

## Book Review

*Grading Education: Getting Accountability Right*  
Richard Rothstein, Rebecca Jacobsen, and Tamara Wilder  
(Teachers College Press, 280 pp., \$19.95)

Rebecca Hinze-Pifer

---

Accountability has been a watchword in education reform for the past decade. In 2002, when Congress passed *No Child Left Behind Act of 2001* (NCLB) with bipartisan support, its backers argued that holding schools and individual teachers accountable for their students' test scores would incentivize improvement. The Department of Education is currently using "Race to the Top," a small grant program created by the *American Recovery and Reinvestment Act of 2009* (ARRA), to encourage school districts to adopt performance pay models for educators. The underlying assumption of these reform movements is that policymakers and education leaders understand performance data and educational effectiveness well enough to evaluate schools and teachers based solely on standardized tests. *Grading Education: Getting Accountability Right*, by Richard Rothstein, Rebecca Jacobsen, and Tamara Wilder (2008), challenges this assumption and proposes an alternate model for school and teacher evaluation.

Rothstein frames his analysis in terms of a single question: what are the goals of the American educational system? Through an historical review, he identifies eight areas which education leaders and political figures have consistently considered important: basic academic knowledge and skills, critical thinking and problem solving, appreciation of the arts and literature, preparation for skilled employment, social skills and work ethic, citi-

zenship and community responsibility, physical health, and mental health. He supports these divisions with an interesting array of historical evidence, from Thomas Jefferson and Horace Mann to *A Nation at Risk* and modern Supreme Court rulings on educational adequacy. Together, these examples paint a long-term picture of the balance and tension between public education's sometimes disparate goals.

Based on his historical analysis and the eight identified goal areas, Rothstein surveyed the general public, school board members, and state legislators to determine their opinions of the relative importance of each goal area. When participants were asked to apportion 100 points between the goals, there were few differences between the three groups, and no goal area received an average of fewer than seven points. Methodologically, one weakness of this design is that it implies that all eight are important and may sway respondents to assign some worth to everything. Furthermore, the relative importance of each goal area is influenced by the choices the authors made in articulating the goals. For example, the physical health and mental health goal areas may have received fewer points total had they been combined. With that caveat, Rothstein's argument relies on the survey results only insofar as they support his contention that schools have important goals not considered in current accountability models. Through the historical review and survey, Rothstein effectively makes the point that the American education system has a wide variety of goals, of which basic academic learning is but one.

At its core, *Grading Education* focuses on how recent education reforms distort the American educational system. Rothstein argues that standardized tests are not reliable enough to be the sole measures of educational success and suggests that educational practice is warped by accountability systems which measure some goals and ignore others. His argument is not that standardized tests should not be used, nor that schools should not be evaluated and held accountable. Instead, he suggests a broader model of assessment would paint a holistic picture of a school's performance across the goal areas.

The chapters *Grading Education* devotes to critiquing recent reforms are the strongest in the book, if not the most original. Rothstein presents a compelling, well-supported argument that if schools are held accountable

for only a few goals, they will focus on those goals, often to the detriment of other, untested goals. Mixing anecdotal evidence from interviews, more rigorous studies of school and state education decision making, and examples of perverse incentives created by numerical accountability systems in other fields, Rothstein convincingly argues that struggling schools divert resources away from untested goals like art and history for the sake of maximizing test preparation. Although this argument has been made repeatedly by others, Rothstein's discussion of the goals of education strengthens his thesis and underscores the broader consequences of goal distortion.

