bias estimates for some of the most historically marginalized subpopulations (e.g., middle school boys of color).

Assessment scores are often used in evaluations of teachers, schools, and programs under the assumption that these scores accurately reflect what students know. However, test disengagement is recognized as a potential problem for inferences made based on observed test scores, since scores may be negatively impacted when student responses are disengaged.

Researchers have developed disengagement metrics for computer-based tests using assessment metadata, including item response times, to identify when students are disengaged. This study examines approaches for identifying rapid-guessing, accounting for rapid-guessing in models, and, as an illustrative example, explores how disengagement might bias estimates of educational effectiveness.

Rapid guessing: a disengaged test taking behavior in which a student responds to a test item so quickly that they could not have understood its content. Rapid-guessed responses are correct at a rate no better than by chance.

Specifically, the study examines data from a cohort of 22,055 fourth grade students from 343 elementary schools through eighth grade in a state administering computer based tests, to examine whether rapid-guessing may impact estimates of:

- school contributions to student growth
- achievement gaps
- summer learning loss

For each, estimates were created in ways that did, and did not, address rapid-guessing. The research also examined whether “motivation filtering”—a common approach that removes data from students who rapidly-guess on a high portion of test items—may introduce bias into estimates of achievement, and so explores tradeoffs associated with different approaches to addressing rapid-guessing. The results and discussion provide insight for program evaluators and policymakers to help them examine the impact test disengagement may have on measurement in their own context.

Disengaged test taking is common and varies across groups

As computer-based tests have become increasingly common, researchers have new data on how long students take to respond to test items. Research has shown that rapid-guessing is common, and that

¹ Emerging research uses terms such as Latinx, a gender-neutral term, for individuals who origin or ancestry is in Latin America (Salinas & Lozano, 2017). Our use of the term Hispanic in this research brief is consistent with the terminology used at the time of data collection. Because Hispanic refers to persons who speak Spanish natively and/or have Spanish-speaking ancestry, while Latino/a/x is more frequently used to refer to individuals of Latin American origin and ancestry (Alcoff, 2005), we wanted to maintain the accuracy of the measure reported by schools and districts.

Citations:

Alcoff, L. M. (2005). Latino vs. Hispanic: The politics of ethnic names. *Philosophy & Social Criticism*, 31(4), 395-407.

Salinas Jr, C., & Lozano, A. (2017). Mapping and recontextualizing the evolution of the term Latinx: An environmental scanning in higher education. *Journal of Latinos and Education*, 1-14.

students who rapidly guess on 5 to 10% of items on a test may have observed test scores biased downwards compared to their true scores, often by more than .2 standard deviations. Some research also shows that the rate of rapid-guessing differs across student subgroups, and so may affect achievement gap estimates for some students. The data from this study support earlier findings: disengaged test taking, defined here as students who rapid-guessed on 10% or more of test items, was more common in Black students than in White or Hispanic students. Asian students showed the lowest rates of test disengagement. Rapid-guessing rates increased as students moved from fourth to eighth grade and was higher in fall than in spring.

Grade	White	Black	Asian	Hispanic
4	0.05	0.08	0.02	0.04
5	0.04	0.08	0.02	0.04
6	0.09	0.15	0.04	0.12
7	0.1	0.14	0.01	0.1
8	0.1	0.14	0.02	0.09

Proportion of students with disengaged test events by gender, race, and ethnicity in spring term testing. Fall data were also examined.

Rapid-guessing did not show a large impact on measures of educational effectiveness

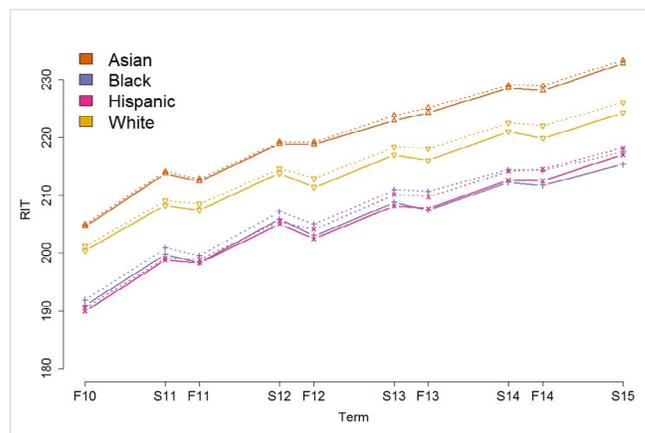
A modeling approach was used to investigate how rapid-guessing behavior affects estimates of schools' contributions to student growth, racial/ethnic achievement gaps in student learning, and summer learning loss. Two methods of accounting for rapid-guessing were used: motivation filtering, and effort-moderated scoring. The first approach filters out scores from examinees showing high rates of rapid-guessing and relies on the assumption that there is no correlation between true achievement and rapid-guessing. While many studies support this assumption, there is a growing recognition that there may be some correlation. Effort-moderated scoring addresses this by keeping the data for all test takers but adjusting scores to not count disengaged items. Results showed that, for this sample, inferences about schools' contributions to growth, achievement gaps, and summer learning loss were mostly unaffected by rapid-guessing behaviors.

