
THE PREDICTABLE FAILURE OF CHICAGO'S STUDENT RETENTION PROGRAM

**Ernest R. House
Professor, School of Education
University of Colorado – Boulder**

November 1998

TABLE OF CONTENTS

SUMMARY	1
THE PREDICTABLE FAILURE OF CHICAGO'S STUDENT RETENTION PROGRAM	6
The New York City Retention Program	7
The Chicago Retention Program	9
Lessons from New York and Chicago	12
Retention/Promotion Decisions.....	12
Claims about Summer Progress.....	13
The Program Offered to Retained Students	16
The Costs of Retention Programs	17
National Research About Retention	17
Retention in Other Developed Countries	20
The Climate of the Times.....	21
Critical Next Steps.....	23
REFERENCES.....	26

SUMMARY

Chicago has initiated a massive program to retain low-achieving students in their current grade for another year. At first glance, retention seems to make good sense. The argument is especially persuasive when retention is contrasted to “social promotion,” passing students on when they have not learned much. However, this is a false choice. Neither retention nor social promotion is effective.

- Research about the negative effects of retention is so overwhelming that Chicago’s student retention program should never have been carried out.
- Retention has consistently failed to boost student achievement over the long term, and it makes students much more likely to drop out.
- Chicago justifies its retention program by arguing that Chicago’s program has unique features that will allow it to succeed where other student retention efforts have failed. However, there is nothing unique about the Chicago retention program that suggests that the ultimate results will be any different in Chicago.

In fact, the author led a team that evaluated a strikingly similar retention effort in New York City in the 1980s that led to the negative results for students consistently documented in research about retention.

The Failure of a Strikingly Similar Program in New York

In the 1980s, New York City carried out a student retention program called Promotional Gates that has all the major features of Chicago’s retention effort. In New York City:

- Students at grades 4 and 7 who failed to reach a minimum score on a standardized reading test were required to attend summer school. They were retained if they did not pass the cutoff score after summer school.
- Special help was provided to these students in classes of 18 or less during the school year they repeated. (At the time, the average New York City class size was 43 students.) The extra help provided to retained students required hiring 1,100 new teachers at a cost between of \$40 million and \$70 million per year.
- Although school system officials initially hailed Promotional Gates as a success, research proved otherwise. Claimed summer school test score gains actually resulted from mistakes in the school district’s analysis of its data.
- Analysis of the long-term performance of students who were retained shows that they achieved no better than similar low-achieving students who had been promoted before Promotional Gates was initiated.

- And students who were held back were much more likely to drop out later on than similar students who had not been held back.

The parallels to the Chicago retention initiative are obvious:

- In Chicago, students have been required to attend summer school at grades 3, 6, 8, and 9, if they failed to reach a minimum score on standardized reading and/or math tests. The summer curriculum is focused on the content of the test the students are expected to pass. Students who do not pass the cutoff score after summer school are retained.
- Special help for retained students varies from school to school in Chicago. For elementary schools with high percentages of retained students, the school system has hired 150 extra teachers for separate instruction in classes of 15. However, such classes serve less than 25% of the retained elementary students. The rest of the retained elementary students receive part-time tutors and after school programs.
- Eighth graders who are too old to be retained in elementary school are assigned to separate Transition Centers, each serving 200 students.
- As in New York City, the retention program at certain specified grades has encouraged increased retention at other grades (such as first and second grade).

There is nothing unique that is being done in Chicago that gives any reason to hope that Chicago's results will be any different from New York's. In fact, New York devoted a much higher level of resources to helping retained children than Chicago has.

Some Key Lessons from New York and Chicago

The similar retention programs in New York City and Chicago suggest a number of important lessons:

- Both New York and Chicago use standardized tests as the basis for making the retention-promotion decision. Yet the developers and publishers of the Iowa Test, which is used for decision making in Chicago, explicitly state that using their test as the primary basis for retention or promotion decisions is an "inappropriate purpose" for using their test.
- In New York, large gains that were claimed to result from the Promotional Gates summer school program proved to result from misanalysis of the data. Chicago's claims about large summer school gains in achievement are similarly suspect. The source of these claims is not an independent evaluation, but rather a school system press release. These gains may result (1) from common mistakes in analyzing test gains that the author's evaluation team found in New York, (2) from excessive "teaching to the test" that will not improve students' abilities to perform on other types of tests, and/or (3) from the dropout rate from the summer program.

