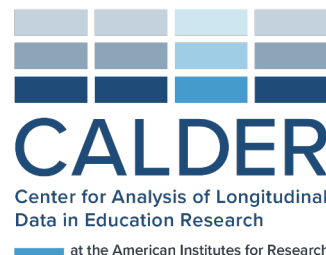

What Can Different Measures Tell Us About the Quality of the Teacher Workforce?



CALDER Research Brief No. 38

Ben Backes, James Cowan, Michael DeArmond, Dan Goldhaber, Roddy Theobald
September 2024

By considering test and non-test outcomes, leaders can deepen their understanding of their workforce and inform efforts to prepare, support, and develop effective teachers.

This CALDER research brief builds on pioneering research by the Measures of Effective Teaching Project (Kane et al., 2013) and Jackson (2018) to look at two measures of teacher quality and what they can tell education leaders about the teacher workforce.

Both measures use a value-added framework to separate out a teacher's contributions to student outcomes from other factors, like a student's previous academic record or economic circumstances. By trying to isolate teacher contributions to student outcomes, the measures speak to the question, *would students be better off with a different teacher?*

This brief looks at value-added measures (VAMs) for both test outcomes and non-test outcomes.

Test-based VAMs have been popular in policy and research since the 2000s (although researchers introduced them much earlier [Hanushek, 1971, 1992]). Despite criticisms of testing, studies suggest these measures tell us something about teacher effectiveness. For example, when teachers are better at improving test scores, their students are more likely to attend a quality college and earn more money as adults (Chetty et al., 2014). Non-test VAMs are newer. They look at outcomes like student attendance, grades, behavior, and grade progression.¹

Together, the two measures offer crucial reminders about the stakes and complexity surrounding effective teachers.

Key Takeaways

Test-based and non-test VAMs underscore three things about the teacher workforce:

- Consistent with prior research, the two measures show that teachers vary significantly in their ability to improve a range of student outcomes. But teachers who are good at improving test outcomes aren't necessarily good at improving non-test outcomes—and vice versa.
- When students have more effective teachers on either measure, their postsecondary outcomes improve. But teachers with different skills impact various postsecondary outcomes to different degrees.
- Teachers' varied effects on postsecondary outcomes depend in part on their skills and the achievement levels of their students.

Go Deeper at <https://caldercenter.org/>

Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2024). How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes. *Journal of Human Resources*, 1023-13180R2. <https://doi.org/10.3368/jhr.1023-13180R2>

TEST-BASED AND NON-TEST VAMS UNDERSCORE THREE THINGS ABOUT THE TEACHER WORKFORCE

1. Consistent with prior research, the two measures show that teachers vary significantly in their ability to improve a range of student outcomes. But teachers who are good at improving test outcomes are not necessarily good at improving non-test outcomes—and vice versa.

The scatter plot in Figure 1 uses a sample of around 8,000 teachers from Massachusetts. It shows teachers' standardized scores for both measures from 2019, with test-based VAM on the horizontal axis and non-test VAM on the vertical.

As the cloud of dots in the middle of Figure 1 suggests, teachers vary on both measures. Some teachers are relatively better or worse at increasing standardized test scores (shown by the horizontal spread). And some are relatively better or worse at supporting students' attendance, grades, and behavior (shown by the vertical spread).

One takeaway from Figure 1 is that the two measures are not strongly correlated. If they were, the dots would form an upward- or downward-sloping pattern. Instead, they are spread out in a cloud in the middle of the figure ($r = .10$). Teachers in the upper left of the figure are good at improving non-test outcomes but not test scores. Teachers in the lower right are good at improving test scores but not non-test outcomes.

Teachers who are strong in one dimension are not necessarily strong in the other.

FIGURE 1. Teachers Vary on Test and Non-test Measures of Effectiveness



2. When students have more effective teachers on either measure, their postsecondary outcomes improve. But teachers with different skills impact various postsecondary outcomes to different degrees.

FIGURE 2. Teachers’ Test and Non-test VAM and Postsecondary Outcomes.

