

Does Studying Student Data Really Raise Test Scores?



—Vanessa Solis/Education Week, Z_Wei/Getty

Researchers say the popular practice is largely ineffective

By Heather C. Hill

February 7, 2020

Question: What activity is done by most teachers in the United States, but has almost no evidence of effectiveness in raising student test scores?

Answer: Analyzing student assessment data.

This practice arose from a simple logic: To improve student outcomes, teachers should study students' prior test performance, learn what students struggle with, and then adjust the curriculum or offer students remediation where necessary. By addressing the weaknesses revealed by the test results, overall student achievement would improve.

Yet understanding students' weaknesses is only useful if it changes practice. And, to date, evidence suggests that it does not change practice—or student outcomes. Focusing on the problem has likely distracted us from focusing on the solution.

"My own recent experiences visiting schools imply this trend continues."

With the birth of large-scale state assessments and widening data availability in the 1990s, school leaders and teachers could access information on student performance that was common across schools and classrooms. Many schools also instituted standardized "interim" assessments, claiming that this periodic, low-stakes testing could help teachers identify difficult content and struggling students before the state assessment, giving both teachers and students a chance to catch up. Over time, educational testing and data companies including the Achievement Network, NWEA, and McGraw-Hill's Acuity began to sell interim assessments to schools and states, making such assessments (and their cousins, test-item banks that support formative assessment) a billion-dollar business.

Currently, a large number of teachers report they regularly get together to analyze student assessment results. In a 2016 survey by Harvard's Center for Education Policy Research, 94 percent of a nationally representative sample of middle school math teachers reported that they analyzed student performance on

 [Printer-Friendly](#)

 [Email Article](#)

 [Reprints](#)

 [Comments](#)

tests in the prior year, and 15 percent said they spent over 40 hours that year engaged in this activity. Case-study research suggests that in many Title 1 schools, this activity is a cornerstone of teachers' weekly or monthly collaborative time.

But here's the rub: Rigorous empirical research doesn't support this practice. In the past two decades, researchers have tested 10 different data-study programs in hundreds of schools for impacts on student outcomes in math, English/language arts, and sometimes science. Of 23 student outcomes examined by these studies, only three were statistically significant. Of these three, two were positive, and one negative. In the other 20 cases, analyses suggest no beneficial impacts on students. Thus, on average, the practice seems not to improve student performance.

One critical question, of course, is, why?

Observational studies suggest that teachers do, in fact, use interim assessments to pick out content that they need to return to. For instance, in a study published in 2009, Margaret E. Goertz and colleagues at the University of Pennsylvania observed teachers planning to revisit math topics using a combination of whole-group and small-group instruction.

But Goertz and colleagues also observed that rather than dig into student misunderstandings, teachers often proposed non-mathematical reasons for students' failure, then moved on. In other words, the teachers mostly didn't seem to use student test-score data to deepen their understanding of how students learn, to think about what drives student misconceptions, or to modify instructional techniques.

My own recent experiences visiting schools imply this trend continues. Field notes from teacher data-team meetings suggest a heavy focus on "watch list" students—those predicted to barely pass or to fail the annual state reading assessment. Teachers reported on each student, celebrating learning gains or giving reasons for poor performance—a bad week at home, students' failure to study, or poor test-taking skills. Occasionally, other teachers chimed in with advice about how to help a student over a reading trouble spot—for instance, helping students develop reading fluency by breaking down words or sorting words by long or short vowel sounds. But this focus on instruction proved fleeting, more about suggesting short-term tasks or activities than improving instruction as a whole.

Common goals for improving reading instruction, such as how to ask more complex questions or encourage students to use more evidence in their explanations, did not surface in these meetings. Rather, teachers focused on students' progress or lack of it. That could result in extra attention for a watch-list student, to the individual student's benefit, but it was unlikely to improve instruction or boost learning for the class as a whole.

In reviewing the research on teachers analyzing student data, I came across a small number of programs that included interim assessment as one part of a larger instructional package. While I excluded these studies from the formal review I undertook for this essay, they are notable nonetheless. One, by Janke M. Faber and colleagues in the Netherlands, focused on a program that not only contained computer-based interim assessments but also provided both instructionally focused feedback to teachers and students and personalized online student assignments. Another study, led by Jonathan A. Supovitz and colleagues at the University of Pennsylvania, examined the Ongoing Assessment Project, a program that helps teachers create assessments and examine the results, then combines this practice with professional development focused on mathematics content and student thinking about that content. Both of these studies saw positive impacts, suggesting that the analysis of data can, when combined with strong supports for improved teaching, shift student outcomes. But the small number of programs that combine the study of data with wider instructional supports limits our ability to draw real conclusions.

About this series



This essay is the second in a [series](#) that aims to put the pieces of research together so that education decision-makers can evaluate which policies and practices to implement.

The conveners of this project—Susanna Loeb, the director of Brown University's Annenberg Institute for School Reform, and Harvard education professor Heather Hill—have received grant support from the Annenberg Institute for this series.

[Read the full series here.](#)

MORE OPINION



[Follow Opinion here.](#)

In total, the research in this area suggests that district and school leaders should rethink their use of state and interim assessments as the focus of teacher collaboration. Administrators may still benefit from analyzing student assessment results to know where to strengthen the curriculum or to provide teacher professional learning. But the fact remains that having teachers themselves examine test-score data has yet to be proven productive, even after many trials of such programs.

For many schools, this news is disheartening. Retooling teacher collaborative time will be a major shift—and that's assuming that schools can first identify more effective ways to help teachers improve their instruction. In our next column, we'll cover possible replacement activities.

Heather C. Hill is a professor of education at the Harvard Graduate School of Education and studies teacher quality, teacher professional learning, and instructional improvement.