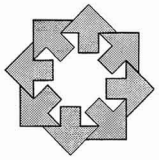
Purpose of this summary: To outline trends in CPS systemwide student achievement over the past decade, based on the findings of a Consortium study of standardized testing. After placing test scores on a more reliable, content-based scale, researchers found that learning gains had improved in both reading and mathematics.



A REPORT SPONSORED BY THE CONSORTIUM ON CHICAGO SCHOOL RESEARCH

Summary

Examining Productivity: Ten-Year Trends in the Chicago Public Schools

March 1998

Introduction

In the report *Academic Productivity of Chicago Public Elementary Schools,* the Consortium on Chicago School Research finds that the testing system used by the Chicago Public Schools (CPS) is limited in its ability to assess the productivity of individual schools. Because different forms of the Iowa Tests of Basic Skills (ITBS) are used from year to year, significant problems arise in drawing inferences about changing school productivity. In addition, the statistical indicator used to measure school productivity—percentage of students scoring at or above national norms on the ITBS—is insensitive to gains of significant numbers of students whose scores are well below national norms but who may have made great improvements.

For these reasons, the Consortium created a constant measurement scale that takes into account the differences among the ITBS forms used in the Chicago schools over the last decade. Using the new scale, the report authors were able to reexamine test scores of CPS students since 1987. The results, which are summarized below, show that year-to-year gains in student learning have improved significantly in the decade since the passage of the 1988 Chicago School Reform Act.

Problems with Nationally Norm-Referenced Tests

The ITBS is the main achievement data gathered annually by the Chicago Public Schools and is the sole information source currently used by the system for school accountability purposes. These tests are inexpensive and relatively easy to administer and score. They are quite useful for the purposes for which they were originally intended—to provide information about how well Chicago's students perform against a national sample of students who

Drawn from the Technical Report **Academic Productivity of Chicago Public Elementary Schools**

Anthony S. Bryk
University of Chicago

Yeow Meng Thum
University of California at Los Angeles

John Q. Easton
Consortium on Chicago School Research

Stuart Luppescu
Consortium on Chicago School Research

took the same test. They were not, however, specifically designed for the purposes for which they now are used—to assess improvements in schools' productivity over time.

By way of background, the ITBS is not a single test, but rather a testing system. It consists of multiple forms that were developed at different points in time. These forms are literally different tests with no overlapping items. Each form consists of multiple levels, each designed to be administered to students at a particular grade. For example, level 9 is designed for grade 3, level 10 for grade 4, and so on. Although it is now an infrequent practice in the CPS, students sometimes have been tested "off level," such as giving level 8 to a very disadvantaged third grader or level 10 to a gifted student at the same grade.

The Non-Equivalence of Grade Equivalence

The ITBS, like most nationally norm-referenced standardized tests, produces a score report called a grade equivalent (GE). GEs have a great deal of appeal to teachers and parents because they appear to describe a child's performance in developmental terms of grade level and months within grades. For example, a student tested in the eighth month of the fourth grade would be "at grade level" or "at the national norm" if his/her score were 4.8. However, because of the different forms and levels of the tests, the scores actually are not comparable from one year to the next and are not designed to trace change over time.

