

How Not to Assess the Situation

Grading and testing have gone astray, but eliminating student performance measures is the wrong prescription

Off the Mark: How Grades, Ratings, and Rankings Undermine Learning (but Don't Have To)

by Jack Schneider and Ethan L. Hutt

Harvard University Press, 2023, \$29.95; 296 pages

As reviewed by Adam Tyner

IN THE YEARS SINCE THE COVID-19 OUTBREAK, the grades and test scores that anchor our education system have been relentlessly disrupted. As the pandemic swept the globe, American schools canceled annual standardized testing, college admissions went “test-optional,” and students were offered “hold harmless” policies that prevented their grades from dropping, regardless of whether they completed assignments or even attended virtual classes. Most end-of-year testing returned to K–12 schools in 2021, but much of the “assessment holiday” has endured. Most colleges continue not to require SAT or ACT scores, states are eliminating high school graduation tests, and grading standards have slipped to their lowest levels on record. States and districts are fueling grade inflation through policies that, in the name of equity, prohibit penalties for late work, recalibrate grading scales in ways that make passing easier, require teachers to assign credit for assignments that aren’t turned in, and even eliminate grading penalties for cheating.

Into this accountability recession arrives a new book arguing that the idea of holding students accountable through measures such as grades and test scores is inherently misguided. Penned by Jack Schneider and Ethan Hutt, two education-school-based researchers, *Off the Mark* is an ambitious volume combining history, policy analysis, and prescriptive recommendations. The authors evaluate the key “assessment technologies” of modern education systems—course grades and external tests—arguing that their presence undermines the aims of education. Although many of the book’s recommendations are sensible, its grandest claims are unsupported by research or contradicted by it.

The role of grades and tests in our education system does need better grounding in theory. Many education writers and researchers assume that these measures serve a single purpose, such as predicting postsecondary success, or that they matter only to one set of stakeholders, such as parents. Schneider and Hutt explain that many measures emerged to serve one

role but now have multiple functions and stakeholders. The authors offer a helpful mnemonic for sorting out the stakeholders, explaining that the assessment strategies convey both “short-haul” messages to parents and students and “long-haul” messages to institutions such as colleges.

Unfortunately, short-haul messages are often garbled by the time they reach parents. Learning Heroes, a nonprofit organization that works to equip parents to support student success, has found via surveys that the “good” grades most students receive have about nine of ten parents convinced their kids are performing at grade level, despite only about one in four of them actually doing so. The organization’s most recent report shows that, even in our era of devastating learning loss,

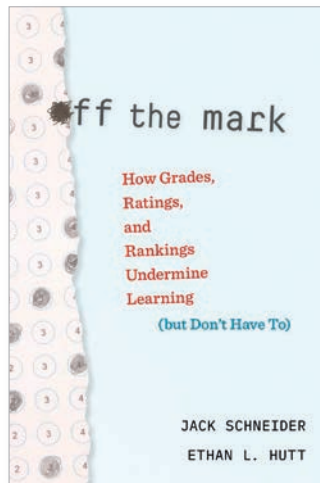
about four in five parents say their child is taking home mostly As and Bs. This disconnect is dangerous, because, as Schneider and Hutt note, “Families want to know how their children are doing, so that they can encourage, coax, and intervene as necessary.”

The book frames the multiple uses of grades and test scores as a dilemma, noting that the measures were not designed to support some of their current uses. The authors’ concern about the long-haul messages is not that they fail to communicate useful information, however. Grade point average and test scores are some of the best predictors of college performance and labor-market success, and the authors acknowledge that the utility of basing college admissions decisions on grades is one of their upsides. Their

critique is that any long-haul message raises the stakes for student performance, as the rating will follow the student far into the future. The authors join prior critics of teacher-assigned grading, including James Coleman and John H. Bishop, in noting how the classroom dynamics around grading help explain grade-grubbing, “nerd harassment,” and other toxic dynamics between students and teachers and between students and their peers.

Yet the authors’ assumption that students’ having greater stakes in their academic performance undermines their learning is at odds with the work of those earlier critics. Indeed, the authors make assertions about grades and test scores harming student motivation that are either unsubstantiated, mostly contradicted by research, or missing analysis of the social dynamics around grades and test scores that researchers have identified.

