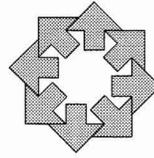


Purpose of this policy brief: To summarize the recommendations of the Consortium's Steering Committee regarding changes in standardized testing in the Chicago Public Schools. The current testing system, the ITBS, does not accurately assess year-to-year changes in achievement, and needs to be replaced with a system that does.



A REPORT SPONSORED BY
THE CONSORTIUM ON
CHICAGO SCHOOL RESEARCH

Policy Brief

Examining Productivity: Improving the Assessment System of the Chicago Public Schools

March 1998

The Chicago Public Schools (CPS) need accurate, fair, and consistent assessments that will give the public credible information for charting academic improvement over time. The current basis for knowing these things—scores from the Iowa Tests of Basic Skills (ITBS) in reading and mathematics—fails us in several key regards. It should be replaced with new assessments that directly align with the new CPS standards.

This conclusion comes after years of study of CPS's current testing and reporting system by the Consortium on Chicago School Research. The Consortium, which has conducted a number of analyses of Chicago's school reforms since 1991, chose now to focus intensely on assessment because it plays such a critical

role in reform. Test scores identify failing schools for probation or reconstitution. They provide a picture of the progress of systemwide improvement. Teachers and parents use these data to judge whether their students are making adequate progress.

The Consortium's research revealed such serious problems with the current assessment system that the authors of the study recommend replacing it with a system more reliable and suited to the current needs of CPS. The Consortium's Steering Committee concurs and endorses a set of recommendations that frame this policy brief. Following the full text of the recommendations is supporting evidence drawn from the Consortium's technical report, *Academic Productivity of Chicago Public Elementary Schools*.

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*

Recommendations of the Steering Committee

Alignment with CPS Learning Goals

New CPS standardized tests in reading and mathematics should be directly aligned with the recently approved CPS Learning Standards. Only if the assessments are specifically developed to achieve this aim and have been demonstrated to be valid in this regard will teachers, students, and parents know whether they are making genuine progress on these important goals. The content of the standards should dictate the content of the tests. A “back-in” solution (i.e., choosing among existing tests the one that comes closest to matching the standards) is inadequate. Under such an approach the test publishers rather than local leaders get to decide the accountability standards for judging schools.

Score Reporting on a Content-Referenced Scale

In addition to a reference to national norms, test score results should be reported in terms of the specific knowledge and skills that students have demonstrated. The content-referenced scales used in this paper illustrate this kind of approach. The results for a sample of actual test questions that were administered to students each year should also be made public. This can help to promote in each school community, and across the city, an educational, as well as a numeric conversation about what students actually know and can do.

A Stable Measurement Ruler for Assessing Academic Progress

For purposes of making judgments about school and system improvement over time, it is essential to maintain a stable testing system that consistently measures students against the same content standards from year to year. This will require that rigorous equating procedures be fully integrated into the overall testing system design. As views about appropriate learning goals shift over time, this will entail developing new assessments and starting new trend lines about progress toward these new standards. The changes in goals should not be obscured by changing forms of the assessment, as has been the case in the past.

An Accountability Focus on Schools’ Value Added to Student Learning

Each Chicago school should be held accountable for the amount of learning acquired by students enrolled in that school. While the current CPS school accountability indicator—the percentage of students at national norms—provides information about the overall attainment of students, it does not tell us how much they have learned in any particular school. Whether gains in student learning are improving over time should play a major role in the school accountability system. Thus, for purposes of reporting on individual school accountability, the CPS should add a focus on gain trends in addition to its current data on output trends.

Evidence Supporting the Recommendations

Judgments about the progress on the reforms of 1988 and 1995 depend on information provided by the annual, systemwide student test score data from the norm-referenced ITBS test system. Using reading and mathematics scores for all students in grades two through eight from 1987 to 1996, this study sought to analyze trend data in a way that enabled the charting of school academic productivity since reforms were initiated in 1988. The data limitations, however, were considerable. The recom-

mendations above for a new assessment system grew out of concerns about the accuracy and suitability of the test score data, as explained in this brief.

Alignment with CPS Learning Standards

The adoption of CPS learning standards compels an assessment system that is aligned to these standards, if they are to be internalized by students, teachers, and parents. The implications of the new standards for a range of assessment options are not covered in this report, because the purpose here is to evaluate

An Inclusive Orientation

The design of new assessments should incorporate a strong commitment to measuring the learning progress of all students. Similarly, schools should be held accountable for the progress of all students, and procedures should be established to minimize the exclusion of students from the accountability system.