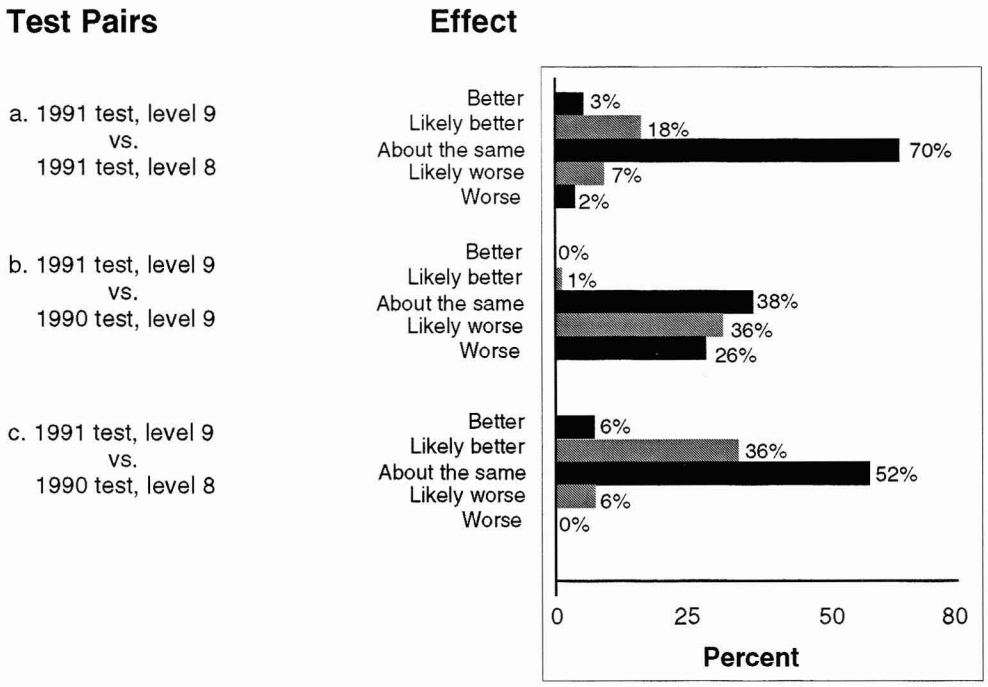
To demonstrate the problems here, we gave a sample of CPS students two different reading and math tests from the ITBS series. For one group, we gave two different levels of the same test form. The second group received the same level of two different forms of the test. Finally, for the third group we switched both the form and the level. As Figure 1 shows, the two tests do not always tell us the same thing about a child's grade level. For example, students who were given the form used by the CPS in 1991 (CPS91), were more than twice as likely to have better GE scores on the higher level test (level 9) than on the lower level test (level 8). (See Figure 1a.) Similarly, consider the students who took the same level of the test from two different years (Figure 1b). These students were much more likely to do better on CPS90 than on CPS91. These differential score effects are equally dramatic when we consider the comparison across forms and levels (Figure 1c). Students were seven times more likely to score higher on CPS91, level 9 than on CPS90, level 8. These empirical examples illustrate a general problem that grade equivalents are both form and level specific and cannot be strictly compared. Clearly, this limits our ability to make accurate statements about how much actual learning an individual student is making over time.

A Value-Added Approach

The report argues that the average achievement level of students in a school is the most useful single statistic for informing us about the overall level of students' capabilities. Unlike the current CPS statistic, percent of students at or above national norms—which is sensitive to the performance of only the students who are near this threshold—the overall mean takes into account the performance of all students. However, average achievement is not an especially good indicator of school productivity and whether this is

The authors of the report would like to thank Rebecca Williams for her help in creating this summary.

Figure 1. GE Test Score Bias Due to Form and Level Differences



Note: *About the same* category is +/- 1 standard deviation from zero.

changing over time. One major problem that this indicator fails to take into account is student mobility. For example, if a group of students enrolls in a school sometime during the academic year, even on the day just before testing, their scores will be counted as part of the overall achievement level for the school. Clearly, the attainment for these students depends primarily on their previous schooling experiences and home background and tells virtually nothing about the effectiveness of the particular school. This concern is especially problematic in urban school districts such as Chicago because student mobility tends to be high. In the typical Chicago elementary school only 80 percent of the students tested in a given year were also tested in the same school the previous year. This means that 20 percent of the students are new each year.

To develop an appropriate indicator for purposes of assessing school productivity and whether this is changing over time, factors such as student mobility and changing school demographics must be taken into account. This process begins with a basic accountability principle: **A school should be held responsible for the learning that occurs among students actually taught in that school**. This suggests that rather than focusing exclusively on the average achievement level at each grade, a good indicator must also consider the *gains in achievement* made by students at each grade in the school.

With these ideas as background, the report proceeds to define a *productivity profile* for each grade in a school. The productivity profile is built from two basic pieces of information for each school grade: the *input status* for the grade and the *learning gain* recorded for that grade.

Across the board, for all elementary grades three through eight, the 1996 learning gains were substantially greater than in 1994 for both reading and mathematics.

To find the *input status*, the authors began by identifying the group of students who received a full academic year of instruction in each grade in a particular school, and then retrieved their ITBS test scores from the previous spring. The average of these students' previous year's test scores is the input status for that school grade. As for the *learning gain* for each school grade, this is simply how much the end-of-year ITBS results have improved over the input status for this same group of students. The grade productivity profiles thus offer more complete information about academic productivity in that they take into account not only whether a school's test scores are rising, but also how much the students in that school have improved over the year.