- With respect to the shortcomings of teaching students to pass a particular test, recent experience in Kentucky provides a striking example. Kentucky students have been prepped for a particular high-stakes test with a distinctive format (KIRIS). Scores rose substantially on the KIRIS test, but only modestly on a national test with a somewhat different format (the National Assessment of Educational Progress).
- The need for an independent evaluation of Chicago's summer school program becomes even more critical because the claimed student test score gains for this summer program are being used to justify a systemwide curriculum — 9,360 daily lesson plans in every academic subject for the regular school year.
- Both New York and Chicago placed retained students into special separate classes with reduced class sizes. A lower teacher-student ratio is certainly desirable. However, placing low-achieving students in separate classes is not good practice. Students in separate schools and classes often become stigmatized and marginalized. For example, special educators have increasingly insisted that special education students be mainstreamed into regular classes. It seems strange that just as Chicago is abolishing separate classes for special education, it is setting up separate classes for retainees.
- Chicago's Transition Centers for overage eighth graders carry separation to yet another degree. It is critical that an independent evaluation be conducted immediately that focuses on how many students from these Transition Centers ever master academic skills and graduate from high school.
- Both the Chicago and New York retention programs entail huge extra costs. In Chicago, the costs of summer school, of over 10,000 elementary school students repeating a year of school at grades 3, 6, and 8, and of extra teachers for the retained students is well in excess of \$100 million per year. These large sums could be spent much more effectively on proven, research-tested initiatives for helping low-income students achieve better.

The Findings of Research Nationally

- There is a large body of research about student retention gathered over 40 years. Although retention sounds like common sense, this research indicates that few practices have such negative effects. This research consistently indicates that students who are retained do not achieve better on standardized tests after a few years and are much more likely to drop out later on. They also experience personal shame and depression.
- A recent study of Baltimore students who were retained claimed to present contrary evidence. But when the Baltimore data were reanalyzed properly, there was no benefit to retention found. (The authors of the Baltimore study acknowledged that the reanalysis was fair.)
- Among developed countries, retention is unique to the United States. In other countries, the practice is virtually unknown.

The Climate of the Times

- Retention is enormously popular despite its proven negative impacts. And Chicago's retention initiative has stimulated a national retention craze in big cities. Why are educators and policy makers so willing to carry out this harmful initiative yet again?
- After watching this process unfold repeatedly in different cities, the author concludes that Americans will support programs and policies that are harmful to minorities, especially African Americans, that they would not support if these same policies were applied to the general population. Neither New York City nor Chicago would have initiated massive retention programs if their students were not overwhelmingly minority.
- By contrast, a survey of fifteen Chicago suburban school districts indicates that those suburban districts typically retained fewer than one percent of their students. It is the inner city school districts with large minority populations where these harmful programs are implemented en masse.

Critical Next Steps

Given this research and these political realities, Chicago's educators, parents, and advocates for urban students should take a series of immediate steps to analyze and challenge current retention policies:

- First, you should publicize the negative research evidence about retention as it has been carried out nationally over a period of decades. Nothing is being done in the Chicago retention program that shows promise of leading to any different results than the strikingly similar retention program that was carried out in New York City twenty years ago.
- Second, you should insist that basic data about the process and impact of Chicago's retention program be immediately made public, so that any interested researchers may analyze it. Given the overwhelming negative evidence about retention, it is astonishing that this program has progressed for nearly three years with nothing more than contradictory and inconsistent data contained in school district press releases as the basis for judging its impact.
- Third, you should focus not only on the impact of retention at the grades where students are being retained as a matter of school system policy, but also on the ripple effect of increased retention at other grade levels.
- Fourth, you should demand that the school system stop using the Iowa Test and TAP Tests as the primary basis for making retention decisions, in ways that are judged inappropriate by the company that publishes these tests. And Riverside Press should be asked to terminate Chicago's use of this test, if Chicago does not immediately agree to stop misusing it.

- Fifth, you should identify better uses for the more than \$100 million that Chicago is now spending annually on retention. Low-achieving students need extra help. Neither retention nor social promotion works. Why not provide coordinated help with proven strategies, rather than those shown not to work?

Some preferable alternatives include fundamental restructuring of schools based on research about the practices of successful urban school where few students score in the bottom quartile on standardized tests; high quality early childhood education; and, as a last resort, promoting low-achieving students but providing them with extra help, rather than retaining them.

THE PREDICTABLE FAILURE OF CHICAGO'S STUDENT RETENTION PROGRAM

Once again we are faced with a massive program for flunking kids, retaining them in their current grades for another year. This time the program is in Chicago's public schools, though Chicago is not the first to employ retention. In fact, the research about the negative effects of retention is so overwhelming that Chicago's retention program should never have been carried out. Retention has consistently failed to boost student achievement, and it makes students much more likely to drop out of school. As I will explain, there is nothing unique about the Chicago retention program that suggests that the ultimate results here will be any different.