The accountability system must take better account of the learning gains of mobile students (i.e. students who have changed schools during the 12-month period prior to spring testing.) The attribution of these students' learning gains to schools, however, must be adjudicated in a fair manner. Procedures can be devised that are both fair to individual schools (i.e., including in their accountability report only the students who have been enrolled long enough for the school to have a measurable impact on their learning) and inclusive in orientation (i.e., where learning gains for over 90 percent of the CPS students are counted.)

Reporting test scores and learning gains for retained students is also an important issue. Incentives implicit in the current system are to retain students because their scores will then count against a lower grade level. Since the progress of retained students is especially important, the reporting system should give them explicit treatment. One possible method would be to analyze the progress of retained students separately (that is, disaggregated from the total); another possibility would be to report scores by age level of students rather than by grade level.

Because of the importance of the first years of schooling for developing a solid foundation of literacy and numeracy skills, Chicago schools need better assessment tools for evaluating the progress of primary grade students toward the CPS Learning Standards. Extant group administered standardized tests, however, are not adequate. A more comprehensive approach, using reliable and valid techniques, needs to be adopted for young students. We specifically caution that assessments in kindergarten and first grade have often been misused in the past to label and track students into weak instruction with low expectations for learning. While it is important that Chicago schools have better data for judging their efforts to educate students during the critical early years of schooling, such data should never be used to limit students' opportunity for learning.

The number of students in the CPS whose home language is not English is now one in six and rapidly growing. We lack good data on the academic progress of these students, and many are currently excluded from annual school accountability reports. The design of a new basic skills assessment should make provisions to meet the linguistic needs of students who are in the process of learning English. Similarly, special education students should be included. In this regard, new tests should be designed consistent with the guidelines for testing and reporting set out in the recent settlement agreement on the education of special needs students in the CPS.

the current use of a norm-referenced system to document progress on reforms and seek ways to improve them. However, it is obvious that assessments used for public reporting must reflect the learning expected under the standards. This would entail substantial changes from current practice.

A Content-Referenced, Rather Than Solely Norm-Referenced, Scale

Currently, the primary measure of school performance in the CPS is the "percent of students scoring at or above national norms." This statistic provides a picture,

although a crude one, of the very real systemwide gap between Chicago's public schools and the national norm, but it also fails to capture significant improvements in the learning of very low achieving and of higher achieving students. It mostly tracks the performance of those students who are clustered close to national norms—yet many Chicago schools enroll large numbers of students in the low-achieving end. Under the current system, schools can avoid sanctions best by focusing their efforts primarily on students close to national norms. This incentive system is skewed and could eventually lead to academic neglect of many CPS students. A

more accurate portrait of student academic progress requires a content-based assessment scale that tells how well all Chicago's students have learned what has been taught in classrooms.

This study explores the effect of changing to a content-based scale for comparing student performance across grades and years. The researchers took ITBS questions

of any new testing program. The post-hoc strategy employed here addresses some of the data problems, but cannot resolve them all.

A Stable Measurement Ruler over Time

The ITBS tests are relatively inexpensive and useful for telling how well students perform against a national sample of students who took the same test. This is how the "norm" is set. They are not appropriate, however, for the uses now being made of them—to assess improvements in school productivity over time.

Granted, the scores from these tests are reported in a way easy for teachers and parents to understand. They are translated into grade-equivalent scores, telling, for example, that 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 the same level of two different forms of the test. A second group received two different levels of the same form. Finally, for a 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 CPS91 (see Figure 1a) 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). 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 what kinds of problems do Chicago students of various grades demonstrate mastery? In what areas are students weak?

from various years and undertook an equating study to place them on a content-based scale. The scale is constructed around the relative difficulty of ITBS test items, regardless of the level or form of the test from which they were taken (issues discussed in the next section). A student's scale scores provide specific information about what the student knows and can do. Simply put, the study looked at items in the six different forms of the ITBS used from 1987 to 1996, equating them vertically (across grades) and horizontally (same grade but different years). A sophisticated statistical equating model used by such large-scale, modern testing programs as the SAT and the National Assessment of Educational Progress drew out the item difficulties. The process changed the ITBS scores into more content-explicit student achievement results. It was now possible to answer such questions as: On what kinds of problems do Chicago students of various grades demonstrate mastery? In what areas are students weak?