Next, the study uses grade productivity profiles to estimate a school's learning gain trend in the form of a *learning gain index* (LGI). The LGI is a measure of the changes in learning gains over the five-year period from 1992 to 1996. Specifically, it assesses the relative change in student learning from the last five years as compared to the amount of learning that occurred across the system in the study's base year, 1991.

Avoiding the Test-Form Problem: 1994 vs. 1996 Gains

Fortuitously, there is one set of comparisons embedded in the post-reform testing series that is not afflicted by the problems of non-equivalent test forms. The test form administered in 1993, Form K, was repeated in 1995, and the form used in 1994, Form L, was repeated in 1996. As a result, the 1994 and 1996 learning gains are directly comparable because they are based on the same pairs of test forms and levels. (For example, the third grade gain in both years is based on second grade students taking level 8 of Form K, followed the next year by level 9 of Form L.) For this reason, the analysis of systemwide trends begins by focusing on the strongest piece of evidence where results are not contingent on the completeness of the Consortium's test equating study (discussed below).

Across the board, for all elementary grades three through eight, the 1996 learning gains were substantially greater than in 1994 for both reading and mathematics. (See Figures 2a and 2b. The vertical axis of the figures represents how far students have advanced along the 100-point scale developed by the Consortium to measure test content difficulty. The same pattern occurs in GE scores.) In relative terms, student gains in 1996 rep-

resent improvements ranging from 10 to nearly 40 percent over the 1994 levels. (See Figure 2c.) This is an impressive two-year productivity gain by most any standard.

These results immediately raise a second question: "Is this a one-year phenomenon (e.g., something attributable to the administrative reforms of 1995) or rather a signal of a longer-term improvement trend that links back to the reforms of 1988?" To investigate this question, the authors used the results of the Consortium's equating study in order to examine the 1996 gains within the larger context of the learning gain trends over the last several years. The equating study produced a constant, content-based scale that helps adjust for the differences in difficulty of ITBS test forms used over these years. If the improvement registered in 1996 is a one-year phenomenon, one would expect to find flat or possibly even declining trends in learning gains through 1995 followed by one big upward jump in 1996. On the other hand, the 1996 data may look like part of a longer-term trend.

The study found that taken overall, the 1996 results appear for the most part to be embedded in longer-term trends. There is some evidence, however, of an upward jump in 1996 in selected grades, especially in mathematics.

Systemwide Average Productivity Profiles

Figure 3 displays the overall ITBS mathematics productivity profile for the CPS. The display aggregates the grade productivity profiles from all individual elementary schools in the system. Notice that the output trends in mathematics are up for

