

The 13th Bracey Report on
The Condition
Of Public Education

This year's Bracey Report begins and ends with the item that will continue to dominate education news: No Child Left Behind.

BY GERALD W. BRACEY

TO PRODUCE last year's report required me to sift through 2½ drawers full of materials. This year, there were 3½ drawers. Attention is being paid to public education; much of it is not benign. On the education news front, budgets and No Child Left Behind (NCLB) dominated. Because readers certainly know the budget situation in their own states, we begin with NCLB, in two parts.

NCLB: THE AYP TRAP

Last year's report said of No Child Left Behind, "[It's] a trap, a Trojan Horse . . . choose your metaphor." These days, Americans speak mostly in war images, so I now call NCLB a weapon of mass destruc-

GERALD W. BRACEY is an associate for the High/Scope Foundation, Ypsilanti, Mich., and an associate professor at George Mason University, Fairfax, Va. His most recent book is On the Death of Childhood and the Destruction of Public Schools: The Folly of Today's Education Policies and Practices (Heinemann, 2003). He lives in the Washington, D.C., area.

tion targeted at the public schools in a campaign of shock and awe, which, given the incredible under-reaction of educators, I must conclude is working. I first expressed that sentiment in an article that no one would publish in April 2002. (The article can be found at www.america-tomorrow.com/bracey/EDDRA.) No one seemed to believe it then. They do now.

At the April 2003 meeting of the American Educational Research Association (AERA), several employees of the U.S. Department of Education approached me to express dismay over NCLB and over working for an ideologically driven department. These are sad times, indeed. One career staffer declared that NCLB should be renamed: NCL-B.S.

As all *Kappan* readers surely know by now, all schools must test all students every year in grades 3 through 8 in reading and math, with science to be added in 2006-07. Schools must demonstrate adequate yearly progress (AYP). For a school to show AYP, all ethnic groups, all major socioeconomic groups, English-language learners, and special education students must make AYP separately. Ninety-five percent of each group must be tested, and, if any one group fails to make AYP, the school as a whole fails. The official phrase is “needs improvement,” but headlines across the nation reveal how everyone actually thinks about it: “Most Schools in State Failing,” *Los Angeles Daily News*, 24 July 2003; “State Adds 544 Schools to Failing List — for Now,” *Grand Rapids Press*, 12 July 2003; “It’s Pass

or Fail; All or Nothing,” *Raleigh News Observer*, 13 July 2003.

Schools must continue to make AYP until, by 2014, 100% of a school’s students must score “proficient.” In last year’s report, I contended that, while the law allowed each state to define “proficient,” that wouldn’t last and that the NAEP (National Assessment of Education Progress) definition of “proficient” would come to rule. I claimed that this would happen even though the NAEP achievement levels have been rejected as fundamentally flawed by everyone who has ever analyzed them. This includes the General Accounting Office, the Center for Research in Evaluation, Student Standards, and Testing (CRESST), the National Academy of Education, and the National Academy of Sciences.

Even so, in his presidential address to AERA, Robert Linn, co-director of CRESST at the University of Colorado and UCLA, made no attempt to provide or use any definition of proficient other than that rendered by NAEP.¹ Educational researchers now assume that proficient means NAEP proficient. Secretary of Education Rod Paige has said that he will use the discrepancy between the performance on NAEP and the performance on state tests to shame schools into better performance.²

Certainly those discrepancies will be there for Paige to point to, and in many instances they will be huge. To see why, we need only look at the Princeton Re-

THE GOLDEN APPLE AWARDS

The Accountability, Yes! Award goes to Superior Court Judge Howard Manning, Jr., of Wake County, North Carolina. In most accountability programs, the state holds the kids and maybe their schools accountable for passing tests in order to get promoted or to graduate from high school or to stay accredited. So it was in North Carolina, until Manning upset the accountability apple cart by declaring that, if kids score low on North Carolina’s tests, the state of North Carolina should be held accountable.

The suit that led to Manning’s decision started with four poor school districts claiming the state had failed in its responsibility to educate the children. Manning concurred:

The public schools have no choice but to shoulder the burdens of these at-risk children and are expected to provide them with the equal opportunity to obtain a sound basic education. This is so because, in the eyes of the law, these at-risk children are citizens of North Carolina, and, as citizens, they have the same constitutional rights as those children whose parents are married, have decent educations, earn a decent living, and provide their children with a learning-rich environment from the day they come home from the maternity ward.

The state has filed an appeal with the North Carolina Court of Appeals, claiming that its tests should not be used to hold the state government accountable. The key findings of Manning’s ruling can be found on the website of *Rural Policy Matters*, <http://ruraledu.org/rpm/rpm405a.htm>.

view's comparisons of performance on state tests with performance on NAEP.³ Ironically, the largest discrepancy is for Texas, the home state of Secretary Paige and President Bush. Texas declared 91% of its eighth-graders proficient in math. NAEP says only 24% of Texas students reach that level. Although there are some states where the differences are small and even four states that have tougher requirements than NAEP (Arizona, Louisiana, Missouri, and Maine), the average difference is 23%, and for 22 states, the difference exceeds 30%.

In most states, somewhere between 20% and 30% of the students test proficient on NAEP. (The range is from 8% in Mississippi to 40% in Minnesota.) So what will it take to get 100% of students to the proficient level?

In his presidential address, Linn first asked how long it will take to get 100% proficiency if we continue to improve at the same rate we have improved for the last decade. Looking at mathematics, Linn observed that, at the last decade's rate of improvement, we can attain 100% proficiency in the fourth grade by 2056, in the eighth grade by 2060, and in the 12th grade by 2166.

Linn then asked how much we have to increase our rate of improvement to reach 100% proficiency by 2014. He found that we must ratchet up improvement at grades 4 and 8 by a factor of four. At grade 12, we must increase our rate of improvement by a factor of 12. The projected improvements are shown in Figure 1.

As Linn also said at AERA, just getting 100% of the students to the basic level would constitute an enormous challenge.

NCLB: THE HIGHLY QUALIFIED TEACHER TRAP

While most people are biting their nails over AYP, their anxiety is also mounting over NCLB's requirement for "highly qualified teachers." All teachers hired after the start of the 2002-03 school year had to meet this requirement, and all teachers (not just new hires) must meet it by 2005-06. New York City, almost defiantly it seemed, announced in June that it had hired 3,000 underqualified new teachers. Maybe it was just the district's way of pointing out the ludicrousness of the requirement. According to NCLB, "highly qualified" means a person must

- hold at least a bachelor's degree from a four-year institution;
- hold full state certification; and

- demonstrate competence in the subject areas taught.

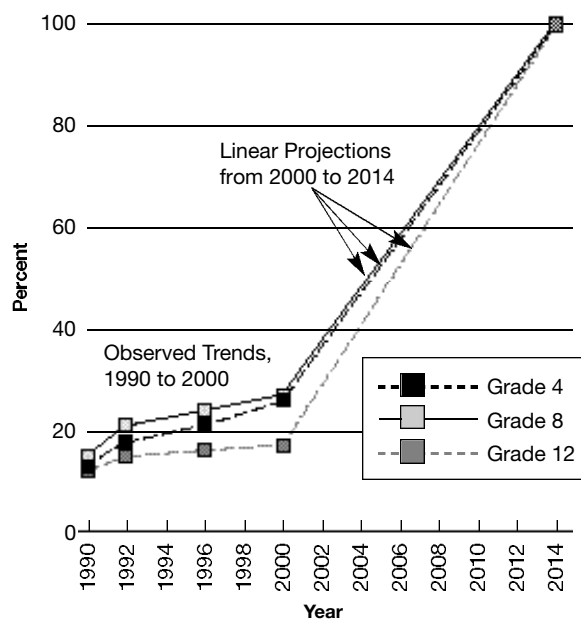
The demonstration of competence involves passing a "rigorous state test" for elementary school teachers and can involve the same for middle and high school teachers. Middle and high school teachers can also demonstrate competence by majoring in the subject they will teach, by earning a graduate degree in that subject, by accumulating coursework equivalent to a graduate degree in that subject, or by attaining an advanced certificate.

