First published 2014
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2014 Taylor & Francis

The right of the editor to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

The big lies of school reform : finding better solutions for the future of public education / edited by Paul C. Gorski and Kristien Zenkov.
pages cm
Includes bibliographical references and index.
1. School improvement programs—United States. 2. Public schools—United States. 3. Education—Aims and objectives—United States. 4. Education—Standards—United States. 5. Educational accountability—United States. 6. Educational change—Government policy—United States. I. Gorski, Paul C., author, editor of compilation. II. Zenkov, Kristien, author, editor of compilation.
LB2822.82.B595 2014
371.2'07—dc23
2013042694

ISBN: 978-0-415-70793-0 (hbk)
ISBN: 978-0-415-70794-7 (pbk)
ISBN: 978-1-315-88636-7 (ebk)

Typeset in Bembo
by Apex CoVantage, LLC

Printed and bound in the United States of America by Publishers Graphics, LLC on sustainably sourced paper.

THE BIG LIES OF SCHOOL REFORM

Finding Better Solutions for the Future of Public Education

Edited by Paul C. Gorski and Kristien Zenkov

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

6

THE TEST DOES NOT KNOW BEST

On Collecting Good Evidence for Student Learning

Deborah Meier

"It's lies, lies, and more lies," I mutter to myself as I read the media's usual account of the current school wars. At what point are we all witlessly drawn in by the winners of the war, the corporate interests, over how to tell the story? The problem is, the language of the winners becomes the language people use for responding to their lies. It begins to influence even me as I try to rebut increasingly familiar stories about "the decline of American public education" and how our public schools have aided and abetted the country's waning global economic competitiveness and standard of living.

So many lies fill the current "school reform" agenda, and they have a hold on the public. This has been true in some form or another for the past 60 years. I have two *Life* magazines, one from the 1950s and one from the 1970s, whose covers point to the "The Crisis of American Education." The first compares us with Russia, and the second focuses on the economy.

Once, when I challenged Al Shanker, former president of the American Federation of Teachers, for participating in the fabrication of these "crises," he agreed that it was nonsense. But after all, he argued, the more attention we pay to public schooling, the more funding we'll get, and that will be good for kids and teachers. Unfortunately, you reap what you sow, and what we're reaping now is one great big mess.

In contrast, my father used to say that if someone starts calling everything a "crisis," check your wallet and then check the other person's. And always follow the money.

One of the big lies misguiding today's conversations about strengthening public schools is that America's schools used to be really, really good, that they made it possible for the poor immigrants arriving on our shores or already living in our midst to join the great middle class or even the wealthier classes. Another big lie is

that our economy's past and future rested and continues to rest on the work of our schools—that we "educated" ourselves into being the central world power. So, if our global dominance is being threatened, it must be because our schools no longer are effective, that our public school system no longer is successful. And even if this story isn't totally true, the argument goes, we're heading toward shaky times in the future, so we better fix our schools before it's too late. Sounds scary, right?

All of these lies are made to appear like reasonable concerns by a system of measuring success and effectiveness that is, itself, a big con. The worst thing about our increasing reliance on standardized testing to measure "achievement" is that so few people, including well-educated people, understand how the system of testing we've chosen operates and why it will not—why it *cannot*—work, especially if one of our goals is equal educational opportunity. It's a system for rank-ordering children that, as I used to say when I blew my whistle at the end of recess, always produces exactly the same number of children who were first, second, third, and last in line. And, lo and behold, it always produces a predictable lineup, with the most privileged students at the front of the line.

The Ease and Implications of Buying In

It might interest you to know that I, too, once was fond of standardized tests. It was hard for me to overcome that fondness, even though I did so under ideal circumstances, where few stakes were involved in my success or failure. Besides, if Diane Ravitch could change her mind about tests, it must be possible to "convert" just about anyone.

The problem has been trying to convert a lot of anyones. Part of the problem is that we reconfirm our beliefs in tests all the time, even if ostensibly we abhor them. I see it all the time: people using test scores as their evidence when making a case for some better, more equitable approach. Maybe it's because determining what *does work* when it comes to creating and sustaining schools in which all students can thrive can seem like an overwhelming task. Perhaps some of us think, in the face of that task, that it would be so much easier if only we had a clear and concise numerical system for keeping score.