At first glance retention is a practice that seems to make good sense. If kids don't master the knowledge and skills they are supposed to, they will fall further behind their classmates. Why not hold them back for a year to catch up on the material they have not learned? Then they can make good academic progress. Retaining students will be beneficial, even if it costs the school district extra money.

This argument is especially persuasive when retention is contrasted to "social promotion," passing students on when they have not learned much, or watering down the curriculum to something they can pass. The President of the United States, the Mayor of Chicago, and the Illinois Board of Education oppose social promotion. And although there are effective alternative strategies for educating students other than social promotion or retention, often retention is justified as the only alternative to social promotion, a false choice.

During the past few decades, another, often unspoken, justification for retention has had a major impact. Students were seen by some as being out of control. One way to discipline them was to flunk them, put them on notice. This attitude has become especially strong in large cities, where the students are mostly African American and Latino. Embattled officials, frustrated with poor student

performance (and operating with declining resources) turned to tough love. Well, maybe it wasn't love, but it was meant to be tough — perform or flunk.

Unfortunately, flunking doesn't work. Whatever might be said for flunking students as an outlet for frustration, retention is a punitive policy which has the opposite effects of what it intends. The evidence against retention is unequivocal and overwhelming. The conclusion that retention does not improve student achievement over the long run and leads to a much higher dropout rate is one of the clearest findings from research. I will illustrate this finding by describing the history of New York City's massive retention program, which is strikingly similar to Chicago's and which I helped to evaluate.

The New York City Retention Program

In 1981 Frank Macchiarola, chancellor of the New York city schools, launched a large-scale retention program that held back 25,000 students the first year, one-fifth of the students in fourth and seventh grades, because they could not meet the cut-off scores on the citywide reading tests. New York City's program was called "Promotional Gates." Macchiarola himself came from a tough part of the city and figured that teachers and students were not trying hard enough. Tough standards would shape them up, he thought. Both students and teachers needed more discipline to improve their performances.

Once identified as being behind on the tests, students were sent to summer school. If they did not attain the required cut-off scores there, they were put in special classes with teachers who had been trained to use one of four compensatory education strategies that had previously been employed in New York. At that time New York City elementary classes averaged 43 students, eighty percent of whom were minority. Teaching students to read in classes of 43 was not easy. The retained students were placed in classes of no more than 18. The first year more than 1,100

additional teachers were required, at a cost of between \$40 and \$70 million. From what I could tell, the compensatory components of the Promotional Gates plan were carried out (New York City Schools, 1982, 1983).

Some city officials, especially Deputy Mayor Wagner, did not think the program would work. However, Chancellor Macchiarola was a friend of Mayor Koch, and he prevailed on the Mayor for support. As a compromise, the Mayor's office, which supplied half the funding for the schools, insisted that an evaluation be done. The next year the program was to be expanded to math, then to other subject areas and grade levels. The city hired me and two colleagues to oversee the evaluation (House, Linn, and Raths, 1981-1982).

When we arrived in New York, the district was proclaiming large test gains from their summer sessions, heralded as proof that the retention program was working. Several months test gain for a few weeks in summer school seemed too good to be true. Indeed, it was too good to be true. The district had made serious statistical errors in their analysis of the summer test data. When the proper calculations were done, the average student had made no gains at all.

After two years the test scores of the students who had been retained were compared to those of similar low-achieving students from previous years who had not been retained. It was found that there were no substantial differences between the students who had been retained and similar low-achieving students who had been passed in the years before the program. In other words, students did just as well if they were passed and received the education provided before Promotional Gates existed.

Furthermore, many students were failing to meet the cut-off scores even after a year of retention. Some were being retained in fourth and seventh grades for two or three years. The school district faced the prospect of having to promote these students or having students shaving in fourth grade. The Promotional Gates Program

began to look like the Boulder Dam program, with tens of thousands of students backed up at fourth and seventh grades.

About this time Chancellor Macchiarola was offered a job by billionaire David Rockefeller (which serves as a rebuttal to those who think nothing good comes from school reform). A succeeding head of the schools quietly put the Promotional Gates Program to sleep without fanfare. Several years later, evaluations by the New York school system indicated that the retained students dropped out at much higher rates than similar low-achieving students who had not been retained. Forty percent of those retained dropped out of school compared with 25% of those of a similar group who had not been retained (Pick, 1998). The Promotional Gates Program had retained tens of thousands of students at huge dollar and human costs without benefits.

The Chicago Retention Program

In 1996, twenty years after New York's failure, Chicago launched a very similar program. Chicago is also identifying tens of thousands of low-achieving students based on test scores and sending them to summer school. If they don't achieve minimum cut-off scores on the standardized tests after summer school, they are retained. The Chicago program is amazingly similar to the New York City program of the early 1980s. Let me point out some similarities, as well as a few differences.