A recent report from the U.S. Department of Education lays out all this and more. The document estimates that in English, math, science, and social studies, no more than 50%, 47%, 55%, and 55%, respectively, of the teachers currently in the schools meet the criteria for being highly qualified.⁴

To get to 100%, the department is pushing a "higher standards, lower barriers" approach to teacher certification. It notes, rather coyly it seems to me, that NCLB defines explicitly what "highly qualified" means

FIGURE 1.
NAEP Mathematics Trends with Projections to 2014

(Percentage Proficient or Above)



Source: Robert L. Linn, "Accountability: Responsibility and Reasonable Expectations," presidential address to the annual meeting of the American Educational Research Association, Chicago, 23 April 2003. Used with permission of the author.

THE GOLDEN APPLE AWARDS

The Investigative Reporting May Yet Live Up To Its Name in Education Award goes jointly to Michael Winerip of the *New York Times* and to Doug Oplinger and Dennis Willard of the *Akron Beacon Journal*. Many of us were stunned and dismayed when the *New York Times* dropped Richard Rothstein as its education columnist. He is still missed. For a while, various *Times* reporters wrote each Wednesday's column, but then the paper settled on Michael Winerip. It was a great choice.

Winerip doesn't so much bring to the task Rothstein's penchant for analytic reasoning as he does feisty, old-fashioned, let's-find-out-what's-really-happening-here journalism. Winerip exposed the New York Department of Education as once again sanitizing literary passages for its Regents exams.¹ He did so courtesy, once again, of Jeanne Heifetz, the literature-loving mom from Brooklyn who discovered the bowdlerization the first time. (See "The Get Thee to a Nunnery, Rick Mills Award," from the 12th Bracey Report, in the October 2002 *Kappan*, p. 138.)

Winerip also made public the fact that Paul Peterson had omitted more than 40% of the data in his study of vouchers in New York.² His article on the 70% failure rate on New York's Math A Regents exam led the New York Department of Education to toss out the results — and to toss out an expendable assistant commissioner, as well.³

Repeat winners Oplinger and Willard couldn't deliver an opening line in 2003 as potent as last year's "In Ohio, businesses now gets more property tax breaks than God." But they continued to investigate and reveal political and financial shenanigans and point to warped fiscal priorities in Ohio.⁴ They noted how entrepreneur David Brennan and the Catholic Conference of Ohio pushed a \$10.5-million expansion of Ohio's voucher program — in spite of the fact that the test scores of Cleveland's public schoolers improved much more from first grade through third grade than did those of the voucher recipients. As predicted last year, businesses now exceed public schools (previously number one) in the value of tax-exempt property they own.⁵ We need more of these guys.

1. Michael Winerip, "How New York Exams Rewrite Literature (A Sequel)," *New York Times*, 8 January 2003, p. B-7.

2. Michael Winerip, "What Some Much Noted Data Really Showed About Vouchers," *New York Times*, 7 May 2003, p. B-12.

3. Michael Winerip, "A 70% Failure Rate? Try Testing the Testers," *New York Times*, 25 June 2003, p. B-9.

4. Doug Oplinger and Dennis Willard, "More Money for Vouchers," *Akron Beacon Journal*, 29 June 2003, p. A-1.

5. Doug Oplinger and Dennis Willard, "Over Last 15 Years, Businesses Pay Less and Individuals Pay More," *Akron Beacon Journal*, 8 June 2003 p. A-1.

but is silent on what it means to have full state certification. Later, the department's report lists a number of "promising" alternatives to full state certification, including Teach For America and Passport to Teaching Certification, a joint project of the department, the Education Leaders Council, and the American Board for Certification of Teacher Excellence.

The document mentions that "the research on Teach For America is encouraging. Analysts at the Hoover Institution found that in grade three through five, Teach For America members elicited greater achievement gains among their students" than did other new teachers or all other teachers in Houston.⁵ Not really. Using other new teachers, only the math comparison was significant (at the .05 level), not the reading comparison. And using all teachers in Houston, neither the math nor the reading comparison was significant.

Of course, given that Houston is an urban district, a sizable portion of new teachers these days are not certified and don't have bachelor's degrees, while all TFA teachers have at least a bachelor's degree (something noted in the report and then, oddly, denied by it).

Despite the report's glowing words about TFA and despite TFA leader Wendy Kopp's close personal ties to George W. and Laura Bush, in July 2003 TFA's \$14-million funding from AmeriCorps got axed. An article by Joe Klein in the 17 August 2003 issue of *Time* magazine quoted the form letter Kopp received: "We regret to inform you that your application was not selected for funding." While other programs faced reductions, TFA was simply zeroed out.

Note that, from the preceding paragraphs, it's clear that the U.S. Department of Education supports a rather right-wing agenda, presents "facts" that are not so, and

THE GOLDEN APPLE AWARDS

The Jonathan Swift Lives Award goes to Don Orlich of Washington State University. While many people used 2003 to reflect on the 20th anniversary of *A Nation at Risk*, Orlich donned the mantle of the satirist and rewrote the report for *Teachers College Record* online. Some of the Orlich Report simply uses the actual words of *A Nation at Risk*. For the rest, Orlich changed a few nouns and verbs to make it more germane to the nation's risk today, as ex-CEOs go to jail, the unemployment rate soars above 6%, 401Ks become 201Ks, and in two years the Bush Administration turns a projected \$5.6-trillion budget surplus into a projected \$4-plus-trillion deficit (quite an accomplishment). Here is a sample of Orlich:

Our Nation is at risk. . . . What was unimaginable a generation ago has begun to occur — our captains of industry and Wall Street investment bankers are creating economic chaos and unprecedented decline in market capitalization. If an unfriendly foreign power had attempted to impose on America the mediocre and unethical economic performances that exist today, we might well have viewed it as an act of war.

The entire satire can be found at www.tcrecord.org; search on “Orlich.”

accepts a study that is hardly an apples-to-apples comparison and thus does not come close to the department's own gold standard of “scientifically based research.”

But it gets still worse. The document misstates a conclusion by Dan Goldhaber and Dominic Brewer.⁶ The document claims certification is unrelated to achievement (p. 44). Goldhaber and Brewer actually examined the impact of advanced degrees on achievement. It further states, “The bulk of evidence on this policy [incentives for advanced degrees] is that there are no differential gains across classes taught by teachers with a master's degree or other advanced degree in education compared to classes taught by teachers who lack such degrees.”⁷ There is not enough literature to constitute

a “bulk,” and the department's report cites none of it.

It argues instead that content knowledge matters, and it cites a study by Brian Rowan to the effect that experience counts, too. It doesn't mention that in that same study Rowan found that “students who were taught by a teacher with an advanced degree in mathematics did worse than those who were taught by a teacher not having an advanced degree ($d = -.25$).”⁸ An effect size of $-.25$ is one that most researchers feel has substantial practical significance.