The authors’ antipathy toward the use of grades and test scores as motivators stems from their unarticulated theory



of learning—a version of the pop romanticism that is often attributed to philosopher Jean-Jacque Rousseau but is better represented by self-help and education writing of the last few decades, such as Daniel Pink’s *Drive* and Alfie Kohn’s *Punished by Rewards*. The pop romantics contend that the use of incentives in education undermines students’ intrinsic desire to learn. In the 1970s, psychologists, including Richard Ryan and Edward Deci, whom Schneider and Hutt cite, found that, under certain conditions, incentives can backfire. The pop romantics, though, have reduced these findings to a simplistic dichotomy: intrinsic motivation is good, and extrinsic motivation is bad. In fact, psychologists have demonstrated that educators can leverage both kinds of motivation. Many studies show benefits to students when they are held accountable for their academic performance, whether from strict-grading teachers, large cash incentives for academic success, or classroom reward systems. *Off the Mark* fails even to mention this body of research, let alone engage with it to synthesize a new approach to assessment.

Schneider and Hutt also object to grading interim assignments such as homework. “If students are going to receive cues about the kind of work that is important in school, those cues should point to substantive knowledge and skills,” they argue. Yet one could counter that accountability for short-term performance serves a valuable purpose. Grading such work motivates students and deters them from procrastinating. Without shorter-term goals, even motivated students may wait until the end of the semester to cram for the final exam. By ignoring the substantial body of scholarship connecting student academic motivation to accountability, Schneider and Hutt’s analysis is left undertheorized and incomplete.



Jack Schneider Ethan Hutt

The authors’ recommendations for change include both level-headed suggestions and ideas that are less compelling. They make three main proposals for reform: allow students to “overwrite” prior grades; base assessment on “a common set of performance-based tasks . . . aligned with a common set of competencies”; and deepen the information that transcripts convey by making them “double-clickable.” As an example of the latter, they recommend the work of the Mastery Transcript Consortium, which places students’ secondary school experiences into a format akin to “a high schooler’s LinkedIn.”

The idea of overwriting grades offers a distinction without much difference, because transcripts already reflect observable progress (or lack thereof) in each subject a student takes.

The “good” grades most students receive have about nine of ten parents convinced they are performing at grade level, though only about one in four are actually doing so.

If the student earns a C in Algebra I and an A in Algebra II, the progress is obvious; students are free to highlight it, and college admissions officers are free to take it into account. Making grades “overwritable” adds another mechanism for inflating grades while encouraging students to procrastinate. “I’ll figure out how to factor polynomials later,” an Algebra II student might well conclude.

Their second suggestion, basing grades on “performance-based tasks,” is akin to using portfolio assessments. This concept is controversial, but if it is part of “a system that incorporates both grades and portfolios”—and some external tests—it could encourage students to focus on developing skills that other assessments might miss while conveying more qualitative information to stakeholders. In other words, if digital portfolios complement the traditional assessment technologies rather than displace them, they could add real value. Schneider and Hutt point to Advanced Placement and International Baccalaureate as examples of programs that, at least for some subjects, successfully combine a variety of

assessments, including examples of student work.

As for their third recommendation, consolidating the information in digital portfolios with student transcripts, the Mastery Transcript Consortium they suggest as a model has already pivoted to a format that better incorporates traditional transcript material, ensuring that GPA and assessment outcomes from AP and college admissions exams are available alongside the new, qualitative elements.

The key reason the analysis in *Off the Mark* falters at times is that even as the authors view students as rational and strategic, they oppose leveraging those qualities to incentivize greater learning. They offer no evidence to suggest that relying on intrinsic motivation alone can address students’ disinterest in academics and today’s skyrocketing absenteeism. In their recommendations chapter, they write that “addressing extrinsic motivation [by removing stakes attached to grades] at least opens the door for conversations about how to foster intrinsic motivation.” Ignoring the idea that education systems might need to engage both types of motivation, the dichotomy leads the authors to recommend “minimizing, to the extent possible, the use of carrots and sticks.”

Left off the menu are reforms to address the faults of current accountability measures, such as improving standardized tests so they rely less on multiple-choice questions or separating teaching and assessment so as to disrupt the morally hazardous dynamic between students and teachers. Both could help solve the problems Schneider and Hutt identify in their book. Unfortunately, the authors’ distorted view of human motivation too often leads their analysis astray.

Adam Tyner is national research director at the Thomas B. Fordham Institute.