Hurrah for Houston

Congratulations to Education Next for its trio of articles on the positive reforms in the Houston Independent School District, carried out in large part by Secretary of Education Rod Paige while he was the district’s superintendent (see “Houston Takes Off,” Feature, Fall 2001).

The article by Jane Hannaway and Shannon McKay of the Urban Institute showed how mandatory testing can serve as a lever to create structural change in an otherwise entrenched education bureaucracy. The piece by Marc Kanstoroom highlighted the benefits not only of testing students, but also of assessing their teachers.

Houston’s reform program—systematically tying standardized tests to a back-to-basics curriculum and combining teacher training with recognition—does not involve punishing teachers whose students fare poorly. Instead, as Kanstoroom reported, the teachers receive the extra training they need to improve.

Larry Parker
Alexis de Tocqueville Institution
Arlington, Virginia

Vouchers versus class size

The statistics reported in Dan Goldhaber’s otherwise thoughtful article (see “Significant, but Not Decisive,” Research, Summer 2001) inadvertently tilted the comparison between vouchers and class-size reduction in favor of vouchers. When comparable samples and measuring sticks are used, the improvement in test scores for black students from attending a small class based on the Tennessee STAR experiment is about 50 percent larger than the gain from switching to a private school based on the voucher experiments in New York City, Washington, D.C., and Dayton, Ohio.

In the D.C. voucher experiment, African-American students in grades 2 through 5 reportedly increased their scores by an average of 10 national percentile points in mathematics and 8.6 points in reading after two years of private schooling. These percentiles are based on the national distribution of scores on the Iowa Test of Basic Skills. Goldhaber scaled these gains by dividing them by the standard deviation of percentile ranks for African-American students in the control group and concluded, “This represents a gain of about 0.5 standard deviation relative to African-American students whose applications for vouchers were unsuccessful.”

He compared this figure with Jeremy Finn and Charles Achilles’s finding that attending a smaller class in the Tennessee STAR experiment raised reading scores for black 2nd graders by one-third of a standard deviation. Finn and Achilles, however, measured test scores using a different metric—namely, “scale scores” on the Stanford Achievement Test. Because the scale scores follow a bell-shaped distribution, while percentile ranks are uniformly distributed between 0 and 100, a one-standard-deviation increase in scores does not imply the same improvement in achievement in the two measures. Furthermore, these effect sizes are not comparable because the standard deviation used to scale the voucher results is from a much less diverse sample: low-income, inner-city students who participated in the experiment.

A nother problem is that the effect sizes Goldhaber took from the Washington, D.C., voucher experiment were adjusted to account for imperfect compliance—the fact that not everyone offered a voucher attended private school, and some of those who weren’t offered a voucher nevertheless attended private school. If the same approach is applied to the STAR sample to adjust for the fact that some students did not enroll in the class they were assigned to—and a comparable sample of low-income black students is used—the gains in test scores after two years of attending a small class (average of 16 students) as opposed to a regular-size class (average of 23 students) is 9.1 national percentile ranks in reading and 9.8 ranks in math.

Note also that Goldhaber compared the STAR results with voucher results for just Washington, D.C., where the gains were higher than in Dayton or New York. Across all three cities, the average effect of switching from a public to a private school for black students was 6.3 percentile ranks in both math and reading. This would seem a more appropriate comparison.

Then there is the question of how much of the students’ gains in private schools are attributable to the fact that private schools had smaller class sizes in all three cities. In Washington, for example, the average class size attended by students who switched to private school was 18, compared with 22 for those who
remained in public school. The gain attributed to private schooling may be due to smaller classes.

Similar issues arise in William H. Howell et al.’s article reporting results from the voucher studies (see “Vouchers in New York, Dayton, and D.C.,” Research, Summer 2001). They scale the gain in black students’ scores by the standard deviation of test scores computed for a select sample of students, and observe that the gain in their scores due to attending private school is “roughly one-third of the test-score gap between blacks and whites nationwide.” The nationwide gap, however, is presumably scaled by the larger nationwide standard deviation. The standard deviation H. Howell et al. used to scale gains was around 19, while the standard deviation of national percentile scores is necessarily 28.9, because percentile ranks follow a uniform distribution. Using the national standard deviation to scale all scores, the effect of attending a private school on black students is only one-fifth to one-quarter as large as the black-white gap.

Lastly, making vouchers available does not assure that students switch to private schools. The estimated gain from being offered a voucher is only half as large as the gain from switching to private school (in response to being offered a voucher), so the estimated impact of offering vouchers is no more than one-eighth as large as the black-white test score gap.

Alan B. Krueger
Princeton University
Princeton, New Jersey

William Howell, Patrick Wolf, Paul Peterson, and David Campbell reply:
Alan Krueger does not question the quality of our research or the integrity of our findings. And for good reason. In his reevaluation of the Tennessee STAR study, he relies on the exact same research design and uses many of the same statistical procedures that we did in our studies of school vouchers in New York City, Dayton, Ohio, and Washington, D.C.