The document is selective, sloppy, and incomplete. Instead of revealing what the research says, it pushes, not very subtly, a particular view of what is important in teaching. As a final example, it says, “There is little compelling evidence that certification requirements, as currently structured in most states, are related to teacher effectiveness.” It cites all of two articles as evidence, one of which does not support the statement. This is unfortunate.

The document summarizes in a theoretical graphic how much various teacher characteristics affect student achievement. “Cognitive ability” (undefined) dominates as the most important characteristic — more than twice as important as “focused training.” Focused training, in turn, is twice as important as either experience or content knowledge, which are equal in importance. Experience and content knowledge are four times as important as certification or master's degrees. Lowest on this ladder are workshops, which are hardly visible. I daresay no “scientifically based research” knowledge base undergirds this model.

VOUCHERS

After the Supreme Court ruled the Cleveland voucher program constitutional and after the Republican gains in the 2002 election, some predicted a flurry of voucher activity. It didn't happen. Enormous state budget deficits killed proposed voucher programs. Texas, Louisiana, and Colorado considered statewide programs, but only Colorado passed one, and Colorado's is a bit different from any existing program. It applies only to Denver (with some opt-in conditions elsewhere), but its major difference is that only students who perform poorly on certain state tests become eligible. Some wags claim this will cause students to deliberately act dumb when taking the test so they can get a voucher — worth 75% of the state's per-pupil expenditure — to go elsewhere.

Early in the summer, the Manhattan Institute re-

requested that I be on a panel with Jay Greene to debate whether vouchers would be a good thing for the District of Columbia. I agreed. Greene's general thrust was similar to what he had argued in a *Wall Street Journal* op-ed several months earlier, the principal purpose of which was to attack Alan Krueger's reanalysis of Paul Peterson's data from New York City.⁹

In his op-ed essay, Greene also contended that rigorous evaluations had been conducted of voucher programs in Charlotte, Dayton, Milwaukee, New York, and the District of Columbia. "None of them finds students harmed by receiving a voucher." By that, he meant test scores had not declined.

This was not wholly true. In New York, test scores did decline for ethnicities other than black. In Dayton, some grades showed declines even for blacks, and test scores for blacks had declined in one year of the D.C. program. It is significant that Greene did not mention Cleveland, where public school students started behind voucher recipients but had caught and overtaken them in math and had almost closed the gap in reading and language. (In Cleveland, budget woes had forced the city to abandon all-day kindergarten; the cost of the voucher program would have permitted its restoration.)

We will consider Milwaukee here only in summary, as it has been dealt with extensively elsewhere. For instance, Princeton's Cecilia Rouse, perhaps the most rational analyst of the Milwaukee data, found a positive effect for voucher students in math, but not in reading.¹⁰ She also found that public school students in small Milwaukee classes outscored the voucher kids.¹¹ I consider the various analyses of Milwaukee to be exercises in frustration because the data are so poor. The General Accounting Office (GAO), in its review of Cleveland and Milwaukee, also said of Milwaukee, "None of the findings can be considered definitive because the researchers obtained different results when they used different methods to compensate for weaknesses in the data. . . . [The studies] suffered from missing test score data, low survey response rates, and the loss of students from program groups and comparison groups over time."¹²

Dayton and D.C., as noted, produced inconsistent outcomes across grades and over time. In New York City, William Howell, Patrick Wolf, Paul Peterson, and David Campbell purported to find a positive effect for black students.¹³ Although they have no explanation for why vouchers should improve black students' achievement and not that of any other group, they

have taken to reporting results solely by ethnicity. This is problematic since there is considerable variation from grade to grade. Indeed, in New York the effect for black students was produced solely by an increase in test scores in the sixth grade. The sixth-grade growth was large enough to make the average for black students for all grades statistically significant. "An average is an average," Peterson told the *New York Times*.¹⁴ As will be seen below, even the claim for sixth grade is suspect.

The GAO examined the results from New York, Dayton, and D.C. and gave this summary conclusion:

The New York Study found consistently greater improvements in math and reading achievement for African American elementary students using privately funded vouchers. Vouchers in Dayton showed no significant improvements in reading or math test scores. The Washington, D.C. study demonstrated positive effects for African American students in the second year of the study, but these disappeared in the third and final year of the study.¹⁵

Not exactly a ringing endorsement. Indeed, the GAO analysis contains a footnote reporting that a reanalysis of New York data by Alan Krueger raises "doubts about the size and significance of earlier findings" (p. 16). The Krueger analysis apparently arrived too late to be incorporated into the body of the text.

In Charlotte, Greene himself conducted the evaluation and claimed to find positive effects.¹⁶ However, Greene's evaluation lasted only one year, which is a problem in that single-year gains in Dayton and D.C. had proved ephemeral. When I asked colleagues at the University of North Carolina, Charlotte, if there was more information, I was told that the voucher people were being very secretive about it. Since private dollars funded the vouchers, they argued, they didn't have to make the data public. Hmmm.

The Peterson-Krueger encounter was the most interesting voucher event of the year. In his *Wall Street Journal* article, Greene had accused Krueger of making "poor research choices" in his re-analysis of the Peterson data. Actually, it was Peterson who had chosen badly. Peterson had elected to omit more than 40% of the data from his analysis. These included large numbers of students for whom background data were incomplete or for whom no prior test scores existed. They also included some students whom Peterson had categorized as "white." In Peterson's system, if a child

had a black mother and a white father, the child was black. A child with a white mother and black father was white.

Worse than omitting these data, Peterson had not told anyone in any published report that he had done so. It was only when Krueger obtained the raw data from David Myers of Mathematica Policy Research, a co-investigator with Peterson on the New York study, that he discovered the existence of the missing data. When these data were added to the mix, the results ceased to be significant even for black students. The ethical standards imparted to me by my mentors require researchers to disclose such data manipulation in their reports.

Peterson and Howell, joined by Greene, claimed that the background data and test scores were needed in order for a student's score to be included.¹⁷ This was very curious. As Greene observed in the *Wall Street Journal*, the students in New York had been randomly assigned. Earlier, Greene and Peterson had written, "Analysis of randomized experimental data does not require controls for background characteristics or test scores. Such controls are necessary only when one doubts that the experimental data are truly random."¹⁸ No one has expressed any doubts about the randomness of the New York data.

For his part, Mathematica's Myers called Krueger's "a fine interpretation of the results." Of Peterson's study, he went on to say, "it is not a study I'd want to use to make public policy. I see this study, and I go 'Whoa.'"¹⁹ Myers' whoa did not bring Peterson to a halt. The man who once referred to himself and other voucher advocates as "Jedi attackers" galloped to a press conference at the National Press Club and presented a paper accusing Krueger of "rummaging theoretically barefoot through the data in hopes of finding desired results."²⁰

As for Cleveland, an ongoing evaluation by Kim Metcalf and his colleagues at Indiana University found that public school students had started first grade with test scores below those of voucher students. By the end of third grade, the public school students had narrowed the gap from 14 points to just 3 points. In language arts, the gap closed from 11 points to 5, and in math the public school students had not only closed

a 9-point gap but had overtaken the voucher students to lead by 2 points. Said the evaluators, "The most recent results do not reveal any significant impacts of participation in the voucher program on student achievement."²¹

Despite this lack of impact, the Catholic Council of Ohio and David Brennan, author of the Cleveland voucher plan, pushed through — literally in the dark of night — a \$10.5-million voucher expansion.²²

All of this dancing around about whether vouchers work is, of course, largely irrelevant to the larger aim of voucher advocates: to privatize the public

schools. The grandfather of the American voucher programs, Milton Friedman, titled one of his articles "Public Schools: Make Them Private."²³ William Bast, CEO of the Heartland Institute, has said that vouchers are merely a temporary expedient on the way to full privatization. The efforts to privatize public schools and the motives behind the efforts to do so have been well summarized in *The Voucher Veneer*.²⁴

The various Peterson papers contain acknowledgments of who funds the research. Mostly they are the same right-wing foundations that paid Charles Murray a million dollars for *The Bell Curve* and bestowed \$2 million on David Horowitz' Center for the Study of American Culture.