While the new scale is an improvement, the actual mechanism for test equating needs to be built directly into the design

The authors of the report would like to thank Anne Lewis for her help in crafting this policy brief.

on CPS91. These differential score effects are even more 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 achieving over time.

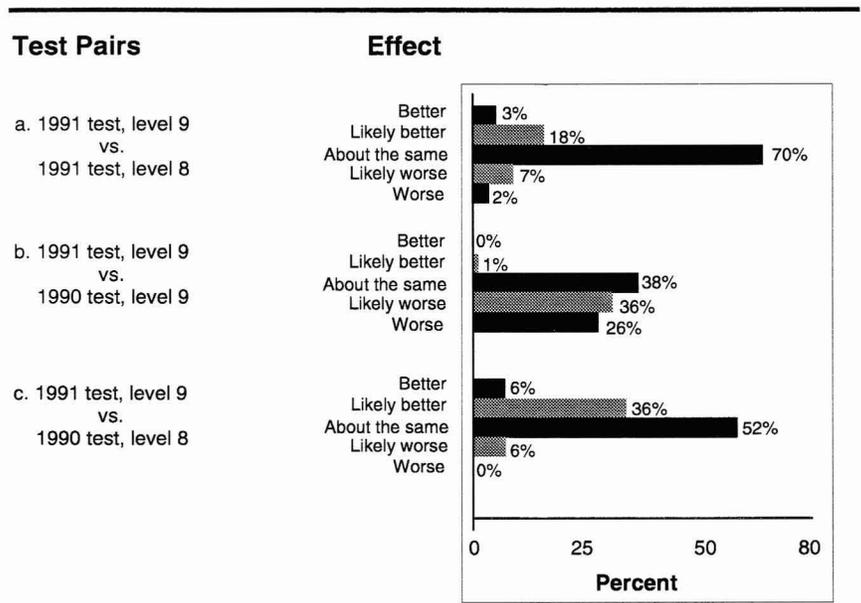
Another problem with the testing system is that the skills it assesses do change over time. Thus, the standards embedded in the tests are a moving target, but because the tests are “secure,” educators are unaware of the differences. The authors of this report, for example, grouped math test items into major categories such as addition, word problems, and data-related concepts, finding a clear shift in content beginning in 1993 that introduced new concepts and decreased the emphasis upon basic computation (see Figure 2). Even though a teacher may determine that his/her students know more math than the previous year’s students, their standardized test scores still may be lower. While the test developers obviously meant to reflect nationwide changes in math content, their tests necessarily must be written broadly to accommodate all types of school districts and cannot be tailored toward the curriculum of one specific school system.

In sum, the CPS relies on a testing and reporting system that is unstable over time and cannot fairly determine the progress—or lack of it—made by schools. Accountability of the schools requires a constant measurement ruler.

Focusing on What a School Adds to Student Learning

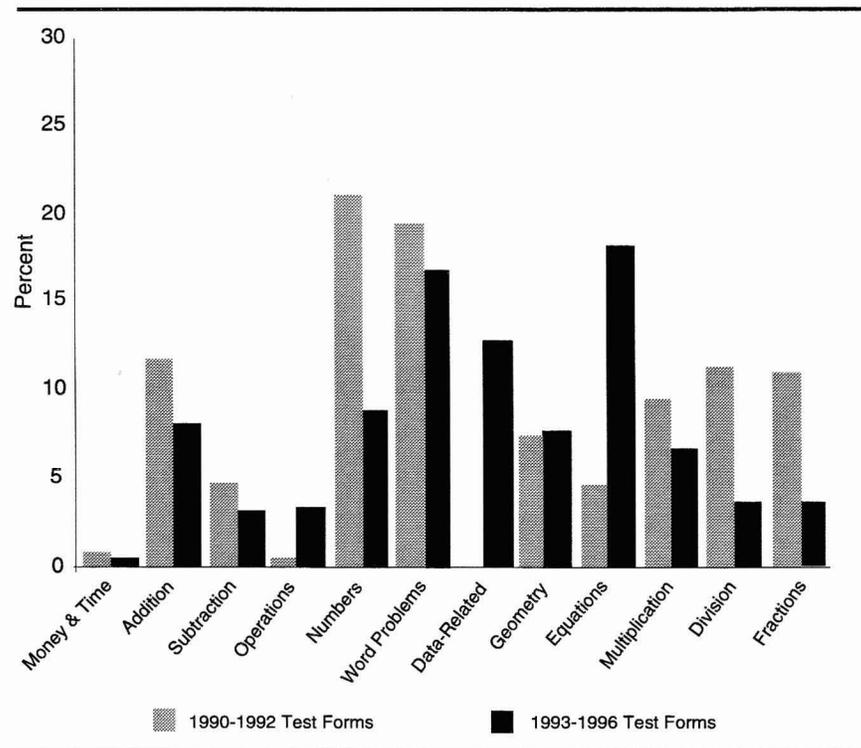
While the average achievement level is the best single statistical summary of the

