CENTER FOR DEMOGRAPHY AND ECOLOGY

The University of Wisconsin-Madison
1180 Observatory Drive
Madison, Wisconsin 53706

Robert M. Hauser
Vilas Research Professor of Sociology
(608) 262-2182, 262-8400 (FAX)
HAUSER@SSC.WISC.EDU

September 7, 2000

The Failure of Test-Based Retention in Chicago

I have read the December 1999 and September 2000 reports of the Chicago study carried out by the Consortium on Chicago School Research (CCSR), in light of the debate over ending "social promotion." The latest, sad findings reconfirm and strengthen those of the first report.

- The combination of test-based retention with a narrowly test-based curriculum has no remedial educational value.
- Over a three-year period, students who have been retained gain no more on the ITBS than students who have passed promotional gates or who have failed to meet the cut-score on the summer examination and been waived into the next grade. But those retained students lose a year of their lives, their retention adds to the cost of schooling, and they have a higher risk of dropping out before completing high school.
- Low-scoring eighth graders are somewhat more likely to drop out within two years under the new policy than under the policies of 1995, and dropout is especially high among low-scoring eighth graders who were retained or sent to transition centers in 1997.
- There are higher rates of passing the promotional gates at grades 3, 6, and 8, but they are no reason to celebrate. Gains have been very small between 1998 and 1999. Part of the gain in grade 3 is explained by more retention in grades 1 and 2. Part of the gain in all grades is due to increased exemption from the testing program. Morever, fewer waivers have been issued after 1997, so retention rates remain very high.
- Most important, repeated use of the same forms of the same tests invariably leads to test-score inflation, and that may well account for the observed increase in pass rates.

The meager impact of the Chicago reforms is especially evident in comparisons between third-graders who failed to meet the cut-score on the ITBS in the spring of 1997 and third-graders in the spring of 1995 whose scores were so low that they would have failed to meet the cut-score. (That is, I ignore the differences among students who passed after the 1997 summer bridge and who were retained or waived thereafter.) From Figure 2-9 of the first Consortium study, I calculated that the small initial difference between the two low-scoring groups (0.1 ITBS Grade Equivalents in the December 1999 report and 0.0 among continuing students in the September 2000 report) was unchanged a year later – after the 1997 students had gone through the first round of testing, the Summer Bridge program, and the following academic year. This finding is reinforced by the additional year of data: As shown by Figure 14 of the September 2000 report, the average three-year gain in GE is 2.7 among students who would have failed in 1995, and it is 2.7 among all students who failed to

meet the cut-score on the ITBS in the spring of 1997 (in the categories "passed after the bridge," "waived," and "all retained"). That is, the overall test-score gain of initially low-scoring 3rd graders was essentially unaffected by selection after the 1997 summer bridge into groups that were passed, waived, or retained. The program had no overall effect on test-score gains of low-scoring third-graders after three years, but we can expect that it will eventually increase school dropout among the retained students.

Much the same finding holds among students at the sixth grade level. There is virtually no difference after three years between the test-score gains of low-scoring sixth graders in 1995 and the aggregate of those who initially failed to meet the cut-score on the ITBS in the spring of 1997 and were subsequently successful or unsuccessful in the summer bridge program. Again, the program had no effect on test-score gains of low-scoring sixth graders after three years, but we may expect it to increase school dropout.

Both among all third-graders and among all sixth-graders, there was a small overall increase in test-score gains between those who first completed the grade in 1995 and those who first completed the grade in 1997. But the gain is due entirely to the larger share of students who initially passed the ITBS. Yet we know that the increased passing rates are of doubtful validity because of the repeated use of the same forms of the same test after 1995.

In its 1999 report on high stakes testing, the National Research Council (NRC) stated a few basic principles of appropriate test use: (1) Any particular test has validity only in relation to specific uses; (2) tests are not perfect, but neither are the alternatives to tests; (3) no high-stakes educational decision about a test-taker should be made solely or automatically on the basis of a single test score; other relevant information should also be taken into account; (4) tests should be used for high-stakes decisions only after students have been taught the knowledge and skills on which they will be tested; (5) neither test scores nor any other kind of information can justify educational decisions that are not beneficial for students.

