

More Facts, Fewer Hopes

Evidence fails to sway in testing policies

Uneducated Guesses: Using Evidence to Uncover Misguided Education Policies

By Howard Wainer

Princeton University Press, 2011, \$24.95; 200 pages.

As reviewed by Mark Bauerlein

What caring educator would not favor tests that allow students a choice in what they must answer?

What responsible college admissions officer wouldn't grant applicants the right to withhold their SAT scores?

What committed Advanced Placement (AP) teacher wouldn't expand access to as many students as possible?

What enlightened test developer wouldn't prefer tests that identify each test-taker's actual knowledge and skill levels and not those that just deliver a numerical score?

These aren't just personal attitudes. They sway large organizations as well. The College Board repeatedly talks about "equitable access" to AP courses, and a 2008 report by the National Association for College Admission Counseling solemnly states, "there may be more colleges and universities that could make appropriate admissions decisions without requiring standardized admission tests such as the ACT and SAT," and further, "some control clearly rests in the hands of postsecondary institutions to account for inequities that are reflected in test scores."

But what happens when SAT scores are optional in college applications? When Bowdoin College allowed it, two results emerged, both predictable. One, applicants who withheld their numbers scored on average 120 points lower than did those who submitted their scores.

Their withholding hence improved their applications, and it also boosted Bowdoin in the all-important *U.S. News & World Report* rankings (by making the average SAT score of the entering class look higher). But, two, the "withholders" hurt Bowdoin, for they performed 0.2 grade points worse than "submitters" did in first-year courses.

Or, what happens when tests allow students to choose the questions they answer, for instance, presenting a pool of essay questions from which test-takers choose two? First of all, you end up with inconsistencies: some questions are harder than others. And second, students often choose poorly, selecting the harder questions. Indeed, one study discovered, "the more that examinees liked a particular topic, the lower they scored on an essay they subsequently wrote on that topic!"

These outcomes belie the policies behind them, and they frustrate the generous souls who crafted the plans. Students end up performing worse than officials expected. Who wants to hear the bad news, though? Not many, and that's precisely the complaint of distinguished statistician Howard Wainer in his book *Uneducated Guesses*, in which the preceding quotation and the Bowdoin case appear. Most educators stick to their faiths rather than follow the evidence, Wainer complains, and their stubbornness necessitates this blunt retort to education policies founded on bad evidence and good intentions. The volume's subtitle, *Using Evidence to Uncover Misguided Education Policies*, describes the method. In 11 curt chapters, Wainer analyzes actual

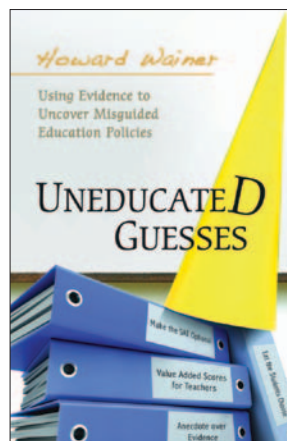
data and uncovers glitches, quirks, misconceptions, and unintended consequences of one practice after another, particularly those related to tests.

Each practice, from Computerized Adaptive Testing (CAT) to coscaling achievement tests, aims to solve a problem or address a need, but under Wainer's withering assembly of numbers (scores, dollars, demographics), they collapse. He notes the discomfort people feel with the exclusive nature of AP courses, but wonders if it's right to open them to students who have little chance of passing the exam. On principle, many would answer, "Give everyone a chance!" But, Wainer replies, such principles aren't free. He takes the case

of AP Calculus results in Detroit and estimates that if the city were to restrict the course to students who score 66 or above on the PSAT Math test, then the resulting cost per passing score on the AP test would be \$1,167. If the city set the eligible score much lower, at 31, the cost per passing score would reach \$4,513. "Would it be a better use of resources to provide a more suitable course for

the students who do not show the necessary aptitude?" Wainer suggests.

In the case of CAT, educators favor the format because it calibrates questions to a student's ability. If a test-taker misses a question, the next question shifts downward in difficulty. If he aces a question, the next one shifts upward. After a few dozen questions, the test identifies the competency level of the student—a better diagnostic than a simple percentage score. But it doesn't allow test-takers to review and



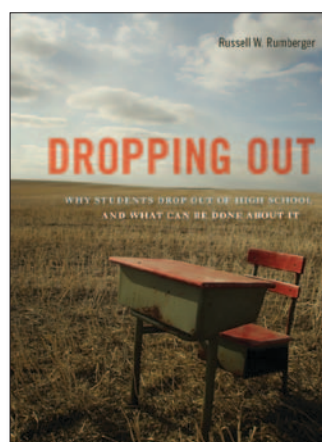
change an answer, which assessment experts consider important to accurate testing. If CAT does incorporate question review, Wainer warns, then when a subject finds an easy question pop up, he assumes he got the previous one wrong and backtracks to change it. Or worse, he deliberately answers every question wrong, ensuring easy questions all the way through. At the end, he returns to the beginning and answers every question correctly, yielding a near-perfect score. In other words, the very customization that educators praise allows savvy students to game the test. Wainer issued that caution in 1993, and he advises that we keep the original CAT because the benefits of item review

don't outweigh the risks of its abuse. Nevertheless, he notes, test specialists have pressed forward with item review since then—another case of hope overriding evidence.

However sharp and persuasive these exposés, though, they stand at a disadvantage, and Wainer knows it. This idea sounded so right, that innovation so sensible and fair, and watching them fail is depressing. Wainer summons evidence and reality against the modifications, but at stake is not just this and that policy but avid social hopes, sympathy for students, and feelings of injustice, too. One advocate for question review on CAT tests asserts that students “feel at a disadvantage when

they cannot review and alter their responses,” their feelings apparently forcing a change in format. Another proponent begins with a basic condition of test-taking, namely, stress, leading the authors to craft methods that allow students more control over the test but that identify cheating (one recommendation they make is to limit changed answers to 15 percent of the total number of answers). Wainer cites both, but has only a dry reply taken from Albert Einstein: “Old theories never die, just the people who believe in them.”

Mark Bauerlein is professor of English at Emory University.



DROPPING OUT

Why Students Drop Out of High School and What Can Be Done About It

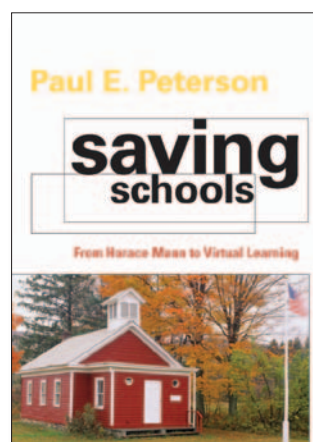
RUSSELL W. RUMBERGER

“The most complete examination of the dropout issue I have ever seen ... Russell Rumberger examines every complex nuance, summarizes every important research paper and

demolishes every Internet myth. His book is a masterpiece, something education wonks will keep close by ... We can't make any improvements, however, without knowing what hasn't helped dropouts, and why. On those vital questions, this book will be the best resource for years to come.”

—Jay Mathews, *Washington Post*

New in cloth | \$35.00



SAVING SCHOOLS

From Horace Mann to Virtual Learning

PAUL E. PETERSON

“In *Saving Schools* Paul Peterson shows how ... public schools became a political football with families and communities on one end of the field and states, courts and federal

governments on the other. Peterson sees virtual learning as the solution—technology could be used to reinvigorate the personalized approach that public education's founders and philosophers dreamed of.”

—Susan Salter Reynolds, *Los Angeles Times*

Belknap Press | New in paper | \$18.95