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



Note 1: *About the same* category is +/- 1 standard deviation from zero.
 Note 2: See endnote 8.

Figure 2. ITBS Mathematics Content Changes: What the ITBS Tests in Grade 6



overall attainment of students in a school, it is not an especially good indicator of school productivity and whether this is changing over time. The problems associated with student mobility have been discussed in the recommendations. Changes in district policy, however, also can influence student achievement results over time. For example, if the procedures for testing bilingual students or the policies regarding grade retentions are changed, these could well affect the gains of students in schools with large numbers of bi-

Changes in district policy can influence student achievement results over time.

lingual students or students in the grade levels affected by the retention policies. Moreover, student demographics in individual schools can change over time, resulting in changes in achievement levels that do not really reflect greater or lesser productivity by the school.

The effects of such factors as student mobility or district policy changes need to be minimized in order to understand exactly what value is added to student learning by each school. Toward this end, this study has developed a productivity profile for each school that includes data on each grade, two through eight. The productivity profile is built from two basic pieces of information for each school grade: the input status and the learning gains.

The *input status* begins by identifying students who received instruction **for a full academic year** in each grade in each school. Then, their ITBS scores are retrieved from the previous spring.

The average of these students' previous year's test scores is the input status for that school grade.

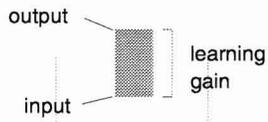
As for the *learning gain* for each school grade, this is simply how much the end of year ITBS results have improved over the input status for this same group of students. Finally, by adding the learning gain to the input status we recover the third piece of information— the *output status*. This tells us about the knowledge and skill levels of these students at the end of a year of instruction. In Figure 3, Panel 3a displays the base year input status (1991), learning gain, and output status for grade six at Fillmore School. Panel 3b adds the data for years prior to and post 1991, showing variations in learning gains over time.

A visual scan of Panel 3b suggests that the input status to grade six at "Fillmore School" may be declining over time. Countering this, the learning gains appear to be increasing and, with this, the outputs also appear to be increasing. To make this clearer, Panel 3c adds an input trend and an output trend to the profile.

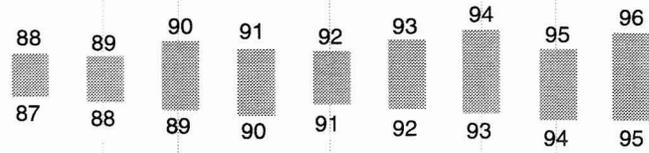
Because learning gain trends can vary considerably from year to year, overall trends tend to be obscured. To highlight the overall trends better, the study computes *smoothed trends* that involve estimating the best summary line that fits the data. These are presented in Panel 3d. To make the trends even clearer, Panel 3e presents the trend lines with the basic data removed. This method also allows adjustments for other factors that might be changing over time besides school effectiveness, such as changes in a school's ethnic composition, percentage of low-income students or retention rates. Generally, these did not make much of a difference in a school's profile; the adjusted trends are quite similar to the unadjusted estimates.

Figure 3. Constructing a Grade Productivity Profile for “Fillmore School”

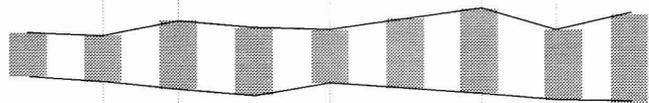
3a. Identify 6th grade scores. Match 5th grade scores from previous year. Include only students in same school.