Figure 2a. 1994 vs. 1996 ITBS Reading Gains

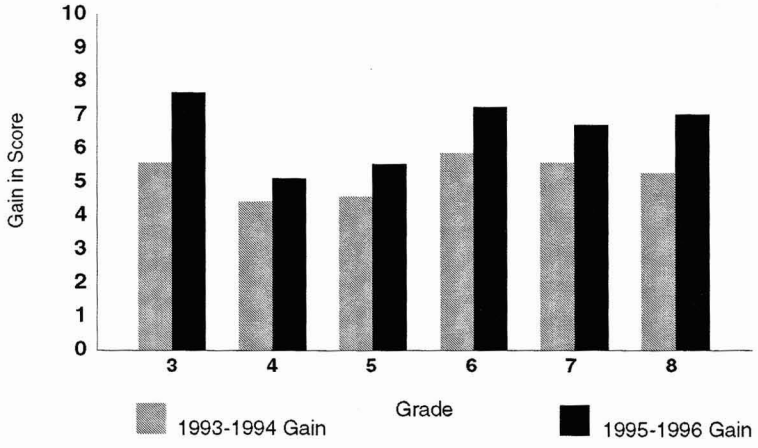


Figure 2b. 1994 vs. 1996 ITBS Mathematics Gains

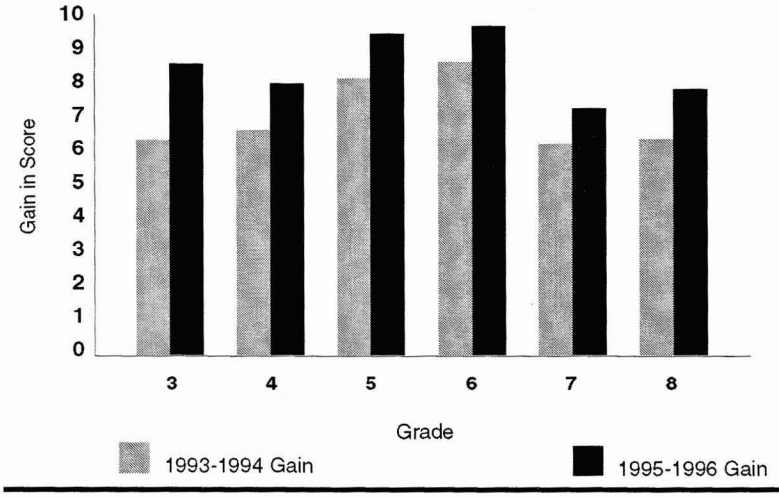
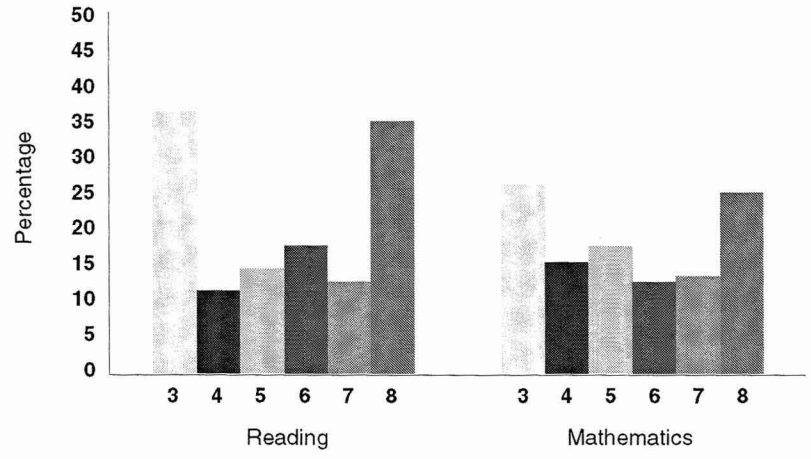


Figure 2c. 1996 Improvements over 1994 Gains



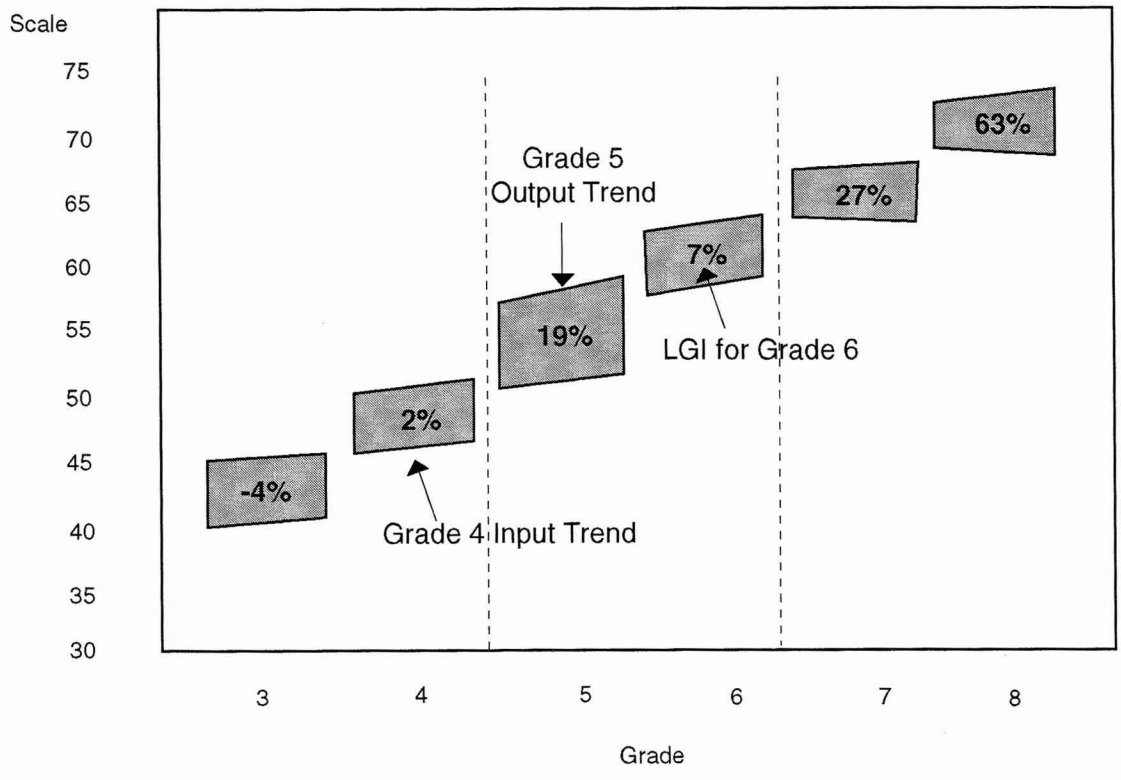
Note: (1996-1994 gain) / average (1988, 1989, 1990 gain)