The Chicago plan operates at grades 3, 6, 8, and 9, while the New York plan focused on grades 4 and 7, with intention of expanding. In Chicago, students in the transitional bilingual programs are excluded (one of six students), and most special education students are included (Chicago Public Schools, no date). In New York about 25% of students at grades four and seven were retained. No independent evaluation has been done of the Chicago retention program, though the Consortium on Chicago School Research has an evaluation in progress, and they have shared

some preliminary data. Of the Chicago students who took achievement tests in spring 1997, 17% of third graders were retained, 12% of sixth graders were retained, and 13% of eighth graders were either retained or sent to separate Transition Centers (Consortium on Chicago School Research, 1998). Newspaper reports indicate that about 74% of ninth graders who failed to pass the tests in the spring were retained because they failed the test again or did not retake the test during the summer (Hendrie, 1997).

In both New York and Chicago, 80% to 90% of the students identified for retention were low-income minorities. For example, the 90 Chicago elementary schools with the highest rates of spring test failure in 1997 were 69% African American, 27% Latino, 3% white, and 94% low-income (Designs for Change, 1998a). And Chicago's Transition Centers (separate schools for the eighth graders who qualified for retention but were too old to remain in elementary school) enrolled 929 African Americans, 330 Latinos, and 34 whites in the 1997-98 school year (Chicago Public Schools, September 30, 1997). In all, about 12,350 third, sixth, and eighth graders were retained in Chicago in summer 1997 (Consortium on Chicago School Research, November 1998), along with 10,600 ninth-graders (Hendrie, 1997).

The New York plan used single test cut-off scores on their standardized reading test (the Metropolitan Achievement Test) to make the retention decision. Chicago also uses a single cut-off score to identify retained students (on the Iowa Tests of Basic Skills, ITBS, and the Tests of Achievement and Proficiency, TAP). In Chicago, the minimum score to avoid retention was 2.8 for grade 3, 5.3 for grade 6, and 7.2 for grade 8 in 1997, essentially a year to a year and a half behind. Similarly, New York used one year behind at grade 4 and one and a half years behind at grade 7 as cut-offs.

New York used the same test form for both spring and summer testing. In contrast, Chicago has used two different forms of the Iowa Test for spring and

summer testing at a particular grade, although there is evidence that some grade-level forms of the Iowa Test being used in Chicago are not equivalent (Bryk et al., 1998). I am not clear exactly how these forms have been employed.

In Chicago, students must meet the test cut-off scores, not accumulate more than 20 days of unexcused absences, and have passing grades in reading and math. If they fail to meet one of these criteria, they must attend summer school and are at risk of being retained. Elementary and high schools are placed on probation in Chicago based on the percentage of students at or above the national norms on the spring tests. New York did not do this.

Students identified for possible retention were sent to summer school in both cities. If they achieved the required score there, they were promoted. Both cities claimed great gains for the summer sessions, months gained on test scores for a few weeks of attendance. The New York gains proved to be fallacious. The Chicago gains also appear suspect to an experienced evaluator, a point I will return to. The Chicago summer curriculum uses fixed lesson plans focused on the topics covered by the Iowa and TAP Tests. Teachers must follow these daily lesson plans, and monitors check to make sure teachers are on the right lesson. New York had a broader summer curriculum.

In New York students who did not achieve the required score in summer were sent to special classes of 18, and the teachers in these classes used one of four compensatory education programs that had previously been employed in the district. As noted earlier, teaching these classes required the hiring of 1,100 additional teachers. In contrast, Chicago students who did not exceed the cut-off score after summer school and had to repeat grades have received varying levels of extra help. For elementary schools where a high percentage of students have been retained, the school system has hired 150 extra teachers to provide instruction in separate classes of 15 students, and has provided after-school programs for the retained students

(Chicago Public Schools, Office of Management and Budget, November 3, 1998). In the retention classes, teachers employ a prescribed test-focused daily curriculum, similar to the summer school curriculum.

However, the roughly 2,250 students who can be served by the 150 extra teachers in classes of 15 are less than one quarter of the more than 10,000 students in third, sixth, and eighth grades who have been retained in Chicago's elementary schools each year for the past two years. For the rest of the retained students, the schools have been provided with part-time tutors and after-school programs by the central administration, but not additional teachers. Students who were too old to remain in eighth grade have been sent to separate Transition Centers, each serving 200 students. At these centers, students learn in classes of twenty in a mandated instructional program that includes double blocks of reading and math, plus, world studies, computer, and physical education (Fields, June 1997).

Another point about both the Chicago and New York retention initiatives is that they catalyzed increased retention at other grades. For example, a number of Chicago teachers and principals indicated that they intended to hold back more first and second graders as well. The retention rate in these other grades has not been released by the Chicago school system.