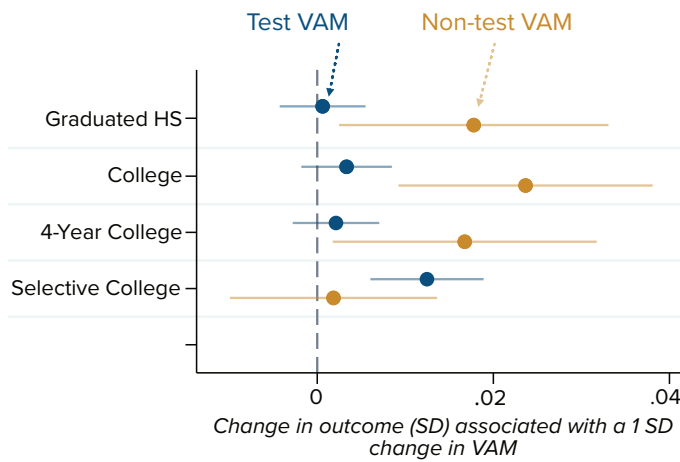


Figure 2 shows the relationship between both types of VAM and four important outcomes: high school graduation, attending college, attending a 4-year college, and attending a selective college.

As the figure suggests, teachers who are good at improving test scores increase the chances a student attends a selective college. Teachers who are good at improving non-test outcomes enhance the likelihood a student enrolls in college and attends a 4-year college.

The reasons for these differing effects are unclear. One possibility is that teachers skilled at improving non-test outcomes have a greater impact on students who require assistance in those areas, and that improving those outcomes subsequently raises the likelihood of those students enrolling in college. These divergent effects merit further investigation.

Why Multiple Measures Matter

Many states added student growth measures to their teacher evaluation systems in the 2010s during the federal government’s Race to the Top initiative. While some places have rolled back these policies, 27 states still permit or mandate the use of student test scores to evaluate teachers (National Council on Teacher Quality, 2022). There are good reasons for this. Despite criticisms of testing², research suggests that test results give leaders valuable insights about teachers. As noted earlier, studies suggest students are more likely to attend a quality college and earn more money as adults when teachers are better at improving test scores.³

But test outcomes obviously do not tell us everything we want to know about teacher quality. After all, teachers do many things—like instilling good habits, cultivating strong classroom relationships, and fostering a safe and inclusive learning environment—that go beyond what achievement tests measure. With this, some states and districts have expanded how they understand teacher quality. For example, Massachusetts, the site of the study behind this brief (Backes et al., 2024), requires evaluators to include evidence of teacher impact on student learning in performance ratings as one of multiple measures. Indeed, in its latest teacher evaluation rubric, Massachusetts expands the meaning of “impact on student learning” to include academic and non-academic outcomes, such as student engagement and sense of belonging.

3. Teachers' varied effects on postsecondary outcomes depend in part on their skills and the achievement levels of their students.