The principal force behind the privately funded voucher programs is Children First America, which began life as the Children's Educational Opportunity fund or CEO America. If one goes to CFA's website (www.childrenfirstamerica.org) and clicks on "links," one finds a list of more than 20 right-wing organizations, such as the Manhattan Institute, the Institute for Justice, the Mackinac Center, the Goldwater Institute, and the Heritage Foundation, which *Slate* editor Michael Kinsley once called "a right-wing propaganda machine masquerading as a think tank."²⁵

Of course, the privatization fox is now in the public education hen house. Eugene Hickok, who as secretary of education in Pennsylvania helped bring Edison Schools, Inc., to that state and tried repeatedly to get vouchers installed, is acting deputy secretary of education. Nina Shokraii Rees has brought her ideology from the Heritage Foundation to the U.S. Department of Education as deputy under secretary for in-

All of this dancing around
about whether vouchers work is, of
course, largely irrelevant to the larger
aim of voucher advocates: to
privatize the public schools.

novation and improvement. From this perch, she doles out an “Innovation of the Week.” She awarded one to Bill Bennett and his K12, Inc., which, according to Rees, provides a “world class” curriculum. Since the curriculum has no research base, much less a “scientific” research base, one wonders how she comes to judge it as world class.

CHARTER SCHOOLS

The late 1990s produced a flurry of charter school evaluations.²⁶ Then the number of studies dropped to near zero. The reasons seem to be a lack of interest coupled with a lack of money. The 1990s evaluations found that charters looked very much like regular public schools. Their advocates had not, by and large, used them to produce “laboratories of innovation” that would serve as models for other schools. Nor had charters achieved their principal promise: an increase in student achievement. One evaluation contended that the absence of novelty in charters stemmed from the risk of a real not-with-my-child-you-don’t backlash if they deviated too much from what parents perceived as a “school.”²⁷

This year, though, has seen another deluge of charter school data. Three studies looked at charters in California, while two others examined outcomes in selected charter-active states. Two of the California studies produced contradictory results, an outcome that appears to stem from methodological flaws in one of them.

The most comprehensive of the three California studies emanated from the RAND Corporation.²⁸ RAND compared traditional public schools to the two charter types that California permits: start-ups and conversions. For the first six years of charter schools in California, start-ups and conversions grew equally fast. In

istrator's credential or to have a master's degree or a doctorate.

On 10 measures of behavioral or health problems, traditional public schools and charters differed little except that students in charters were twice as likely to drop out — 8% versus 4%.

The two studies of the “same” data that reached contradictory conclusions came from Margaret Raymond at Stanford's Hoover Institution and David Rogosa, a Stanford statistician.³⁰

Raymond examined data in a variety of categories from 1999 to 2002. Unlike the RAND researchers, she did not examine membership by various ethnic groups, but simply by minority versus majority in elementary, middle, and high schools. Minorities have been less represented in charter high schools than in traditional high schools, but not by much. In middle schools, minorities are consistently overrepresented, but not by much. Elementary charter schools, on the other hand, have been getting whiter each year since 1999.

Poverty is often confounded with minority status, and this might be why there are more students eligible for subsidized lunches in charter middle schools, while there are fewer in charter high schools and fewer each year in charter elementary schools. It appears that increasing whiteness in these schools is also associated with increasing affluence.

Raymond and Rogosa collide over test scores. Raymond examined test scores for three categories of schools: traditional, charter, and what she termed “local competitor.” These last were traditional schools operating in a district that contained at least one charter. For elementary and middle schools, traditional schools started in 1999 and finished in 2002 with the highest API (Academic Progress Index, an accountability index made up of test scores, sometimes called the Affluent Parent Index because of its high correlation with wealth). Local competitors had the lowest API.

When Raymond looked at API gains over time, a different picture emerged. The three school categories did not differ at the elementary level. Charters had the smallest gain at the middle school level and the great-

est at the high school level. Raymond writes, “The striking finding is that their [charter high schools'] average improvement is more than twice that of conventional high schools.”³¹ For some reason, she does not find it equally striking that gains in charter middle schools are less than half of those of traditional and local competitor schools.

When Rogosa examined the results only for students with large socioeconomic disadvantage, the gains favored charters slightly at the elementary school level, but noncharters showed even more dramatic gains at the middle and high school levels.

But it doesn't matter, according to Rogosa. Raymond's method is wrong, so her results are wrong. Rogosa starts with a grade-by-grade analysis. In the elementary grades, he finds some grades in which charter school students gain more and some in which noncharter (his term) students gain more. Grades 2 through 6 taken together show an unimportant gain advantage for noncharters, 96.3 points versus 94.3. In the middle grades, charters fare poorly. Grade-7 charter students

gained 17.9 points, while noncharter seventh-graders gained 53.1. Eighth-grade charter school students gained a mere 0.25 points, while their noncharter peers gained 38.9. Rogosa's analyses replicate Raymond's at the middle school level.

It is at the high school level that Raymond and Rogosa split. His analysis for grades 9-11 favors noncharters at all three grades. Together, noncharter students in the three grades gained 24.9 points; charter students, 14.8. When Rogosa examined the results only for students with large socioeconomic disadvantage, the gains favored charters slightly at the elementary school level, but noncharters showed even more dramatic gains at the middle and high school levels. Indeed, the “gain” for disadvantaged charters at grade 8 was -12.1 points.

Raymond's analysis started with the mean scores of schools, then averaged them across charters and noncharters. And therein lies the problem. When Rogosa conducted this analysis, he replicated her findings. But the problem with using a school mean as the unit of analysis is that schools vary in size. In such an analysis, individual student scores in a small school count for more than individual scores in a big school, and charter schools, as we saw above, are mostly small.

Here is an example from another arena that will clarify why Raymond's analysis was misleading. The average 2002 SAT verbal score for New Jersey was 498,

and that for Mississippi was 559. This produces an “average” for the two states of 529. This average makes no sense, though, because 80% of the seniors in New Jersey take the SAT, while only 4% of the seniors in Mississippi take it. Fifty-nine times as many kids sit for the SAT in New Jersey as in Mississippi: 71,163 versus 1,213. If we were to use the state as the unit of analysis, the 1,213 kids in Mississippi count as much as the 71,163 kids in New Jersey. To get a meaningful average, we must multiply New Jersey’s 498 by 71,163 and multiply Mississippi’s 558 by 1,213, add those two products together, and divide by 72,376 (the total number taking the test in both states).

This procedure yields an average of 499, quite different from the 528 that results when the two states are treated as if they had the same number of students taking the test. This applies to Raymond’s approach. Charter schools in California vary in size by a factor of 20. Raymond’s data are further distorted by the fact that not all schools designated as elementary, middle, or high have the same grade configuration.

