

Measuring Transfer Effects on Student Reading Comprehension

The Model of Reading Engagement (MORE) is a content literacy program that leverages engaging thematic lessons in science and social studies and rigorous books to build connections between taught words and knowledge of a topic (e.g., dinosaurs) to improve reading comprehension in new topics. A key component of evaluating the effectiveness of MORE is administering reading comprehension tests to MORE students. Typically, researchers would use



the total test score to measure the effectiveness of the MORE program. This study explores a new statistical method to measure MORE's effectiveness on each individual item in the test to provide a more fine-grained understanding of how MORE improves reading comprehension by promoting learning transfer.

How can we measure learning transfer in reading comprehension?

We collected data from 7,797 third grade students, who each completed an online reading comprehension test with thirty items. All students in the study received a 12-day sequence of science lessons. Students in the control group received two additional science vocabulary lessons, and students in the treatment group received two additional social studies extension lessons. The test was made up of three reading passages representing different degrees of transfer from the MORE curriculum. The “near transfer” passage included seven vocabulary words explicitly taught in MORE, the “mid transfer” passage included four, and the “far transfer” passage included none. Furthermore, the near transfer passage included only science content, but the mid and far transfer passages included science and social studies content. We predicted that treatment students would perform better on the mid and far transfer passages because they had practice transferring their learning in science to new contexts in social studies.

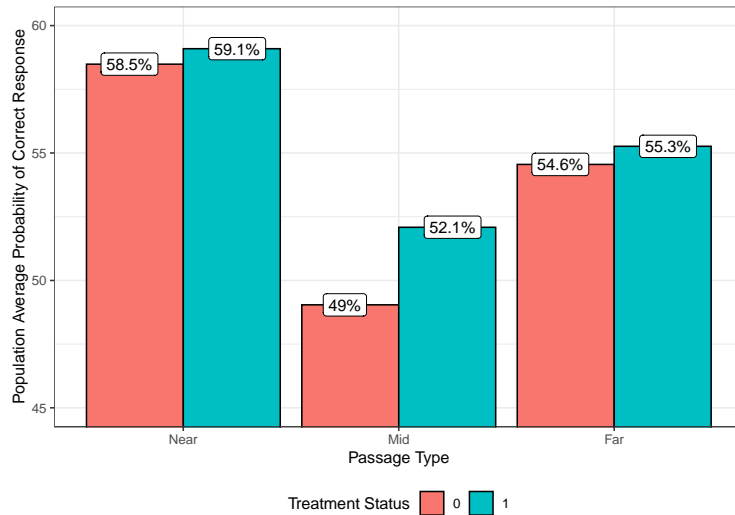
We find no treatment effect on overall reading comprehension, however, there were large positive treatment effects on the “mid transfer” items

When we estimate the treatment impact on the reading comprehension overall test score, the effect is positive (meaning that treatment students performed better than control students) but not statistically significant (meaning that the results could have been due to chance). Without considering the possibility of transfer effects on the different test passages, the analysis might have stopped here. When we allow for different transfer effects on the near, mid, and far transfer passages using the new

This brief describes work done for READS Lab at the Harvard Graduate School of Education based upon Gilbert, Kim, and Miratrix (2023), *Modeling Item-Level Heterogeneous Treatment Effects with the Explanatory Item Response Model: Leveraging Large Scale Online Assessments to Pinpoint the Impact of Educational Interventions*. The research reported here was supported by the Chan Zuckerberg Initiative.

statistical method proposed in this study, we find that there is a large, positive, and statistically significant effect of MORE on items from the mid transfer passage, but not the near and far transfer passages. This result shows that treatment students were able to successfully transfer their understanding from science to social studies.

The graph shows the accuracy rates for treatment and control students for the near, mid, and far transfer test passages. The y-axis of the graph represents accuracy rate, or the percent of students who answered the items on that passage correctly. The x-axis shows the type of test passage. We can see that the difference in accuracy rates between the control (red) and treatment (green) students is much larger on the mid transfer passage than the near or far transfer passages, demonstrating successful learning transfer.



Policymakers should consider how interventions affect individual test questions, not just the overall test score

When research measure the impact of educational programs like MORE, they often use an overall test score. This study shows that only using an overall score would have resulted in an incomplete conclusion about how MORE affects student reading comprehension. In this study, we used new statistical method to measure how MORE can impact different portions of a test and showed that treatment students were able to transfer their learning from science to social studies. Other researchers can apply our model to better understand on which items or types of items treatment effects emerge, allowing us to better understand how interventions affect test outcomes.