Lessons from New York and Chicago

What might we say about the retention programs in New York and Chicago?

Retention/Promotion Decisions

First, a single standardized test score is not a legitimate way to make the critical decision about whether a student should be retained or promoted. A single test score is too meager a piece of evidence. There is too much error associated with individual test scores to use them this way, and other information should be used in addition. The manual for the Iowa Tests specifically states that using this test as the

primary basis for retaining or promoting students is an “inappropriate purpose” for using the test (Hoover et al, 1996a). The head of Iowa testing, H. D. Hoover, has been quoted as saying,

A single test should never be used as the sole basis to make a decision such as promotion and retention. That’s because you have other information available from what the teacher knows....The teacher has been with the kids all year, and that should be taken into account (Newstips, September 3, 1998).

Presumably, the school district is in violation of the agreement it signed with Riverside Press, the publisher of the Iowa Test, which states that test purchasers must “avoid labeling students based on a single score” and “administer, score, interpret, and use tests exactly as specified in the manual.” (Riverside Press, no date). As H. D. Hoover states, one needs several pieces of information when considering retention, and the decision should be made with strong professional and parental involvement.

Claims about Summer Progress

Having the identified students attend summer school is a good idea used by both cities. Extra attention helps. However, the gains reported during the summer sessions are too large to be taken at face value. For example, in summer 1997, Chicago claimed reading gains of 4.4 months for third graders, 7.0 months for sixth graders, and 9.9 months for eighth graders who failed to meet the cut-off on the Iowa reading test in spring 1997, and then retook the test after the 1997 summer school. Yet these claimed results were not presented in an independent evaluation, but rather in a school system press release (Chicago Public Schools, August 14, 1997). How seriously can we take these claimed gains?

One possibility is that the data have been misanalyzed. A statistical phenomenon called regression toward the mean is a frequent problem in evaluating the progress of low-scoring students. If you separate out a lower scoring portion of

students, then give them the same test again, they will score higher on the average without receiving any instruction at all. This artifact can be corrected for, but often it is not. This error explained the summer test gains initially claimed for the New York retention program. Whether the same problem exists in Chicago depends on their test use and data analysis. This is only one of many such technical pitfalls which must be guarded against when tests are used in this fashion.

A second possibility is that the summer curriculum is so narrowly-focused that it is preparing students to pass one particular test with a certain format, without really helping students master general academic skills. What is wrong with this? Let's take an example from language arts. In the early grades, the Iowa Test in reading consists of short paragraphs, usually followed by two questions (Hoover et al., 1996b). Students are not required to write essays as part of the Iowa Tests. A curriculum focused on preparing students for short reading passages with multiple choice questions may not have a lasting effect on students' basic reading competence (for example, a student's ability to understand a short story several pages long or an instruction manual). And the fact that the Iowa Test does not require students to write will mean that writing is deemphasized in the test-preparation curriculum. Research indicates that students exposed to such a test-focused curriculum may not perform better on other kinds of tests a year or two later (Linn, 1998).

For example, Figure 1 reveals what has happened in Kentucky, which has a test-centered curriculum.

As the figure shows, students performed better and better on Kentucky's high-stakes KIRIS tests while at the same time failing to perform much better on the National Assessment of Educational Progress Tests (NAEP). As Bob Linn, the leading test expert in the country, says,

...reliance on a single test for repeated testing can distort instruction and lead to inflated and non-generalizable estimates of student gains in achievement (Linn, 1998, p. 7). Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high-stakes accountability uses are attached to them....Don't put all the weight on a single test (Linn, 1998, pp. 7, 28, 29).

In the extreme, teachers may obtain copies of the exact tests being used and teach children the specific answers to test questions. Although Chicago has used different versions of the Iowa Test, the versions used to judge summer school progress have all been used before in Chicago. Chicago has made some effort to guard against outright cheating by using alternative forms. So I have focused more on the problem of teaching to the test format. However, I can cite many examples of direct cheating elsewhere in the country. A good evaluation must take such possibilities into account.

A third problem with Chicago's claims about test gains is that a substantial number of students who failed to make the cut-off in the spring did not complete summer school and retake the test. In spring 1997, for example, Education Week used Chicago school system data to estimate that about 9,000 of the 41,000 students who failed the spring reading or math test at third, sixth, eighth, and ninth grades did not retake the tests in the summer (Hendrie, September 10, 1997). It is quite plausible that the students who were not retested were less motivated and had more severe learning problems than the students who completed summer school and were retested. This would inflate results. An impartial evaluation of the retention program must take into account all students who did not pass the test in the spring, including those who dropped out of the summer school process.