So first I want to tackle, in a rather personal way, how tempting it can be to imagine standardized test scores as a euphemism for "achievement." This view has been prominent for the past half century (and probably longer), even back when schools used far fewer standardized tests. The search for a supposedly neutral, objective way of assessing who is and isn't intelligent and who is and isn't well educated (the two often are conflated, which is part of the problem) became a national obsession as soon as we rhetorically broke with our historical class/caste rhetoric. Few and far between were people in the Western world in the 19th century who didn't assume that poverty and skin color weren't either the same as or closely tied to intelligence, brain power, and even moral and ethnical understanding. Richard Herrnstein and Charles Murray, after all, wrote a persuasive and popular book, *The Bell Curve,* only 20 years ago in which they claimed to prove scientifically that race was causally related to intelligence—and that the reason white and wealthy people scored at one end of the normal curve was a simple fact of biology.

Today, amid demands that *all* children achieve at high levels alongside denials that racism or poverty affects achievement, we've placed schools in a paradoxical position, relying more than ever on a numerical tool and complex numerical equations for measuring the intelligence of students and the effectiveness of schools and teachers. In the process we have forsaken human judgment, such as that which would allow much more robust assessments of student learning and teacher effectiveness and that which would help us to recognize that, despite Herrnstein and Murray's claims, *race* does not explain the *achievement* gap; rather, *racism* explains the *opportunity* gap. It's hard to measure that on a multiple-choice test.

I actually grew up in a relatively testless setting, although apparently my parents once had my brother and me tested for IQ. I don't recall the results.

I didn't question the idea of measuring intelligence at the time. I viewed it as a scientific way of sorting test-takers' answers into rights and wrongs and in doing so finding the deeper and broader patterns could help rank us all with regard to our intelligence. To my mind, science was good; I was predisposed to it. Even the progressives of the time, including John Dewey, accepted most of the rhetoric on scientific standardized psychometrics. Most progressive educators were not even leery of tracking, separating schoolchildren by "intelligence" on the basis of early IQ testing. Achievement tests grew from these IQ tests and followed the same psychometric rules in the interest, at least ostensibly, of giving parents, educators, and the larger society information that would help us provide education and training that matched students' abilities and society's needs.

This was confirmed to my satisfaction when I took a standardized test in order to transfer into the graduate history department at the University of Chicago. It even covered topics that I never had studied or been particularly interested in. Still, I did well. I had a "knack," it seemed, for choosing "right" answers. That "knack," I assumed, was "native intelligence."

Coming to See the Trouble with Testing

I assumed the same would be true of my own three children. So I was amazed and unprepared when my older son's third-grade teacher informed me that he needed remedial reading help. He was a fluent and sophisticated reader, devouring books right and left, so I was dumbfounded. I didn't give in and was labeled in what were then secret cumulative records as a "troublemaker." Later I found an old copy of a third-grade test and administered it to my son. We discussed his answers and his approach to taking the test. He insisted that it was only fair to cover up reading passages before trying to answer the questions about them, which I explained he didn't need to do. He also wrote little notes next to some

of his answers in which he explained why he chose what, he surmised, likely were not considered the "right" answer. Fortunately I convinced him not to do that on future tests, but not without a struggle over his own notion of what was morally right and wrong.

This experience coincided with my starting a new job teaching kindergarten in central Harlem. In my third year, a group of us, all viewing ourselves as "progressives," convinced the principal to let us transform the early-childhood wing of the school into a pre-K through second-grade cluster. Testing started at second grade at that time, so we were eager to see the results of our four years' worth of collaborative work. Oh, dear. They didn't match our expectations. Our students did not fare better or worse than other second graders. More startling, though, was our discovery that several youngsters we knew to be very good readers scored as nonreaders!