The more compelling and unique portions of Rothstein's thesis are that standardized tests are unreliable measures of either a child's knowledge or a teacher's skill, and that high-stakes accountability should be based on broader assessment tools. Recognizing that almost all students can reach certain benchmarks, it would be fallacious to argue that all students have equal abilities. A classroom of twenty-five students usually includes significant variation in student ability. Furthermore, classes vary significantly from one year to another, and a student's performance on the day of the test may be significantly above or below his or her average performance. These variations are statistical realities. On the school level, there are enough students in the sample that statistical fluctuations average out and test scores can reliably measure school success in certain goal areas. However, NCLB requires that test scores be reported by grade level, by subject, by teacher, and by subgroups including special education students, racial groups, and English language learners. As these data are parsed into smaller categories, the results necessarily have a greater degree of uncertainty. Reform-minded policymakers want to hold teachers accountable for student achievement; Rothstein makes a clear case that standardized tests are insufficient tools for this task.

In addition to concerns about test reliability, *Grading Education* also outlines perverse incentives created by testing. For example, NCLB-mandated standardized tests set a single score cutoff. Below the cutoff, students "fail"; above it, they "pass." The percentage of students who pass the test is the only measure of teacher performance. As a result, teachers have little incentive to provide challenging material beyond the remedial level of the tests or to establish advanced instruction for gifted students. Even more

disturbingly, teachers with a large number of underperforming students have an incentive to focus on “bubble kids,” or those who perform just below the cutoff, and ignore students who are so far behind that no amount of additional instruction is likely to bring them up to passing within the span of a year. Outside of the classroom, state test designers have incentives to write easy tests. There is no academic consensus on what children of any given age should know and be able to do, and NCLB gives states almost complete freedom to design their own tests. As a result, tests focus on the most basic subject matter, with little attention to higher order thinking skills or difficult subject matter. In states with relatively weak education systems, more rigorous tests would label a huge number of schools “failing” and overwhelm the available resources for reform. Rothstein presents evidence that many states have made their tests easier over time and that tests have radically different levels of difficulty from one state to another. These perverse incentives distort classroom practice and devalue the non-academic goals of education.

On the balance, Rothstein’s argument is that teachers and schools should be assessed holistically. The latter portion of the book is devoted to exploring holistic accountability programs used elsewhere and sketching the outlines of such a program that could be used in the United States. Rothstein’s proposal fuses the multifaceted approach adopted by the original version of the National Assessment of Educational Progress (NAEP) with the existing optional school accreditation process and the inspection-based assessment model used in Great Britain.

In the model Rothstein proposes, the federal government would be limited to two roles: expanding the NAEP testing program to produce reliable state-level data in all eight goal areas, and equalizing educational resources between states. School and teacher standards, assessment, and accountability, he argues, should be established at the state level. In this model, state leaders would use NAEP results to inform their education policy and funding choices. The expansion of NAEP is a reasonable proposal; a relatively modest investment would produce a broader and more robust understanding of the state of America’s schools. However, only a small percentage of students take the NAEP test every year, so states

would still have to develop their own standardized assessments. The high cost of test development currently causes states to rely on similar test questions from one year to the next, making teachers more likely to distort their instruction to focus on the test. Currently, all 50 states separately invest huge amounts of money to reinvent the wheel: if states combined their test development efforts, assessment would be more reliable, less predictable, and dramatically less expensive.

In the two years since *Grading Education* was published, the National Governor's Association and the Council of Chief State School Officers have formed a coalition to support common assessment and Secretary of Education Duncan has set aside \$350 million in ARRA funding to support the effort (Cavanagh 2009). This March, they released draft standards in reading and mathematics (National Governors Association, and Council of Chief State School Officers 2010). Although Alaska and Texas were the only states that did not participate in developing the standards, it is unclear how many states will adopt the final version. Rothstein's support for state-based assessment was grounded in the belief that political polarization was less likely to scuttle the process than if it were conducted on the national level. Through inter-state collaboration, the common core movement will hopefully be able to exploit the financial efficiency of a single test while avoiding the polarization concomitant with federal policymaking.