I might add that it is astonishing that a program that is so controversial and that costs so much money has not been independently evaluated after three years, and

that only press releases and conflicting data in newspapers constitute the available information about such a massive retention effort. The need for an independent evaluation of Chicago's summer school program becomes even more critical because the claimed test score gains are being used to justify a systemwide curriculum of 9,360 daily lesson plans in every academic subject for the regular school year (Duffrin, September 1998). Last week, the head of the Chicago school system responded to a study of elementary schools by using the alleged summer school gains to justify this lock-step curriculum.

Vallas said efforts are under way to make sure teachers in all schools are provided with lesson plans and curriculum standards that will put schools "on the same page." "That's why the eighth graders the last two summers have shown one year's growth in reading because of the improving quality of our curriculum...." (Metsch, November 6, 1998).

The Program Offered to Retained Students

As I pointed out earlier, both New York and Chicago have placed retained students into special separate classes with reduced class sizes. A lower teacher-student ratio is certainly highly desirable. However, placing low-achieving students in separate classes is not good practice. Students in separate schools and classes often become stigmatized and marginalized (Oakes, 1985). For example, special educators have increasingly insisted that special education students be mainstreamed into regular classes. It seems strange that just as Chicago is abolishing separate classes for special education, it is setting up separate classes for retainees (Martinez and Poe, November 7, 1997). Chicago's Transition Centers carry separation to yet another degree. It is critical that an independent evaluation be conducted that focuses on how many students from these Transition Centers ever master academic skills and graduate from high school. I would predict enormous drop-out rates.

The Costs of Retention Programs

Both the Chicago and New York retention programs incurred huge extra costs. Holding kids back means Chicago students will spend an extra year in school at \$4,641 per year per student (\$6,941 per year counting categorical programs). If we use the smaller per pupil figure, retaining 10,000 elementary students in Chicago is costing \$46 million this year alone. In addition, there are the costs of summer school and additional teachers for smaller classes during the regular school year. In New York, the cost was \$40 to \$70 million the first year for additional teachers (depending on who was estimating).

In Chicago the summer schools cost \$25 million in 1996, \$34 million in 1997, and \$42 million in 1998. Chicago's extra teachers and after-school programs for retained elementary students cost at least \$12 million. A conservative estimate is that Chicago's retention initiative is costing in excess of \$100 million per year. These are heavy extra costs indeed. After reviewing other research about the effectiveness of retention, I will return to the question of whether these huge outlays could be more productively spent in other ways.

National Research About Retention

If we move away from New York and Chicago, what does research from across the country say about retention? This research indicates that few practices have such negative effects. Researchers use a process called "meta-analysis" to combine data from a number of studies on a particular topic, like retention. Meta-analysis indicates that retention is either harmful or ineffective (Holmes, 1989). Students retained are a quarter of a standard deviation worse off on educational outcome measures than comparable students who are promoted. These negative effects are even stronger for academic achievement alone. When children of the same age were compared, the retained group lost .45 standard deviation in achievement on average.

Evidence indicates that failing a grade is strongly tied to dropping out of school later. Being retained is as strong as low achievement in determining whether a student drops out or graduates. For example, in Austin, Texas, repeating a grade increased the chances of a white female dropping out by 17% and increased an African American male's chances of dropping out by 38% (Grissom and Shepard, 1989). This is a very powerful negative effect.

In fact, previous research in Chicago's schools indicated the same thing: the level of reading achievement and the student being overage (an indicator of flunking) were the best predictors that the student would drop out (Hess and Lauber, 1985). The dropout rate was 37% for those not retained, 59% for those retained once, and 69% for those retained twice. This study also concluded that students who were retained in elementary school were more likely to drop out, even when the retained student was reading significantly better than a student who entered high school at the normal age.

Research about the effects of retention on personal adjustment also show negative results. In some cases the stigmatizing effects on the children are striking. In one study girls who had been retained refused to identify themselves as having been held back, even though they could name others who had been. Not only did students conceal that they had flunked, they were ridiculed by peers. Fully 84% reported feeling "sad, bad, or upset." Children said their parents were "mad" (48%) and "sad" (28%), and half reported being punished. Flunking evokes ridicule and punishment, shame and humiliation (Byrnes, 1989).

Not every study reports negative effects. Of 63 studies in the meta-analysis, nine reported some positive results. However, these retention practices were in suburban settings in which retained students were put in special classes, given lots of help, and mainstreamed. It might be the case that retention helps some students some times under some circumstances. In a review of twenty-one studies between

1982 and 1992, I found that two showed positive effects, nine showed no effects, and ten showed negative effects.

There is a recent challenge to the consistent conclusion that retention is ineffective or harmful for urban low-income students. Alexander, Entwistle, and Dunbar (1994) followed a sample of students in the Baltimore schools for eight years, 53% of whom were retained at some point. They claim their study shows that students who were retained gained academically and showed no ill effects in other ways. Unfortunately, the authors' own data do not support such claims. They became confused analyzing gain scores (tricky business) and derived conclusions contrary to their own data.