So we bought tape recorders and proceeded to interview students individually and in small groups. I wrote a pamphlet describing the results and began speaking everywhere about it. This was the late 1960s and early 1970s. I got a lot of applause and smiling faces, but New York City and the nation as a whole proceeded to require more testing, not less. It took a dozen or more years for my opinions about the ineffectiveness of these tests to be echoed widely by others. But by the 1980s, the anti-testing sentiment had become part of the educational mainstream. It wasn't the dominant part, but it was growing.

We even thought by 1990 that we had won! I remember Linda Darling-Hammond, one of our academic heroes, saying that Edward Thorndike—the early nineteenth-century educational psychologist and staunch advocate for quantifying, well, just about everything—had finally lost and that standardized testing was on its way out. Unfortunately, the reality is that we had blinded ourselves to another train of thought that was coming down the track. That train of thought was building steam and soon would bump us even further into the margins of the national conversation on what it means to be well educated.

The renewed and far more vigorous and rigorous use of standardized testing that returned during those years made some of us forget that the testing focus wasn't a new phenomenon. Its pernicious effects had been apparent for a long time. Of course, given the increasingly conservative political and economic climate in which we were living, we probably shouldn't have been so surprised by the renewed efforts to infect public schools with more and more testing regimens. Tests were being pitched as the free-market answer to the schools' and even the nation's woes.

Occasionally I wonder how it could have happened that the likes of Ted Sizer, Al Shanker, and me ended up preaching together to largely converted businesspeople and industrialists about what the new wave of schools would and should look like: schools that knew their students through their work, not their test scores. "Maybe," I'd say under my breath, "what's good for General Motors is good for progressive education."

Turns Out We Were Wrong, and I Should've Known Better

We slowly began to recognize the tenaciousness of the underlying assumptions that test scores confirmed. What was handy was that, generally speaking, students' test scores correlated with our expectations—the rank order looked more or less right. There were occasional kids, like my son, for whom this wasn't the case, but, the thinking went, these kinds of aberrations were rare and could be ignored as outliers in the vast piles of data.

Of course, expectations always are colored with bias. I soon learned from colleagues at Educational Testing Service, one of the companies often contracted to create and deliver standardized tests, that if the correlations didn't hold up—didn't closely match IQ scores, achievement test scores, and general expectations—there'd be trouble. It would be a sign that psychometric errors had been made. Sometimes this meant merely reconsidering particular misleading items and replacing them with others in the pool that "worked" properly. If the rank order was "off," it meant that either the scoring rubrics were off or that there was some irregularity in the test setting, like cheating.

If an item was answered correctly more often by low-income Black children than by upper-class White ones in the field tests, it would be struck from the test. It was clearly an outlier. Something must be wrong with it, they figured, even though nobody quite could say what.

I also learned that traditional psychometrics placed scores along a normal curve, so that the median score was at the top of the curve and labeled "grade level." In fact, later, when the idea that properly designed tests had to fit neatly onto that curve—that half of all test-takers would always score below and half would always score above grade level—seemed unacceptable, other ways of benchmarking the rank-ordered scores replaced the normal curve. "Norming" remained, but the percentile labels that accompanied the curve as a way of recording one's status changed. We invented many new names that lumped scores to fit new paradigms: not meeting expectations, meeting expectations, going beyond, and so on. I thought, at first, that these new paradigms would make a positive difference as they didn't pretend to be quite so precise. In many localities today the scores are divided into four groups, so we have 1s, 2s, 3s, and 4s. Which refers to the highest scores and which refers to the lowest is determined by local authorities or by the state. But kids in many districts commonly became known as "4s" or "2s," including in high schools that accept only the highest-ranked students, for example.

The scoring was confusing, too. Teachers and parents often had no idea what a score of 5.7, for example, meant under the old system. Even when told (it was the norm for the seventh month of fifth grade), few teachers could even pick out a book that they thought was 5.7 without external aids. The new system seemed easier, but it still used a formula that relied on field-tested norms to decide what "grade level" meant. Noting of substance changed.