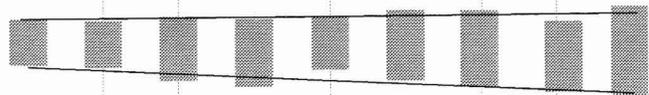
3b. Repeat for all years beginning in 1987.



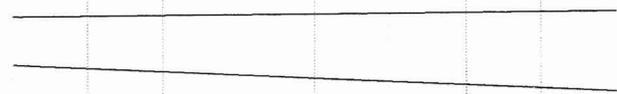
3c. Add input and output trends to the profile.



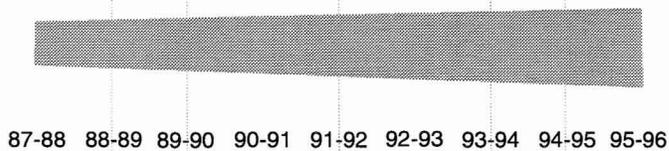
3d. Compute smoothed trends using "best fitting" summary line.



3e. To make trends clearer, remove basic data.



3f. Final Productivity Profile

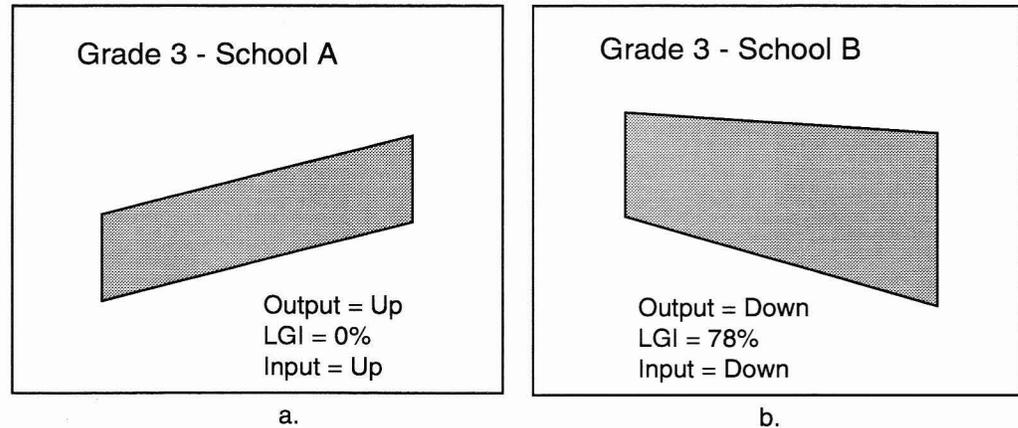


Each grade profile computed in this way consists of three trends: input, learning gain, and output trends (Panel 3f). Much of the recent literature on school accountability emphasizes only one trend—learning gains—to judge productivity, but this study contends that approach is too narrow on both educational and policy grounds. For example, suppose a grade shows a strong increase in the output trend, but the input trend is also increasing, so the estimated learning gain trend is zero (see Figure 4a). Focusing only on the learning gain trend might lead us to conclude that no significant change in

instruction had occurred in this school grade. In another example (Figure 4b), both the input and output trends were declining, but the input was declining at a faster rate. This would show up as improved learning gains, but it would not be a good example of what the reforms want to accomplish.

Productivity profiles are computed for each grade, but to create a more stable estimate of school productivity and to encourage faculty to accept collaborative responsibility for improving student achievement, this report recommends that grade level profiles be divided into three

Figure 4. Grade Productivity Profiles



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

groups: summary profiles for grades three to four, grades five to six, and grades seven to eight.

* * * * *

This study, on which these recommendations are based, set out to provide an accurate summary of the progress of Chicago schools over the last decade. Literally thousands of alternative analyses were conducted over several years. Nonetheless, we encountered the limitations of the extant testing system. The report indicates that much better information about productivity for policy and practice is possible—if the CPS could stretch

beyond the limits of its current testing system. It is frustrating to know what answers the CPS reforms need but not be able to develop them more accurately. This is not a criticism of the current testing system (ITBS) per se. Rather, the CPS is trying to use these tests for purposes for which they were not designed.

Even the sophisticated statistical designs used in the study are no substitute for a new testing system constructed to provide stable, reliable, inclusive, and content-related information useful to policy makers, educators, students, parents, and the public as they work to reform the schools.

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