Shepard, Smith, and Marion (1996) subsequently converted the Baltimore data to national percentile ranks and within-grade standard scores. Their analysis showed that retainees did improve their test scores during the repeated year itself. For example, first-grade retainees went from the 18th percentile in reading comprehension at the end of their first year to the 59th percentile at the end of their retained year (taking the same curriculum and tests). This looks good at first. Such short-term gains are the reason that architects of retention programs can point to higher test scores in the first year or two.

However, the important question is whether the students maintained such an advantage later. Alas, they did not. They were back at their same lower percentile level by second grade. Shepard, Smith, and Marion (1996) concluded that there is no effect from retention one way or the other in the Baltimore study. (Alexander's response to this critique was to acknowledge that their re-analysis was "fair and accurate" but that Shepard and her colleagues exaggerated his enthusiasm for retention.)

Why would scholars arrive at conclusions about retention contrary to their own data? Alexander and his colleagues supply one clue:

We wonder whether the studies now in the literature overrepresent children who are especially susceptible to retention's damaging effects. Retention no doubt is harder on children in some circumstances...and it could work differently in schools where just a handful of retainees stand out in comparison with a large majority of successful students. Little retention research has been conducted on children like those in [the Baltimore study], that is, minority and disadvantaged student populations in urban school systems where retention rates are high overall, and where many students fit the so-called risk profile. ...it seems plausible that social stigma would not be as much of a problem. (Alexander et al., 1994, p. 216-217).

Indeed, 55% of the students retained in Baltimore were African-American. But is it not particularly dangerous to conclude that African-American children have different sensibilities so that they won't feel stigmatized or that retention won't harm them as it does others?

Retention in Other Developed Countries

U.S. views about retention contrast sharply with those in other countries. Few other countries tolerate flunking students. As one British educator observed to me: I've been thinking about the issue of holding students back a grade. I've talked to some people here about it but we simply have no experience of the practice. As far as I can tell, it doesn't happen here. As far as possible, there is a strong preference for keeping children in their age cohort, not least because of friendship groups. Some primary schools are vertically grouped (very common in rural schools), which is to say that a class will have children across the age range for that phase of education, e.g., 5-8 or 8-12. There are also occasions when children are moved up a year because they are too advanced for their age group and this is causing problems. Holding back, however, is very rare.

And an Australian educator suggested why Australian teachers do not endorse retention:

Because they have this view that any targets are going to be too simplistic, that they are never going to measure the full extent of a kid's capacity, that different kids learn at different rates and eventually catch up. We know there is strong evidence that girls perform better up to a certain age, then boys catch up and sometimes overtake them. Kids from working class backgrounds respond better to different styles so having simplistic targets doesn't help, doesn't give you a true picture. And, of course, the bottom line is it would cost a fortune to hold kids back. We would prefer to have remedial teachers than keep half a cohort back.

The Climate of the Times

If the evidence about retention is so overwhelmingly negative, why does this practice persist? We might learn from the nineteenth century science of “craniometry.” Scientists who were convinced that whites had superior intelligence and that men were smarter than women cast about for evidence to support their biases. They found that whites had larger brains than other races and that men had larger brains than women. Based on these findings, they developed a pseudo-scientific theory. Larger brains meant more intelligence, they argued — after all, that relationship seemed to hold from one species to another among animals. The scientists then developed methods to measure the brain sizes of humans, with the connection to intelligence that they thought this entailed (Gould, 1981). These findings appeared in technical journals and the popular press, much like Herrnstein and Murray’s The Bell Curve (1994).

Stephen Jay Gould (1981) studied these craniometric methods and concluded that precision of measurement could never overcome the inherent racism and sexism

implicit in the beginning assumptions. The scientists started with the assumption that the human races could be ranked on a linear scale of mental worth and explored any method they thought might yield the proper ranking. They explained away any exceptions they found contrary to their own beliefs. Although craniometry seems ludicrous now, gross injustices were perpetrated on vast numbers of people. From this history, we might draw two conclusions: first, the precision of methods is no guarantee of impartiality. Second, the ideological climate of the age can seriously affect the conclusions reached.

The prevailing climate regarding retention is favorable, from the President to the Mayor to the public, even to many educators. In one large city 74% of the principals, 65% of the teachers, and 59% of the parents thought students should "always" or "usually" be retained for lack of basic skills. Teachers worried, but reassured themselves that it would be to the children's ultimate advantage. Teachers have deep-seated beliefs about child development. Half in one study believed that children develop in a linear fashion, "unfolding" through set stages when they are ready, and that this unfolding occurs outside the control of teachers and parents. Schools with teachers of this view held back 30% of their students, while teachers who believed that they could influence children's academic progress held back less than 1 or 2% (Smith, 1989).