In sum, Rogosa’s analyses show students in charter and noncharter elementary schools improving by about the same amount over time, while students in non-charter middle schools and high schools both gain more than those in charters.

A Manhattan Institute study by Jay Greene, Greg Forster, and Marcus Winters focused on “nontargeted” charter schools. These are charter schools not designed to serve a special population, be it mathematically precocious students or musically talented students or low-income students.³² Schools that had enrollments of many low-income students could be included as nontargeted if they happened to be in neighborhoods with many low-income students and did not make a special effort to recruit such students. The authors refer to their study as “national,” but they end up using a selected sample.

The Manhattan Institute team began with 15 states that appeared to offer sufficiently large samples after targeted schools were eliminated. These states were Arizona, California, Florida, Texas, Michigan, Ohio, Colorado, North Carolina, Minnesota, Pennsylvania, Wisconsin, Illinois, New York, New Jersey, and Indiana. Eventually, though, the last four states were eliminated because the database proved to be too small. The researchers also eliminated “virtual schools” because they had no real location and conversion schools, such as those in California, because they remain neighborhood schools in their student makeup, with perhaps

a few neighborhood students opting out and a few from outside the neighborhood opting in. Although data from 11 states are used to calculate the “national” result, only five states are reported individually.

The report is called a “working paper,” and the authors write, “A working paper is a common way for academic researchers to make the results of their studies available to others as early as possible. This allows other academics and the public to benefit from having the research available without unnecessary delay.” This strikes me as unmitigated nonsense because the data presented are so skimpy that there is no meaningful way the paper can be said to make the research available.

The researchers *tell* you what they did, but they don’t *show* you what they did. They dropped six states from the state-level analysis because they had insufficient data. They say. They present no figures, nor do they even tell readers what decision rules they used for including or excluding a state.

What they do show is of dubious value. Greene and Paul Peterson, his partner in a number of voucher

THE GOLDEN APPLE AWARDS

The Concerned Dad Doggedness Award goes to Martin Swaden of Edina, Minnesota. Swaden's daughter failed the test Minnesota uses to determine who graduates from high school. Swaden figured the best way to help her pass the next time around was to look at the test and her answer sheet and see which questions had stumped her. The state refused to supply the materials and told him to tell his daughter to study harder. Swaden persisted. The state refused. Swaden threatened to sue. Swaden is a lawyer. The state sat him down in a room with a state official, the test, and his daughter's answer sheet.

Swaden soon found a question that his daughter got right but that the scoring service, National Computer Systems, said she had gotten wrong. Then he found another. And another. Six in all. Enough to raise her above the cut score. When it was over, the state determined that errors by NCS had caused 47,000 Minnesota students to get lower scores than they deserved, 8,000 to fail when they should have passed, and 525 seniors to be unjustly denied diplomas.

The students sued. NCS admitted the error but argued against permitting punitive damages. It maintained that it was a one-time affair that couldn't happen again, found a scapegoat, and fired him. The judge concluded that NCS had a long history of shoddy quality control and had scrimped on hiring sufficient — or sufficiently skilled — employees to cure the problem in order to maximize company profits and make NCS an attractive takeover target. (Just after the error was acknowledged, the company was bought by the Pearson empire for \$2.5 billion; the CEO was said to have pocketed \$50 million.) The judge permitted the students to sue for punitive damages, and NCS settled. The kids collected more than \$8 million, with those who had missed graduation receiving \$16,000 each.

studies, are the only people I know who report data as "significant" at the .10 level. In this paper, none of the data are significant at that level, so one must assume that the inclusion of a statement on what .10

means is there to amplify the perceived importance of the few results they show that are significant at the .05 level.

According to the report, students in charters score three percentile ranks higher than students in traditional schools in math and two percentile ranks higher in reading. Wow. This isn't even a single item's worth of difference on most standardized tests. These results are significant at the .05 level, but that's only because there are more than 2,000 "cases" for reading and math each. The authors do not explain how 11 states with about 600 schools generate 2,000 cases, but the reader should recall that the likelihood of finding significant results varies with sample size. The larger the sample, the smaller the effect needed to attain statistical significance.

Apparently this report is not aimed at other researchers but at methodologically naive readers. What researcher would insult other researchers by declaring, "Because these results are statistically significant we can be very confident that the charter schools in our study did have a positive effect on test scores"?³³ Very confident? At the .05 level with huge samples?

At the state level, none of the results are significant for Arizona, California, or North Carolina. Not even at the .10 level. Only two of four results for Florida are significant, and those two are not subject-consistent across tests. The impact of charters is largest for Texas, and the impact on TAAS (Texas Assessment of Academic Skills) reading and math scores is significant. The TAAS no longer exists, and it is too soon to know if the charters are doing well on its replacement, the TAKS (Texas Assessment of Knowledge and Skills).

TESTING

As I sifted through the dozens of articles that I had pulled from the Net and from newspapers, magazines, and journals, one word kept emerging: turmoil. Test protests here, test boycotts there, arguments about the value of all the testing all over the place. And endless debates. High-stakes tests increase dropouts. No they don't. High-stakes tests produce improved learning. No they don't. The double-issue Winter-Spring 2003 *Fair Test Examiner* (www.fairtest.org) provides a good summary and overview of what the fuss is all about.

Among the most debated issues is whether or not the scores from high-stakes state tests generalize to anything else. Audrey Amrein and David Berliner contend that they don't. Margaret Raymond and Eric Hanu-

shek claim that they do. Martin Carnoy thinks they might.

Anthony Ralston, a computer science professor at SUNY Buffalo, penned “Next Disaster in American Education: Rising Test Scores” and offered a typical observation: “Improving education will always improve scores on well-designed tests. But when the central aim of educational change is just to improve test scores, improved education is seldom the result.”³⁴ Similarly, CRESST’s Eva Baker told the *New York Times*, “The most perverse problem with high-stakes tests is that they have become a substitute for the curriculum instead of simply a measure of it.”³⁵

A reasoned but largely overlooked analysis raises the question of whether or not state tests are aligned with state standards. Robert Rothman and his colleagues at Achieve, Inc., along with Lauren Resnick of the University of Pittsburgh, examined this alignment. States claim such alignment, and NCLB requires it. But Rothman and his co-authors can’t find it except at the most trivial level.

The researchers rated state test questions on the quality of the match between the content of the standard and the content of the question; on how well the cognitive demand of the question matches that of the standard (e.g., whether it requires a student to “analyze” or merely to “select”); on the level of challenge in the standard compared to the level in a set of questions; on the balance between the content coverage of the standard and the content coverage of items; and on the range of the standard’s coverage compared to the range of the items. They found, generally, that the items were aligned with the standards, but

the good news ends here. With few exceptions, the collections of items that make up the tests that we examined do not do a good job of assessing the full range of standards and objectives that states have laid out for their students. What is included and excluded is systematic: The most challenging standards and objectives are the ones that are undersampled or omitted entirely. Standards and objectives that call for high-level reasoning are often omitted in favor of much simpler cognitive processes — low- or non-inference questions in reading and routine calculations in math, for example.³⁶

As test-based accountability becomes more stringent, they continue, “schools and teachers will match their curriculum and teaching ever more closely to what is

on the tests rather than to what the standards say ought to count.” Ah, yes.

Some of the year’s events in testing are reported here in various Golden Apple segments. Others were described in the September 2003 Research column. Over the year, the victims of two of the more egregious crimes committed in the name of testing got their day in court — and won.