As Krueger correctly notes, the best estimates from our research come from taking the weighted average of the individual effects in the three cities. The result is that African-American students who switched from public to private schools scored, on average, 6.3 points higher than their public school peers; by contrast, Krueger reports effects of between 9.1 and 9.8 points for African-Americans placed in smaller classes.

Krueger wonders whether the 6.3 point impact is due simply to the smaller classes in the voucher schools. We were interested in this question as well; and, as we report elsewhere, the data do not support his conjecture. The size and significance of voucher effects for African-Americans appear unchanged after controlling for the class sizes in the public and private schools students attended.

The cloud hovering over the comparison between voucher and class-size research is not the intervening effect of class size on our findings, but rather the possible bias in Krueger’s own estimates. The problem lies in the data collection procedures used in the Tennessee STAR study. School administrators, rather than trained social scientists, assigned students to large and small classes, and no one collected baseline test-score data. As a result, there is no way of telling whether reported gains (most of which accrued in the program’s first year) are based on a true random assignment protocol or whether they reflect improper assignments of subjects to treatment and control conditions. (In our research, we ourselves assigned students randomly to test and control groups, and we collected baseline test scores to verify the lotteries’ success.)

We agree with Krueger that expressing impacts in terms of effect sizes critically depends on the population of students one considers. We followed the standard practice of dividing the estimated impact for the sample population by its standard deviation. Obviously, scaling this impact by different standard deviations from alternative populations of students—as Krueger does—would yield different effect sizes.

Krueger further notes that “making vouchers available does not assure that students switch to private schools.” This, of course, is true. But in a large-scale program, we do not know what proportion of the treatment group would actually use a voucher offered to them. For this reason, we focus on the experiences of those students who used the treatment as prescribed. And depending upon which scale one uses, vouchers appear to reduce the black-white test score gap by either one-quarter or one-third.

Too late
David Elkind’s arguments against academic training for young children make a great deal of sense, but...
I know so many individuals, myself then put them together. Years can begin to read larger words and at ages “four or five” and in the following mathematics and reading. He explains that based discussion of skills like mathematics, needs, and interests. “Unfortunately, respond “to the child’s developing ability, truth, noting that curriculum must correspond to the child’s developing abilities, needs, and interests.” Unfortunately, he continually lapses back into an age-based discussion of skills like mathematics and reading. He explains that children will learn to read phonetically at ages “four or five” and in the following years can begin to read larger words and then put them together.

I am compelled to disagree because I know so many individuals, myself included, who were reading at age four what Elkind would keep from us until the 2nd grade. First grade was rather meaningless for us. All this because America is tied to an outmoded system of age-based education that demands conformity rather than celebrating the unique gifts of the individual.

Trent England
Arlington, Virginia

David Elkind seems to think that literacy training has to replace hands-on learning. On the contrary, any good preschool program has literacy activities that appeal to the child’s senses. No one was a bigger advocate of early literacy than Maria Montessori. She started the “language readiness” activities with hands-on activities. Then she brought in objects to match with pictures, moving toward more abstraction slowly. Then the pictures would be matched with words, and so forth. A child’s first direct experience is with sandpaper letters. The child runs her fingers over the letters. After that there is the movable alphabet. Children make words with hands-on exercises.

Elkind warns of the “dangers” of introducing children “to the world of symbols too early in life.” But children are introduced to symbols as soon as they are given a name for an object or person. “Mommy” comforts me when I call out; a “bottle” soothes my hunger; a “teddy bear” sleeps with me at night. The concept that a word can stand for a thing is not new to a preschooler, or a toddler, or a baby for that matter.

The next level is understanding that spoken symbols can be represented by writing. The basic difference between the ability to speak a word and reading or writing it is a child’s level of phonological awareness, or his ability to understand the relationship between spoken sounds and their written symbols. Most children are ready to deal with phonological awareness in their preschool years. It is labeled “developmentally inappropiate” to teach early phonological awareness to poor children, but well-to-do families practice these skills at home all the time. It doesn’t seem to hurt their children.

Abida Ripley
Alexandria, Virginia

Define “evidence”

As a long-time teacher educator specializing in reading instruction, I was intrigued by your comments regarding “evidence” and education reform (see “Evidence Matters,” Letter from the Editors, Spring 2001). I find that certain terms used in the article, such as “evidence-based,” “scholarly integrity,” “the facts,” and “worthy research,” no longer have single, generally accepted meanings in my field. For example, investigations into how students best learn to read are carried out in two distinct ways.

One is the traditional experimental or empirical approach, which uses scientific methodology. The other is qualitative research. The latter gathers non-numerical, anecdotal data. It thus eschews statistical analysis in favor of subjective judgments. It is deliberately designed not to be replicable.

In studies of how children learn to read, conclusions from these two types of research consistently refute one another. The “whole language” approach to teaching reading cites qualitative evidence as proof of its effectiveness. However, none of the unique principles or novel practices of whole language reading instruction is corroborated by experimental investigations.

I hope that future editions of Education Next make it clear to readers which kind of research “evidence” is under discussion. At least in the context of research on reading, “evidence-based” is simply too imprecise an expression these days.

Patrick Groff
San Diego State University
San Diego, California

Feedback from a Young Reader

Cameron Paranzino
Rockville, Maryland