At the outset, the Chicago plan violated these basic principles. The Iowa Test of Basic Skills (ITBS) was not designed to be used as the sole criterion of grade-promotion decisions. Test reliability was not considered in setting arbitrary cut scores, and there was no systematic second chance to take the spring exam. No other clearly specified criteria were used at first – though we now have learned from the Consortium study that thousands of students have actually been promoted, despite failing scores. The ITBS has no verified relationship to previously established curricular goals and standards of the Chicago Public Schools. In fact, the NRC Committee concluded that Chicago's regular year and summer school curricula were so closely geared to the ITBS that it was impossible to distinguish real subject mastery from mastery of skills and knowledge useful for passing this particular test. Again, this problem is exacerbated by Chicago's repeated use of exactly the same forms of the test.

This latest evaluation of the Chicago plan reconfirms that it has also failed the last of the NRC's

criteria; there is no credible evidence of lasting educational benefit, either to those who passed or those who failed to exceed cut-scores on the ITBS. Use of an independent, external standard of academic achievement – not the ITBS – is essential to a valid evaluation – but the Consortium study includes no independent standard of achievement.

The failure of the Chicago plan will come as no surprise to anyone who has followed research on the consequences of grade retention. While retention sometimes leads to short-term gains for retained students – and even these are lacking in Chicago – study after study shows that the costs of retention are high, both to school systems and to students, and that retention leads to lower academic achievement and greater high school dropout. And a worrisome level of school dropout is already signaled by the Chicago data. These findings reconfirm the failure of retention as shown in New York's Promotional Gates program of the 1980s and by more recent studies in Baltimore and in Texas.

The Consortium reports merely tell us, again, what has happened in the past, not what need happen in the future:

- Chicago's school leaders should ask themselves, what are the likely consequences, both immediate and in the long-term, of their well-intentioned efforts to raise educational standards?
- Chicago's leaders need strong evidence that reforms will work before they put them in place on a large scale.
- Chicago's leaders should maintain a commitment to measure reforms and their consequences as they take place. The Consortium studies are a commendable move in this direction, but their value is reduced by the absence of an independent criterion of achievement and by other problems of research design and analysis.
- Chicago's leaders should remember what the problem is as they seek to solve it. The problem is not social promotion; it is low academic achievement.

During 1998, I served as Chair of the Committee on Appropriate Test Use of the Board on Testing and Assessment at the National Research Council. The National Research Council is the operating arm of the National Academy of Sciences, which was chartered by Congress in 1863 to advise the government on matters of science and technology. I was elected to the Academy in 1984. The Committee on Appropriate Test Use prepared its report, *High Stakes: Testing for Tracking, Promotion, and Graduation* (National Research Council, eds. Jay M. Heubert and Robert M. Hauser. Washington, DC: National Academy Press, 1999), in response to a Congressional charge "to recommend appropriate methods, practices, and safeguards to assure that existing and new tests … are not used in a discriminatory manner or inappropriately for student promotion, tracking, or graduation, and existing and new tests adequately assess student reading and mathematics comprehension in the form most likely to yield accurate information regarding student achievement of reading and mathematics skills" (P.L. 105-78, Sec. 309). The NRC panel was a diverse group of 15 scholars from all across the country. My comments on the Consortium study are informed by the NRC study, but

they are mine alone. Neither the NRC nor my fellow committee members bear any responsibility for my opinions.

When used appropriately, high-stakes tests can help promote student learning and equal opportunity in the classroom by defining standards of student achievement and by helping school officials identify areas in which students need additional or different instruction. When used inappropriately, high-stakes tests can undermine the quality of education and reduce opportunities for some students, especially if results are misinterpreted or misused, or students are relegated to a low-quality educational experience as a result of their scores.

One of the NRC's strongest recommendations was that "Accountability for educational outcomes should be a shared responsibility of states, school districts, public officials, educators, parents, and students. High standards cannot be established and maintained merely by imposing them on students" (p. 5). Chicago has tried to raise achievement largely by imposing arbitrary standards on students. It is time for Chicago's educational leaders to become accountable, to recognize the failure of their "get-tough" policies, and to turn in new directions.