However, before concluding that test disengagement generally is unlikely to impact metrics relevant to educational effectiveness, it is important to note that the rapid-guessing thresholds used here are very conservative to avoid wrongly discarding engaged responses for individual students. For program evaluation, where individual scores may be less

important than aggregates, evaluators might be more willing to risk discarding engaged item responses to reduce bias in the aggregate. Second, rapid-guessing is only one type of disengaged testing behavior, so more impact might be seen if other disengagement could also be quantified. Results may also differ for other assessments that are not adaptive or for high-stakes tests. These approaches are worth investigating in other evaluation contexts as a sensitivity check to test whether observed improvements in student achievement are biased by differential test engagement.

Filtering unengaged test takers may result in biased estimates in some circumstances

While filtering is a common approach for addressing rapid-guessing, this sample showed that students who were filtered on this had much lower effort-moderated scores than students who were not. Students who rapidly-guess often had much lower mean achievement, even when their scores were adjusted for rapid-guessed items. This raises a concern that filtering students with high rates of rapid-guessing from a sample may introduce bias by removing students who have low true achievement, are racial minorities, or both.



Comparisons of average student trajectories by race/ethnicity and whether filtering was used (where the solid line represents all students, and the dotted line represents only engaged students). This figure shows both the variation by subgroup and the impact of filtering.

RECOMMENDATIONS

Use metadata to provide more information on students

Response time metadata collected in computer-based tests provide new opportunities for researchers, policymakers, program evaluators, and others to understand student test disengagement.

Explore the relationship between test disengagement and key outcomes

We examined two approaches (effort-moderate scoring and filtering) to account for rapid guessing in important measures of educational effectiveness and did not find strong impacts in our study. However, there may still be large impacts in other settings, such as intervention studies that are targeted to middle and high school students.

Measure test disengagement in other assessments

With a computer-based test, item-level data, and associated metadata, student disengagement can be quantified by:

- setting response time thresholds to separate engaged and disengaged responses
- re-estimating scores to correct for rapid guessing. Using an effort-moderated approach may be more defensible and prevent the bias that may be introduced by filtering, but it is more labor intensive and can introduce greater imprecision in achievement estimates

Acknowledgments

The authors would like to thank Dr. Steve Wise for his feedback on earlier drafts.

This brief describes research documented in:

Kuhfeld, M. & Soland, J. (2019). *Using Assessment Metadata to Quantify the Impact of Test Disengagement on Estimates of Educational Effectiveness*. (The Collaborative for Student Growth at NWEA Working Paper).

Suggested citation:

Kuhfeld, M. & Soland, J. (2019). *Measuring the Impact of Test Disengagement on Estimates of Educational Effectiveness*. (The Collaborative for Student Growth at NWEA Research Brief).

ABOUT THE AUTHORS

Dr. Megan Kuhfeld is a Research Scientist II for the Collaborative for Student Growth at NWEA. Her research seeks to understand students' trajectories of academic and social-emotional learning (SEL) and the school and neighborhood influences that promote optimal growth. Kuhfeld completed a doctorate in Quantitative Methods in Education and a master's degree in statistics from the University of California, Los Angeles (UCLA).



Dr. Jim Soland is a Senior Research Scientist at the Collaborative for Student Growth at NWEA, and is an Adjunct Professor of Quantitative Methodology in the College of Education at Oregon State University. His research focuses on assessment and evaluation policy and practice, with particular emphasis on measuring social emotional learning, test engagement, and estimating teacher and school effectiveness. Soland completed a PhD in Educational Psychology at Stanford University with a concentration in measurement and policy.



ABOUT THE COLLABORATIVE FOR STUDENT GROWTH

The Collaborative for Student Growth at NWEA® is devoted to transforming education research through advancements in assessment, growth measurement, and the availability of longitudinal data. The work of our researchers spans a range of educational measurement and policy issues including achievement gaps, assessment engagement, social-emotional learning, and innovations in how we measure student learning. Core to our mission is partnering with researchers from universities, think tanks, grant-funding agencies, and other stakeholders to expand the insights drawn from our student growth database—one of the most extensive in the world.



NWEA is a not-for-profit organization that supports students and educators worldwide by providing assessment solutions, insightful reports, professional learning offerings, and research services. Visit [NWEA.org](https://www.nwea.org) to find out how NWEA can partner with you to help all kids learn.

© 2019 NWEA. NWEA is a registered trademark of NWEA in the US and in other countries.

APR19 | KAP3699