In Gwinnett County, Georgia, it will be recalled, teacher James Hope’s wife put six obsolete test items on a website to give parents some idea of what Gwinnett’s Gateway Tests were like. In reaction to her effort at spreading understanding, the county sought to

THE GOLDEN APPLE AWARDS

The Don’t Tell Rod Paige or Gene Hickok

About This Award goes to Gary K. Hart, former state senator and former secretary of education for California, and Duke Helfand of the *Los Angeles Times*.¹ At 59, Hart has returned to teaching and makes \$28,000 a year for four-fifths time, conveying history to ninth-graders. “In the last three months,” says Helfand, “Hart has discovered that it’s one thing to issue edicts about test scores and academic standards from the Capitol and another thing entirely to motivate 37 fidgety teenagers.” Do tell.

In fact, Hart disapproves of his former self. “When he was secretary of education, Hart helped oversee the system that governs how California teachers run their classrooms and what textbooks they use. Now Hart has his own students, and he is finding he wants freedom to innovate.” Do tell.

Hart has abandoned a chronological approach to history for intense study of a few topics. But, “ironically, he works in an educational atmosphere that is more regulated than he would like.” Helfand says that Hart has committed to teach for two years, but he is uncertain about what the future holds. Methinks it would be a good thing if he went back to Sacramento to undo some of his earlier handiwork.

1. Duke Helfand, “Mr. Hart’s Teachable Moment,” *Los Angeles Times*, 19 November 2002.

fire Hope and permanently suspend the teaching license of this former teacher of the year. During the two-year ordeal, Hope had his phone records confiscated, received multiple visits from the police, took a lie detector test, and was told that a police officer was on the way to his classroom to watch him administer the Gateway Test. The Georgia Professional Standards Commission decided that the act merited no more than a six-month suspension.

Georgia judge Gail Tusan thought that even that was too much. “Public policy dictates,” said Tusan, “that Hope, an experienced 17-year veteran who works directly teaching and evaluating the very students to be tested, be able to actively participate in the public debate regarding the test and share with the concerned parties the benefit of his hands-on experience with the

students, the test, and its administration.”

The district attempted to appeal. But the courts denied that, and in February 2003 the Georgia Professional Standards Commission voted to close the case. Incredibly, Sloan Roach, spokeswoman for the Gwinnett schools, said that the district did not agree with the rulings. “School system officials are disappointed,” said Roach.³⁷

In 1999, the Chicago Public Schools (CPS) suspended, then fired, teacher George Schmidt. The district also sued him for \$1 million, upped that to \$1.3 million and later to \$1.4 million. Schmidt had been so appalled at the quality of CPS tests that he had published them in *Substance*, the monthly newspaper that he puts out for Chicago teachers. The CPS suit claimed that it needed \$1.4 million to replace the published

THE GOLDEN APPLE AWARDS

The Breath of “Fresh Air” Award goes to Geoffrey Nunberg of Stanford University. When I saw a comment by reading czar Reid Lyon that the 3-year-old children of professional mothers have larger working vocabularies than “welfare mothers” of 3-year-olds, I put it out on some listservs as the silliest comment of the day. Lyon took umbrage and suggested that, if I consulted a study by Betty Hart and Todd Risley, it would get me “out of the bush league,” a puzzling statement to say the least.

When linguist Geoff Nunberg was made aware of Lyon’s comment, he also took notice that assistant secretaries of education Susan Neuman and Grover Whitehurst had made the same claim. Nunberg used their silly contention to construct one of his many informative and entertaining commentaries for National Public Radio’s “Fresh Air.”

On consulting the study by Hart and Risley, Nunberg and I both found that when 3-year-olds of professional mothers talk to their mothers, they use more words than do mothers on welfare when talking to their own 3-year-olds.¹ The words used when talking to your young child hardly form your “working vocabulary,” a fairly meaningless term itself. Of course, what you say when you talk to your 3-year-old is, for most people anyway, quite different from what you might say to your spouse or any other adult. If it weren’t, we’d be left with the absurd conclusion that professional mothers have working vocabularies of only 2,000 words. That’s how many they used when talking to their 3-year-olds.

Nunberg got on the phone to Hart, who told him that “welfare mothers used far larger vocabularies when talking to their friends, their older children, or the researchers themselves.” He points out, “When you think about it, in fact, the claim that any normally functioning adult could get by on a 3-year-old’s vocabulary is absurd on the face of things.”

Nunberg imagines that none of these “Administration education honchos . . . actually read the study — this has the sound of one of those third-hand factoids that are always making their way around the scientific grapevines.” (Whitehurst, however, makes the claim on the inside cover of the Hart and Risley book, although there he calls it the “spoken vocabulary.”) But Nunberg warns that such claims always carry a tone of condescension. “It’s easier to ignore people’s voices,” he argues, “when you’ve decided they couldn’t possibly have one.” Nunberg’s commentary, “A Loss for Words,” can be found at www-csli.stanford.edu/~nunberg/vocabulary.html.

1. Betty Hart and Todd Risley, *Meaningful Differences* (Baltimore: Paul H. Brookes, 1995).

items it argued were now useless. (Most of them were so awful that, if anyone was to be fired over them, it should have been the testing staff at CPS.) In 2003, CPS dropped the amount to \$500, and this is to be paid only if Schmidt loses all of his First Amendment appeals, which are still in progress. When a group dubbed the Curie Twelve — a dozen teachers at Chicago's Marie Curie High School — called the tests "flawed and invalid" and announced that they would boycott the tests this year, CPS dropped the tests altogether.

Full disclosure: I testified for Schmidt at his suspension hearings and joined FairTest and PURE (Parents United for Responsible Education) to file amici curiae briefs for Schmidt's First Amendment appeal.

Coverage of the Hope and Schmidt cases can be found in the *FairTest Examiner*, Winter-Spring 2003, "Resisters Win Legal Victories," and in *NEA Today*, May 2003, "High-Stakes Tests Spawn High-Stakes Lawsuits."

NAEP AND THE CITY: WHITHER THE 'TEXAS MIRACLE'?

For years, Washington, D.C., suffered a harsh glare from NAEP assessments. The smallest unit of analysis otherwise reported by NAEP was the state, but, as a separate entity, the District received its own treatment, and it did not stack up well against most states. In 2001, though, Congress, persuaded by the Council of the Great City Schools, funded a trial assessment in reading in grades 4 and 8 in five large urban districts: Atlanta, Chicago, D.C., Houston, and Los Angeles. New York City participated at the fourth-grade level.³⁸

In general, the five cities scored lower than the nation as a whole, and the white/black and white/Hispanic achievement gaps were larger than in either the nation as a whole or in central cities considered nationally. Scores for white students in grade-4 reading ranged from 221 in Chicago to 250 in Atlanta. Nationally, whites came in at 227, with central-city whites at 225. (All scores are for public schools only.) For

Secretary of Education Paige used the NAEP urban data to pen a misleading, illogical, and, in places, erroneous op-ed for the *Wall Street Journal*.⁴¹ Paige contended that white fourth-graders in D.C. “score the highest of any subgroup tested.” This was not true. White students in Atlanta earned that honor. He then said, correctly, that black students in D.C. scored 60 points lower than white students in D.C. What he didn’t bother to point out was that the black and white subgroups in D.C. are in no way comparable.

White students in the District scored fully 21 points higher than whites nationally (248 versus 227). Looks like the whites in the D.C. public schools constitute some kind of elite. Two other stats confirm their elite position: D.C. whites constitute only 3% of the sample, while D.C. blacks make up 88%. Even more telling, fully 28% of white fourth-graders in the District scored at the advanced level. Nationally, just 9% attain advanced status.