Rothstein's proposal that the federal government equalize school funding between states is laudable, but would require a significant political shift to be viable. From a national perspective, Rothstein has a point. A 1998 study found that intrastate inequality accounted for only one-third of the total disparity in education funding; the remaining two-thirds resulted from inequality between states (Murray, Evans, and Schwab). Mississippi and Tennessee unquestionably have weaker tax bases than Oregon or New Jersey; poorer states are fundamentally unable to fund education at the same level as their richer peers. However, while the future civic engagement of the American populace and viability of the American workforce is certainly a matter of national concern, one could make similar arguments for any number of government programs that are currently in the purview of the states. Short of a dramatic change in the politics surrounding redistrib-

utive policies, it seems unlikely that the federal government would be able to enact wholesale interstate equalization. One more viable reform would change the equation used to allocate federal education dollars. Currently, states with large tax bases get more education funding from the federal government because it uses a matching grant model. A more equitable system would apportion federal dollars based on the amount of state spending relative to the state's tax base, rather than the absolute dollar amount. While Rothstein acknowledges the political difficulty of the task, his discussion does little to illuminate the path to financial equality.

On the state level, *Grading Education* presents a viable model for school accountability. By expanding, professionalizing, and making mandatory the existing accreditation system, state policymakers would take advantage of existing infrastructure to create a robust mechanism for school evaluation. In its current form, accreditation is optional and usually staffed by volunteer evaluators. Schools in danger of losing accreditation opt out of the system, and the quality of evaluations varies widely from one evaluator to another. In contrast, accreditation plays a vital role in American higher education, where a bad report almost always costs the university president his job, and loss of status can cause a university to shut its doors. Rothstein proposes that reformers develop the existing accreditation system into a British-style school inspection model, in which professional evaluators visit schools and make recommendations on an array of educational dimensions. *Grading Education* presents a grounded discussion of the delicate balance between assessment for the sake of accountability and assessment for the sake of iterative improvement. This level of accountability requires adequate funding, and is one area where set-aside federal spending would leverage state compliance and help address tax base inequality.

Standardized testing is often touted as a tool for teacher accountability. *Grading Education* presents a clear explanation of why single-administration tests are insufficient for teacher evaluation and a solid argument that teachers and schools should be evaluated based on all eight of the goals of education. However, Rothstein's proposed accountability program speaks only to school-level evaluation. Currently, most teachers are evaluated based on a brief annual observation by a school administrator combined with student

test scores and informal feedback from students, parents, and other teachers. This model is woefully inadequate; it is barely sufficient when a school leader wants to fire a bad teacher, and it provides good teachers little meaningful feedback or assistance in professional growth. Teacher quality is an important facet of educational quality, and a complete model for school accountability should include a strong paradigm for teacher evaluation.

Although the accountability structure proposed in *Grading Education* has a few flaws, it would be a significant improvement over the myopic test-driven system currently in use. Standardized tests have a place in performance evaluation, but their exclusive use as performance measures distorts the broader goals of our educational system. As Rothstein convincingly argues, education accountability tools should be holistic and hold schools responsible for educating well-rounded, work-ready, healthy citizens.

## References

- Cavanagh, Sean. 2009. Consortium's Algebra Test Again Yields Poor Results; Sponsors See Evidence of States' Cooperation to Craft Common Exams. *Education Week* 29(6): 12.
- Murray, Sheila E., William N. Evans, and Robert M. Schwab. 1998. Education-Finance Reform and the Distribution of Education Resources. *American Economic Review* 88(4): 789-812.
- National Governors Association, and Council of Chief State School Officers. March 10, 2010. *News Releases*. <http://www.corestandards.org/> (accessed March 13, 2010).
- Rothstein, R., R. Jacobsen, and T. Wilder. 2008. *Grading Education: Getting Accountability Right*. Washington, D.C. and New York, N.Y.: Economic Policy Institute and Teachers College Press.

---

Rebecca Hinze-Pifer is in her first year of the Master of Public Policy program at The George Washington University. She earned a Bachelors of Science in astrophysics and computer science from the University of Wisconsin - Madison and taught science for seven years. In her spare time, she enjoys pottery and hiking.

---