all grades three through eight. The learning gain trends also show marked improvements for the intermediate and upper grades. For grade five, the systemwide improvement was 19 percent over the five-year period from 1992 through 1996. For grades six, seven, and eight the relative improvement in mathematics learning was 7 percent, 27 percent, and 63 percent respectively. In contrast, grades three and four show little change in learning gains over this period.

The grade three mathematics data are quite interesting. The estimated LGI for grade three is actually slightly down (by 4 percent), but the output trend is still positive. This is a case where if only the output trend was considered, as it would be under a more traditional accountability approach, it could be mistakenly concluded that third grades have been improving systemwide. In fact, the registered gains in achievement at the end of grade three appear largely attributable to changes occurring prior to grade three.

Figure 4 displays the systemwide productivity profiles for reading. The results here are a bit more mixed, but still generally positive. The output trends are up at all grades except grade four; the rates of improvement, however, are not as large as in mathematics. Grade four registers a learning rate decline of 22 percent. Grades three, five, six and seven, however, show significant gain trends ranging from 10 to 42 percent; Grade 8 remains basically unchanged.

Figure 3. Mathematics Productivity Profile for CPS, 1987-1996



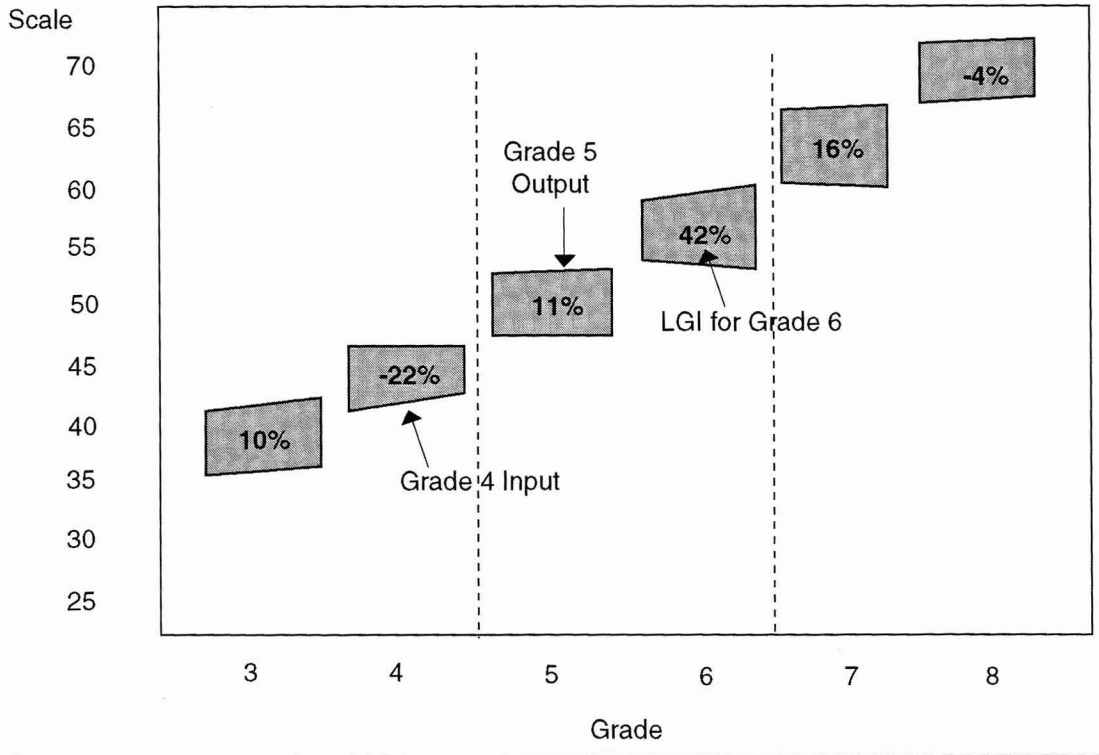
LGI = Learning Gain Index, computed for 1992-1996.