As you might imagine, explaining to parents who thought their children were good readers that these same children had scored poorly on standardized tests produced many a peculiar discussion. It led me to develop new ways to describe a child as a reader to her or his parent or guardian, always with that parent and that child in mind. We began to tape-record children reading from various reading-level-appropriate texts at the beginning and near the end of each school year. We, the teachers, met and listened to tapes to try to decide the language we'd use to describe a child at this or that stage and to discuss how confident we were in our assessments. We even invented ways to assign numbers to the stages we identified and to graph our results. (Parents loved this sort of ongoing "hard data.") The important thing was that it was coming from a full understanding of the student and not a narrow measure from a narrow test.

Building and Rebuilding a Movement against Test-Centric Schools

This way of thinking spread widely to many other subjects; we found ourselves freed from what seemed more and more like the vicious cycle of narrow assessments that standardized testing had created. As part of the Coalition of Essential Schools, we developed, in collaboration with other schools, variations on this, ways to assess when students were ready to graduate from high school and into the big wide world.

Many educators liked our work, but they often thought it was either "utopian" or "flaky." To those who mattered in the policy world, including the media, we were seen at best as "interesting" and at worst as "the enemy." The media enthusiastically hyped schools whose scores had jumped three "years" from one school year to the next but rarely noted when the populations of these schools had changed. Nor did they note that, yes, of course, Stuyvesant High School's students did well on tests. After all, they were at Stuyvesant *because* they were at the very top of the test-score heap. Why couldn't all schools be like Stuyvesant, they asked. They pointed out that Stuyvesant had high-scoring students of every race, class, and ethnic background. They rarely got around to pointing out that those students had to pass a test just to be allowed to attend the school, that it didn't practice open enrollment. Despite all the "miracle" schools acclaimed by the mayor of New York or by the media, it never was disproved that tests were peculiarly sensitive to matters of race and class. These miracle test score gains rarely lasted for more than one year, and rarer still were those that were schoolwide, across race and class.

Many people who were reluctant to change how we measured success viewed what we were doing as a form of affirmative action, which itself was growing less popular. Particular scorn seemed to be pointed at immigrants, and especially English-language learners. We heard insistent claims from third- and fourth-generation Americans that their forebears in the United States had done well, enabling future generations to progress. Everybody seemed to forget that, in point of fact, until World War II most immigrants with low-income roots did not do well at school, most certainly didn't graduate, much less go to college. And they surely did not test well. Jews, for example, had lower-than-average IQ scores as a group when my parents immigrated, as did Asian immigrants. For most groups, economic improvement came first, and then, later, the children did better in school.

By the early 1990s we realized that our advocacy for a different approach had failed and that schools had become more committed to high-stakes standardized testing than ever—more so even than China, where standardized testing had been used for centuries. Scores came to determine a child's future in school, her opportunities to participate in certain prized activities, what school she could attend, whether she could graduate or attend college. Projecting all of this further in the future, one could argue that test scores helped determine somebody's ability to get a job that might pay enough to raise a family. More and more, everything came down to test scores.

Today the reach of these tests is even greater. We have school consolidation and closings based on test score data. Schools are being turned over to publicly funded, privately managed, and for-profit organizations because they did not meet Annual Yearly Progress criteria on test scores. Test scores are being used to justify the elimination of tenure and the firing of entire schools of teachers and administrators, almost exclusively from high-poverty schools with large percentages of students of color. The newer scapegoats include schools of education that prepare teachers for their futures in the classroom. Even in these institutions, filled with professional educators, a majority of whom are, at the very least, suspicious of the high-stakes testing craze, standardized tests are used to decide who should be admitted. Many teacher-education programs exist under the threat of decertification by state agencies if a certain percentage of their graduates do not perform well on the Praxis exams, assessments used to certify teachers.

The combination of these measures likely accounts at least in part for shifting teacher demographics. For example, Chicago's teaching population, which once was 44 percent Black and Hispanic, now is only 19 percent Black and Hispanic.

So Now What?

The future rests, however shakily, on us—"the people." It's not written in the stars how this will all turn out. I do not think we can assume, as I once naively did, that the interests of big business and stockholders are in sync with the interests of all children. In fact, I fear now that quite the opposite is true. I think, for example, that the civil rights advocates who originally bought into No Child Left Behind are gradually seeing more clearly where that paradigm and its obsession with testing lead us. Many have shifted their loyalties and advocacy. Certainly, a growing number of parents are actively opposing the effects this approach is having on

the curriculum in their children's schools, the kind of work being rewarded in classrooms, and the increasing tendency to rank-order children and schools. Even students are beginning to object more noisily.