FIGURE 3. Varied Teacher Impacts (SD) Across Deciles of Student Achievement

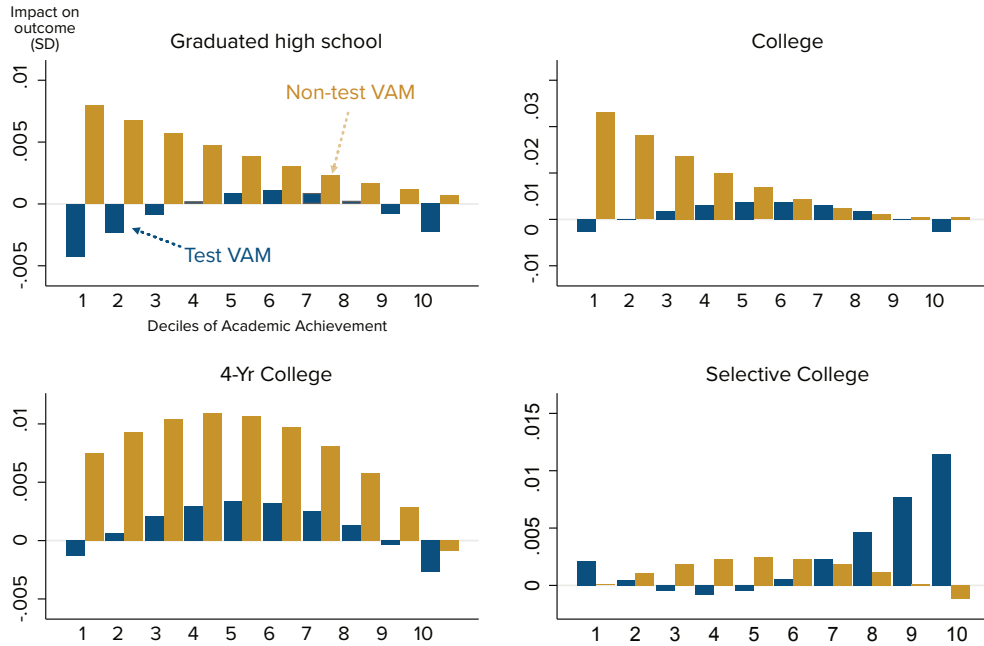


Figure 3 shows how the impacts of the two VAM measures vary across the distribution of student achievement in high school (represented in deciles on the x-axis). The bars in the charts reflect the standardized change in each outcome associated with a 1 standard deviation change in VAM. As with Figure 2, test VAM is shown in blue and non-test VAM is in gold.

The gold bars show that students with lower achievement levels have better graduation and college attendance outcomes when they have a teacher with high non-test VAM. By contrast, high-achieving students have better selective college attendance outcomes when they have a teacher with a high test VAM. These figures suggest teachers with different skill sets may play distinct but valuable roles in educating students across the achievement distribution. We need to know more about how these distinctions play out in practice and what their full implications are for educational opportunity and equity.

The Bottom Line

Given the critical role teachers play in student success, system leaders need to focus on the quality of their workforce and its distribution across students and schools. Leaders can use test-based and non-test VAMs to better understand their workforce by asking questions like:

- How do test and non-test measures of teacher quality vary across our district and schools?
- How do the measures vary based on teacher experience? What are the implications for how we prepare, support, and develop teachers?
- What practices do effective teachers on both measures use in the classroom? What can the system and other teachers learn from the most effective teachers to inform capacity building efforts?
- How do test and non-test measures vary by different pathways into the profession for (e.g., traditional vs alternative certification; different teacher preparation programs)?
- Are we keeping the most effective teachers on both measures? What is the gap in retention between the most and least effective teachers on both measures?

Answers to these questions might help leaders target resources to the places where they can do the most good. Instead of responding to testing critics by abandoning tests as measures of school and teacher performance (or turning to measures with questionable validity), leaders should seek different ways of measuring teacher quality that capture a richer picture of how teachers contribute to student success.

References

- Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, 34(2), 275–318.
- Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2022). Teachers and school climate: Effects on student outcomes and academic disparities. CALDER Working Paper No. 274-1022.
- Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2024). How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes. *Journal of Human Resources*, 1023-13180R2.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4), 1593–1660.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.
- Dobbie, W., & Fryer, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*, 123(5), 985–1037.
- Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, 32(4), 692–717.

Hanushek, E.A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. *American Economic Review*, 61(2), 280-288.

Hanushek, E.A. (1992). The trade-off between child quantity and quality. *The Journal of Political Economy* 100(1), 84-117.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy*, 126(5), 2072–2107.

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, 131(1), 157–218.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. Research Paper. MET Project. Bill & Melinda Gates Foundation.

LaFortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, 10(2), 1–26.

National Council on Teacher Quality. (2022). Measures of Student Growth National Results. State Teacher Policy Database. [Data set]. <https://www.nctq.org/yearbook/national/Measures-of-Student-Growth-95>.

Notes

¹ In other research, we consider teacher impacts on student perceptions of school climate (see Backes et al., 2022). Like their test-based counterparts, both non-test and climate measures are arguably too imprecise for making high-stakes judgments about individual teachers—unless pooled over numerous years of data.

² For example, critics argue that test-based accountability narrows the curriculum and focuses teachers and students on low-value test-taking skills. Others make stronger critiques related to the design and purpose of testing as they relate to racial and social justice. Another limitation: test-based measures are not available for all teachers—most work in grades and subjects without standardized testing.

³ A large body of studies on teachers (Chetty et al., 2014), peers (Chetty et al., 2011), small class sizes (Dynarski et al., 2013), finance reform (Jackson et al., 2016; LaFortune et al., 2018), and some school choice programs (Angrist et al., 2016; Dobbie & Fryer, 2015) support the idea that there is a causal link between what test scores measure and life outcomes.



CALDER

National Center for Analysis of
Longitudinal Data in Education Research

