

The Testing Mess—Why It Won't Go Away



“. . . despite the constant criticisms leveled at state tests, local assessment is arguably the far weaker link in the whole chain of would-be reform”

Grant Wiggins (2010) [“Why We Should Stop Bashing State Tests”](#)

Paul Zavitkovsky

Critics of standardized testing argue that rich, authentic classroom assessments make most current uses of standardized testing unnecessary. And they are right. On average, teachers who use high-quality classroom assessment improve achievement by one to one-and-a-half grade levels. But for close to a century, efforts to scale up rich classroom assessment have never gotten traction with more than [20%-25% of American teachers](#). Others have opted out . . . choosing instead to grade and motivate student achievement [in more conventional ways](#).

It was exactly this weakness that gave standards and accountability reformers a license to opt out on classroom teachers. They argued that most teachers didn't have the chops to assess challenging state standards. Their alternative—outsource big chunks of the job to commercial testing organizations.

If No Child Left Behind taught us anything, it's that out-sourcing assessment to commercial vendors [can't bail us out of this problem](#). But we shouldn't fool ourselves into believing that the problem started with standardized tests. As a culture, we rarely use assessment to help us teach and learn better. Local assessments are, and always have been, [mostly about grading](#). NCLB just upped the ante by doing unto the adults what we have been doing unto students for as long as anyone can remember.

The Cost of Opting Out from High Quality Local Assessment

A few years before NCLB became law, Paul Black and Dylan Wiliam generated big excitement in the assessment community with their now-classic meta-analysis of classroom assessment practices. Published in 1998, [“Inside the Black Box”](#) summarized two decades of evidence that all pointed to a single, spectacular conclusion. Engaging students with frequent,

high-quality feedback about how they are learning and where they are getting stuck improves standardized achievement by an average of 1 to 1 ½ grade levels (an effect size of 0.4 to 0.7 standard deviations).

In 2001, the University of Chicago's Consortium on Chicago School Research echoed Black and Williams' findings in a ground-breaking study called, [*Authentic Intellectual Work and Standardized Tests: Conflict or Co-existence?*](#) Over a three year period, the study assessed the connection between standardized achievement and the intellectual demand of classroom assignments.

A key feature of this study was that it took place in schools with high percentages of Black and Latino students from low-income households. 89% of the students studied were eligible for free or reduced lunch, 53% were Black and 39% were Latino.

Like "Inside the Black Box," *Authentic Intellectual Work* found that students at all achievement levels who were regularly challenged by intellectually demanding assignments achieved about a grade level higher than students who were not. This finding flew in the face of conventional notions that only higher-achieving students can handle and benefit from intellectually challenging tasks.

In 2008, John Hattie published [*Visible Learning*](#), the most extensive meta-analysis to date of factors that impact student achievement. Hattie showed that the effect sizes reported by Black and Wiliam and the Chicago Consortium were about the same as the impact of socio-economic status (SES) . . . 0.57 standard deviations or roughly one grade level. The impact on achievement that comes from engaging *teachers* with high-quality feedback about their day-to-day practice was even higher (0.9 standard deviations, or close to two full grade levels).

Most people understand that frequent, high-quality feedback has a big impact on student and adult learning. But many will find it surprising that the size of this impact is big enough to cancel out the negative effects of low SES. That is hopeful information for an education system that has struggled for decades to increase instructional effectiveness with students from low-income households.

But the opposite is also true. Positive effect size measures impact over-and-above what is typical. When frequent, high-quality feedback gets positive effect sizes of between one and two grade levels, it means most schools and classrooms are opting out from providing that kind of feedback.

Opting In to a Different Mindset about Assessment

In the 1985 movie classic, *Back to the Future*, Michael J. Fox's teen age character, Marty McFly, re-set his family history by traveling back to 1955 and helping his then-teenage father act more heroically at a key moment in his parents' relationship. State and local leaders can't turn back the clock. But big improvements in large-scale test design, and big increases in the flexibility of federal law, offer something almost as good. They offer a once-in-a generation opportunity to:

- Stop pretending we can finesse deep improvements in classroom assessment by sidestepping teachers and outsourcing the work to external testing organizations
- Stop blaming standardized tests for a longstanding culture of assessment that focuses almost entirely on grading and low-level content mastery to motivate student learning
- Stop expecting that individual teachers will somehow be able to change [deeply embedded cultural practices](#) without sustained institutional support at the school, district and state level
- Start making standardized tests part of the solution by reporting their results in ways that give teachers and teacher teams deeper insights into where instruction is getting stuck and why

Common Core assessments like PARCC and Smarter Balanced [aren't simply longer, harder, more expensive versions of NCLB tests](#). Both tests elicit rich information about how students *think*. Thoughtfully reported, this information can support the difficult work of building greater rigor and depth of knowledge into local assessment practices. But most parents and teachers saw none of that in the first round of PARCC and Smarter Balanced test reports. All they saw were new sets of scale scores, new proficiency levels, and much lower success rates.

It's not like we don't have good prototypes for reporting rich, standardized test information in user-friendly ways. The National Assessment of Educational Progress (NAEP) has spent years building [on-line item-maps and released items](#) that demystify what scale scores and proficiency levels mean, and provide concrete examples of how students respond to different levels of academic demand. If we had the political will, systems like these could be scaled up fairly easily, and at relatively low cost, to report item-level information at the school and district level.

In and of itself, better reportage of large scale test information won't solve the local assessment challenge. The solution for that lies where it always has . . . in ongoing collaboration and experimentation by teams of teachers who commit to helping each other build greater rigor and depth of knowledge into the work they do with students. Supporting those teams at scale is the job of states and districts. That job begins by equipping teacher teams with the best information we have about the challenges we expect them to address.

In 2003, the National Research Council offered a [thoughtful set of recommendations](#) for how to bridge the gap between standardized tests and local assessments. Under NCLB, those recommendations were universally ignored. The price we're paying for that [is widespread illiteracy about what large-scale assessments can and can't do](#), and growing distrust of standardized testing as a whole.

Testing in America has reached a crossroad. We can actively support improvements in local assessment by reporting large-scale test results in more meaningful ways. Or we can continue to opt out, and settle instead for another generation of disappointing grades.

Paul Zavitkovsky is a former Chicago Public Schools principal. His work as a leadership coach and assessment specialist at the [Urban Education Leadership Program of the University of Illinois-Chicago](#) is supported by a continuing grant from the [W. Clement and Jessie V. Stone Foundation](#)