In a high school in Seattle, students refused to take one of the annual tests. By doing so, they made a clear statement about what was happening in and to their educational lives. Other schools followed. Diane Ravitch launched a nationwide organization, the Network for Public Education, to bring together the growing number of activists in local communities disturbed by how these assessments were being used to literally close schools in order to replace them with semiprivate charters or to push vouchers.

Perhaps the most exciting development, though, is that school boards, even in places like Texas and Florida, not known to be bastions of progressivism, are wary these days about who is deciding what when it comes to public schools. Business interests and billionaire foundations are spending millions of dollars to elect school board members who will enact business-friendly policies. You read that correctly. They are spending that money on local *school board elections.* But they are often being opposed and beaten by people with far less means.

In my view it's a little early for those of us who oppose this corporate-backed imposition of high-stakes testing on public schools to build a single coherent message, and maybe that's just as well. It's starting to come together, though. People are less cowed now when asked, "What's your alternative?"—which is good. But while we explore alternatives, we can stop the testing that is doing damage to students right now—tests on the basis of which high-stake decisions are made that affect individual people and institutions. We should not excuse the effects of a medicine that we know harms patients because the industry hasn't yet produced a better one that works (and is as cheap).

We are beginning to be able to point to counterexamples. Finland, for instance, did us a big favor by having high scores on international academic measures even though *it doesn't use standardized tests at all* between kindergarten and twelfth grade. Internationally normed tests, such as our one current national test (NAEP, the National Assessment of Educational Progress), do not test each and every child, nor do they report scores by individual students or schools. And there are alternatives for providing "scores" to individuals, and they are neither new nor untested. They involve the development of processes based upon the judgment of adults whom we collectively trust to decide whether Rosanna and Johnny can read. In these processes, the scores are the end result of collective decision making based on a wide range of evidence. We need these sorts of exhibitions of competence, as Ted Sizer called them, which we already use in most arenas of life. These exhibitions "show off" the skill, talent, and knowledge that demonstrate what we "mean" by being "ready to vote," "ready to serve on a jury," ready to take on complex adult tasks. They are like our "road test" for drivers, which I'm sure we'd all agree is more reliable and valid than a multiple-choice standardized test on driving. There is no lack of films, books, and articles describing these assessment approaches in detail—just visit FairTest.org for a reading list.

But just as important as reimagining assessment is joining together to discuss ways to defend public education itself—to work through why we're "incarcerating" our youth during their critical formative years in mandatory, overly standardized, testing-centered schools. We need clear language to describe what public education ought to be. We will disagree, of course, and when we do, *vive la différence.* But let's ask ourselves this: What values, beliefs, skills, habits of heart and mind, and aptitudes do we believe are essential to sustaining the most precious parts of who we are individually? Which equip us to inch closer and closer to a more authentic form of democracy?

Interestingly, most private and parochial schools of the past were built upon statements of commitment to these sorts of goals; many have mission statements to this effect that go back a century or more. Meanwhile, the mission of public education has been watered down to what we can measure cheaply on standardized tests, so that when somebody says "achievement" people automatically think "test scores." Everything else is made to seem soft and fuzzy. If our goal is to nurture citizens who are defined, at least in part, by their ability to exercise judgment in the town hall or the local voting booth, we cannot sit back and rely on some quaint practice of ye olden days to get us there. Nor will we nurture such citizens by handing them over for 12 or 13 years to adults who are forbidden to freely exercise their professional judgment about how to engage young people. As Ted Sizer has said, speaking as a parent, we want the right to look into the eyes of those who are making important decisions about our children.

Democracy flourishes when all of its institutions reward the exercise of judgment and when they all build their reputations on their past records of wise judgments. It remains an exciting and fragile idea to which to hold ourselves and our leaders accountable. Democracy will flourish only when we see it not only as an end goal or as the subject of a Sunday sermon but also as an essential means for carrying out the most important of all tasks: together raising the next generation.