Overall, these analyses indicate broad-based systemwide improvements in student learning, stronger in mathematics but also in reading. Moreover, the authors believe that these data, up through 1996, are a reasonably good indicator of meaningful changes in instruction and student learning because no high stakes external accountability was associated with them. That is, prior to 1996 the main external accountability force over Chicago Public Schools was the Illinois State Board of Education, which based its school rankings and "academic watch list" on IGAP data. Although ITBS scores still mattered to individual schools, no formal consequences were directly attached to them. Beginning in 1996, the CPS instituted its own high stakes accountability system based exclusively on the ITBS; as a result, the future utility of these data as an indicator of broad instructional improvement has become more questionable.

Interpretation

Interpreting these results is complicated by the fact that the ITBS has been a tacit rather than an explicit standard for school performance. While the content of the tests is certainly reasonable, it has never been publicly established as a systemwide standard for subject matter content and sequence. Individual schools may well be working on other academic goals. They might, for example, focus more on higher-order thinking skills or deeper engagement of students with projects and some specific subject mat-

Figure 4. Reading Productivity Profile for CPS, 1987-1996



Note: LGI = Learning Gain Index, computed for 1992-1996.

ter. Some schools may also be teaching the ITBS content, but not in precisely the same sequence. Such schools may record weak test scores in some grades, where the alignment between the test and instruction is poor, and then much better results in other grades where the test better matches with students' actual classroom experiences.

With these caveats in mind, the authors proceed to offer an interpretation. There appears to have been significant improvement systemwide in primary programs through grade three. This is reflected in the improving second grade output trends in half of the schools in mathematics and a third in reading. It also is known generally that the student population entering the Chicago Public Schools has been growing gradually more disadvantaged over the recent 10-year period.[1] Thus, although the report cannot specifically account for the sources of the improvement (e.g., more state pre-kindergarten, improvement in kindergarten, grades one and two curriculum), something positive appears to have happened here.

For the most part, the early gains prior to grade three, however, are not being further advanced during grades three through four in either mathematics or reading. In fact, some of the early improvements appear to be lost. In contrast, the intermediate and upper elementary grades look quite positive. A word of caution is in order, however. If the primary grades suddenly began to promote children with much higher levels of skills, this would challenge intermediate and upper grade teachers to rethink their instruction to build on the advanced knowledge that students now possess. In the absence of such proactive improvement efforts by intermediate grade teachers, i.e., if they simply repeated past practice, their productivity trends would suddenly start to look worse. In short, the positive trends in the intermediate and upper grades may, in part, be a curious consequence of the lack of productivity improvement in primary grades.

Finally, these analyses strongly suggest that Chicago school reform has precipitated substantial improvements in achievement in a very large number of Chicago public elementary schools. The governance reforms of 1988 and 1995 have significantly advanced the learning opportunities afforded to literally hundreds of thousands of Chicago's children. While more improvements are still needed, these results should nonetheless encourage the public that Chicago schools can substantially improve and that this is, in fact, occurring.

Endnote

[1]See Storey, Sandra, John Q. Easton, Thomas C. Sharp, Heather Steans, Brian Ames, and Alicia Bassuk (1995). *Chicago's Public School Children and Their Environment.* Chicago: Chicago Public Schools and Chicago Urban League and the Latino Institute.

See the Consortium's world wide web site for productivity profiles of each Chicago Public Elementary School. These data are considered public information.
http://www.consortium-chicago.org

Consortium on Chicago School Research
1313 East 60th Street • Chicago, Illinois • 60637
773-702-3364 • 773-702-2010-fax