Paige uses the black/white discrepancy to argue for vouchers to give parents more choice. But the numbers could just as logically be used to argue that D.C. schools are wonderful — just look at how well the white kids do. More than three times as many whites in D.C. score at the advanced level as whites in the nation. Indeed, Paige himself, in a speech to the National Urban League the day before his article ap-

peared, seemed to feel that the black kids needed to get themselves in gear. “It’s not a product of their DNA. It is a product of effort,” he said. Standing before the Urban League, Paige did not mention choice or vouchers as solutions.⁴²

D.C. schools have enough problems. The black students in the District, who make up 88% of the sample, scored 10 points below blacks nationally and 39 points below whites nationally. These kids don’t need the secretary of education pandering to the Right with illogical arguments for choice.

EDISON: THE SAGA CONTINUES

While Edison’s stock hovered in the \$1.50 to \$1.75 range, Merrill Lynch, Edison’s number-one fan, announced it would no longer track Edison stock because of a lack of investor interest. Chris Whittle, Edison’s founder, announced that it didn’t matter because he planned to take the company private again. Investors announced that they were suing Edison because what Whittle said he’d pay for their shares was an insult, and Janice Solkov announced that she couldn’t take it anymore. The 50-year-old with 30 years of school experience began the school year as the principal of an Edison elementary school in Philadelphia. Urban reality hit the second day, when a teacher quit in tears. Over Thanksgiving, Solkov thought about it and said she would leave after Christmas.⁴³

In a *Washington Post* piece about her travails, Solkov wrote that “Edison accepted funding of \$880 per student — much less than the \$1,500 per student it had sought — above the regular district per-pupil expenditure. That meant that the full Edison model could not be delivered.”⁴⁴ That sure gave the game away, didn’t it? For a decade, Edison had promised its revolution for free — the wonders of an Edison education would not cost more than the districts’ regular per-pupil expenditures. It was a promise Edison never kept.

A NATION AT RISK TURNS 20

The “Paper Sputnik,” as it is sometimes known, enjoyed its 20th anniversary in April. Reactions were plentiful and predictable. In the April *Kappan* I opined that the document had been fundamentally wrong.⁴⁵ The conservative Koret Foundation, in cooperation with the conservative Hoover Institution, produced a 378-page report that said the document had been, and remained, fundamentally right.⁴⁶ The report listed the

usual litany of selected and spun statistics, as one would expect from Koret's task force on the topic: John Chubb, Bill Evers, Chester Finn, Eric Hanushek, Paul Hill, E. D. Hirsch, Caroline Hoxby, Terry Moe, Diane Ravitch, and Herbert Walberg.

Elsewhere, Pete Du Pont saw "no progress at all," and for Du Pont that meant we need a free-market school system.⁴⁷ *Education Week* presented the most extensive anniversary coverage in its edition of April 23. In it, three reporters/editors ruminated on their high school days, which took place about the time *A Nation at Risk* was being prepared. *Education Week* also conducted some focus groups with current high-schoolers. The results of both are interesting. John Goodlad and Ted Sizer also contributed thoughtful essays to that same issue of *Education Week*, in which they put the report into the contexts of both what preceded it and what followed.

Goodlad said *A Nation at Risk* did what the commissioners were charged with doing, but that "we should have assembled a task force with a charge and resources at least comparable to those that launched our exploration of outer space." In fact, the report kept the Reagan Administration from eliminating some funds for the National Science Foundation, but, as then-Secretary of Education Terrel Bell pointed out, once the headlines died down, so did Reagan's attention to education.⁴⁸

Sizer related *A Nation at Risk* to the 1974 report *Youth: Transition to Adulthood*, produced by a panel headed by James Coleman. "*Risk* looked at schools, *Youth* looked at adolescents," Sizer wrote, and he noted that there is much more to an adolescent's life than school, which, even in 1974, was not "perceived by students as 'real.'" Sizer feels that the progress since 1983 has been disappointing, but "the reason is likely to be found less in the schools and to be largely due to the manner and settings (real and virtual) in which contemporary youths grow up, to the absence of influential adults regularly in their lives, and to the insistent and often engaging pressure of commerce-driven messages that surround them."

Gerald Holton, a Harvard physics professor and a driving force on the National Commission on Excellence in Education, which produced the report, wrote in the *Chronicle of Higher Education* that his handwritten first draft began with "America is at risk."⁴⁹ Given that the report paid no attention to President Reagan's education agenda, Holton thought he needed a shocker to get the President's attention. As they listened to Reagan's speech accepting the report, one commis-

sioner was overheard to say, "We have been had." Reagan's speech totally ignored the report. Holton echoes Sizer when he concludes that any new commission on education would need "the courage to point to what is clearly the most basic flaw in the structure on which the educational system is built . . . the children of America are the most disenfranchised members of society."

Jim Harvey, a member of the staff of the commission and generally credited with the final rhetoric of the famous opening section, thought that *A Nation at Risk* offered a golden opportunity. He believes that we blew it and have been blowing it ever since:

The bumbling began immediately as Reagan started us by hailing our call for prayer in the schools and the abolition of the Department of Education. We hadn't said a word about either.

It continued with corporate calls for improving the skills of American graduates, apparently to equip them for jobs that were being moved offshore as fast as management and shareholders could locate shipping.

The latest policy non-starter is the No Child Left Behind Act, a shock and awe campaign in its own right. . . . Tests have become the nation's latest weapons of mass destruction.⁵⁰

And so we end where we began. NCLB is a weapon of mass destruction used to launch a campaign of shock and awe against the schools and against the children. The year that was, was not a good one.

1. Robert L. Linn, "Accountability: Responsibility and Reasonable Expectations," presidential address to the annual meeting of the American Educational Research Association, Chicago, 23 April 2003.
2. George F. Will, "Shame: School Reform's Weak Weapon," *Washington Post*, 2 March 2003, p. B-7.
3. *Testing the Testers 2003: A Ranking of State Accountability Systems* (New York: Princeton Review, 2003).
4. *Meeting the Highly Qualified Teachers Challenge* (Washington, D.C.: U.S. Department of Education, June 2003), p. 11.
5. Margaret Raymond, Stephen H. Fletcher, and Javier Luque, *The Teacher for America Evaluation in Houston* (Palo Alto, Calif.: Hoover Institution, August 2001), pp. 30-31.
6. Dan D. Goldhaber and Dominic J. Brewer, "Why Should We Reward Degrees for Teachers?," *Phi Delta Kappan*, October 1998, pp. 134-38. For a refutation of their claims, see Gerald W. Bracey and Alex Molnar, "Recruiting, Preparing, and Retaining High-Quality Teachers: An Empirical Synthesis," Educational Policy Studies Laboratory, Arizona State University, Tempe, February 2003, available at www.asu.edu/educ/eps/EPRU/documents. Search on the title.
7. *Meeting the Highly Qualified Teachers Challenge*, p. 46.
8. Brian Rowan, *What Large-Scale Survey Research Tells Us About Teacher*

- Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools* (Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, November 2002), p. 12.
9. Jay P. Greene, "An Unfair Grade for Vouchers," *Wall Street Journal*, 16 May 2003; and Alan B. Krueger and Pei Zhu, "Another Look at the New York City School Voucher Experiment," paper presented at the Conference on Randomized Experimentation in the Social Sciences, Yale University, August 2002.
 10. Cecilia Rouse, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *Quarterly Journal of Economics*, May 1998, pp. 553-602.
 11. Cecilia Rouse, "School Reform in the 21st Century: A Look at the Effect of Class Size and School Vouchers on the Academic Achievement of Minority Students," Working Paper No. 440, Industrial Relations Section, Princeton University, available at www.irs.princeton.edu. Click on "working papers."
 12. *School Vouchers: Publicly Funded Programs in Cleveland and Milwaukee* (Washington, D.C.: General Accounting Office, Report No. GAO-01-914, August 2001), pp. 4-5.
 13. William G. Howell, Patrick J. Wolf, Paul E. Peterson, and David E. Campbell, "Test-Score Effects of School Vouchers in Dayton, Ohio, New York City, and Washington, D.C.: Evidence from Randomized Field Trials," unpublished paper, August 2000, available from the Program in Educational Policy and Governance, Harvard University.
 14. Kate Zernike, "New Doubt Cast on Study That Backs Voucher Efforts," *New York Times*, 15 September, 2000, p. A-26.
 15. *School Vouchers: Characteristics of Privately Funded Programs* (Washington, D.C.: General Accounting Office, Report No. GAO-02-752, September 2002), p. 3.
 16. Jay P. Greene, *The Effect of School Choice: An Evaluation of the Charlotte Children's Scholarship Fund* (New York: Manhattan Institute, Civic Report No. 12, August 2000), available at www.manhattan-institute.org/html/cr_12a.htm.
 17. Paul E. Peterson and William G. Howell, "Efficiency, Bias, and Classification Schemes: Estimating Private-School Impacts on Test Scores in the New York City Voucher Experiment," unpublished paper, June 2003, submitted to *The American Behavioral Scientist*. It is available at www.ksg.harvard.edu/pepg.
 18. Jay P. Greene and Paul E. Peterson, "Methodological Issues in Evaluation Research: The Milwaukee School Choice Plan," unpublished paper, August 1996, available from the Program in Educational Policy and Governance, Harvard University.
 19. Michael Winerip, "What Some Much-Noted Data Really Showed About Vouchers," *New York Times*, 7 May 2003, p. B-12.
 20. Peterson and Howell, p. 6.
 21. Kim K. Metcalf et al., *Evaluation of the Cleveland Scholarship and Tutoring Program, Summary Report 1998-2001* (Bloomington: Indiana Center for Evaluation, Indiana University, March 2003), p. 7.
 22. Doug Oplinger and Dennis J. Willard, "More Money for Vouchers," *Akron Beacon Journal*, 29 June 2003, p. A-1.
 23. Milton Friedman, "Public Schools: Make Them Private," *Washington Post*, 19 February 1995, p. C-7.
 24. *The Voucher Veneer: The Deeper Agenda to Privatize Public Education* (Washington, D.C.: People for the American Way, July 2003).
 25. Michael Kinsley, "A Wife's Tale," *Washington Post*, 18 March 2002, p. A-17.
 26. See, for example, Amy Stuart Wells, *Beyond the Rhetoric of Charter School Reform: A Study of Ten California School Districts* (Los Angeles: UCLA, 1998); Jerry Horn and Gary Miron, *An Evaluation of the Michigan Charter School Initiative: Performance Accountability and Impact* (Kalamazoo: Evaluation Center, Western Michigan University, 2000); idem, *Evaluation of the Michigan Charter School Initiative: Final Report* (Kalamazoo: Evaluation Center, Western Michigan University, 1999); and Casey Cobb and Gene V Glass, "Ethnic Segregation in Arizona Charter Schools," *Education Policy Analysis Archives*, vol. 7, 1999, available at <http://epaa.asu.edu/epaa/v7n1>.
 27. David Arsen, David Plank, and Gary Sykes, *School Choice Policy in Michigan: The Rules Matter* (East Lansing: Michigan State University, 1999).
 28. Ron Zimmer et al., *Charter School Operations and Performance: Evidence from California* (Santa Monica, Calif.: RAND Corporation, 2003).
 29. See Wells, op. cit.; Katrina Bulkley, "Educational Performance and Charter School Authorizers: The Accountability Bind," *Education Policy Analysis Archives*, vol. 9, 2001, available at <http://epaa.asu.edu/epaa/v9n37.html>; and Gerald W. Bracey, *What You Should Know About the War Against America's Public Schools* (Boston: Allyn & Bacon, 2003), chap. 6.
 30. Margaret Raymond, *The Performance of California Charter Schools* (Palo Alto, Calif.: Hoover Institution, May 2003); and David Rogosa, *Student Progress in California Charter Schools, 1999-2002*, available at www.cde.ca.gov/psaa/apiresearch.htm. Look under the category "Other Reports of Interest."
 31. Raymond, p. 2.
 32. Jay P. Greene, Greg Forster, and Marcus A. Winters, "Apples to Apples: An Evaluation of Charter Schools Serving General Student Populations," Working Paper No. 1, Center for Civic Education, Manhattan Institute, July 2003.
 33. Ibid., p. 8.
 34. Anthony Ralston, "Next Disaster in American Education: Rising Test Scores," *Sacramento Bee*, 1 December 2002.
 35. Greg Winter, "Make-or-Break Exams Grow, but Big Study Doubts Value," *New York Times*, 28 December 2002, p. A-1.
 36. Robert Rothman, Jean B. Slattery, Jennifer Vranek, and Lauren B. Resnick, *Benchmarking and Alignment of Standards and Testing* (Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA, CSE Technical Report 566, May 2002), p. 29.
 37. D. Aileen Dodd, "Ruling Clears Teacher Opposed to Gateway Test," *Atlanta Journal-Constitution*, 18 December 2002.
 38. *The Nation's Report Card: Trial Urban District Assessment, Reading 2002* (Washington, D.C.: National Center for Education Statistics, Report No. NCES 2003-523, July 2003).
 39. Diana Jean Schemo, "Questions of Data Cloud Luster of Houston Schools," *New York Times*, 11 July 2003, p. A-1.
 40. "Houston's School Dropout Debate," *New York Times*, 21 July 2003, p. A-14.
 41. Rod Paige, "Where's the Choice?," *Wall Street Journal*, 29 July 2003.
 42. Chris Osher, "Paige Calls for Closing 'Achievement Gap,'" *Pittsburgh Tribune Review*, 29 July 2003.
 43. Sara Rimer, "Philadelphia School's Woes Defeat Veteran Principal," *New York Times*, 15 December 2002.
 44. Janice I. Solkov, "Privatizing Schools Just Shouldn't Be This Hard," *Washington Post*, 2 February 2003, p. B-4.
 45. Gerald W. Bracey, "April Foolishness: The 20th Anniversary of *A Nation At Risk*," *Phi Delta Kappan*, April 2003, pp. 616-21.
 46. Koret Task Force, *Our Schools and Our Future . . . Are We Still At Risk?* (Palo Alto, Calif.: Hoover Institution Press, 2003).
 47. Pete Du Pont, "Two Decades of Mediocrity," *Wall Street Journal*, 5 May 2003.
 48. Terrel H. Bell, *The Thirteenth Man: A Reagan Cabinet Memoir* (New York: Free Press, 1988).
 49. Gerald Holton, "An Insider's View of *A Nation at Risk* and Why It Still Matters," *Chronicle of Higher Education*, 25 April 2003.
 50. James Harvey, "Nation's Students Still at Risk," *Seattle Times*, 4 May 2003. ■