Many teachers endorsed retention, though not all practiced it. Many expressed the belief that next year the retained child would move from the bottom of the class to the top. Retention would save the child from becoming frustrated and failing in the future. Most teachers could not recall a single negative example of harmful repercussions, while reciting stories about children who had suffered after being socially promoted. They routinely misjudged retained children's feelings and the resistance of their parents. They located the child's inability to perform or behave properly in the child's psychological make-up rather than in the school (Smith, 1989).

There is another issue buried here. Sociologist William Julius Wilson contends that Americans will not support public policies that are seen to benefit minorities primarily (Wilson, 1987). For example, Americans will support social security because it is seen as benefiting all people, not just minorities. But they will not support programs like welfare thought to benefit minorities primarily.

I have a corollary to Wilson's thesis. Americans will support programs and policies that are harmful to minorities, especially African Americans, that they would not support if these same policies were applied to the general population. New York would never have had the Promotional Gates Program if minorities had not constituted 80% of the school population. Baltimore would not have carried out its retention program if its student population were not 68% minority. And Chicago would not have its retention program if Chicago's students were not 89% minority. By contrast, a survey of fifteen Chicago suburban school districts indicated that those suburban districts retained fewer than one percent of their students (Ryndar, 1997). It is the inner city with large minority populations where these harmful programs are implemented en masse.

Critical Next Steps

Given this research and these political realities, Chicago's educators, parents, and advocates for urban students should take a series of immediate steps.

First, you should publicize the negative research evidence about retention as it has been carried out over decades. The harm that it has caused nationally has been consistently documented by researchers, and nothing is being done in the Chicago retention program that shows promise of leading to any different results than the strikingly similar retention program that was carried out in New York City.

Second, you should insist that basic data about the process and impact of Chicago's retention program be made public immediately, so that any interested

researchers can analyze it. Given the overwhelming negative evidence about retention, it is astonishing that this program has progressed for three years with no more than contradictory and inconsistent data contained in school district press releases as the basis for judging its impact. Further, you should insist that a truly independent evaluation of the Chicago program be carried out. I know that the Consortium on Chicago School Research is conducting a study of Chicago retention, and they have shared some initial data. However, key data should be made generally available, so that many researchers can analyze Chicago's program, which has stimulated a national mania for flunking urban students.

Third, you should focus not only on the impact of retention at the grades where students are being retained as a matter of school system policy, but also on the ripple effect of increased retention at other grade levels. Side-effects are often the most important outcomes of programs even when unanticipated.

Fourth, you should demand that the school system stop using the Iowa Test as the sole basis for making retention decisions in ways that are judged inappropriate by the company that publishes the test and by the head of the testing program. And Riverside Press should be asked to terminate Chicago's use of this test if Chicago does not agree to stop misusing it.

Fifth, you should identify better uses for the more than \$100 million that is now being spent on retention. The focus of my paper has not been on alternatives to retention, but there are a number of research-based strategies that are effective alternatives to both retention and social promotion. These effective strategies include fundamental restructuring of schools based on research about the practices of successful urban schools where few students score in the bottom quartile on tests (Designs for Change, 1998b); high quality early childhood education (Epstein et al., 1996); and promoting low-achieving students but providing them with extra help, such as tutors (Smith and Shepard, 1989). You should fight for redirecting the huge

amount of money that is being spent on retention towards strategies that will really improve the achievement of Chicago's children.

REFERENCES

- Alexander, K. L., Entwistle, D. R., Daubar, S. L. (1994). On the Success of Failure. New York: Cambridge University Press.
- Bryk, A. S. et al (1998). Examining the academic productivity of Chicago Public Elementary Schools. Chicago: Consortium on Chicago School Research.
- Byrnes, D. A. (1989). Attitudes of students, parents, and educators toward repeating a grade. In Shepard, L. A., and Smith, M. L. Flunking grades: Research and policies on retention. London: Falmer Press. pp. 108-131.
- Chicago Public Schools (no date). 1997-1998 Guidelines for promotion in the Chicago Public Schools. Chicago: Author.
- Chicago Public Schools (1997, August 4). Eighth graders master promotion criteria: students at all levels show growth. Press release.
- Chicago Public Schools (1997, September 30). Student racial/ethnic survey. Chicago: Author.
- Chicago Public Schools, Office of Management and Budget (1998, November 3). Personal communication to Jessica Clarke, Designs for Change.
- Consortium on Chicago School Research (1998, November). Preliminary data from study of Chicago retention policies. Chicago: Author.
- Designs for Change (1998a). Characteristics of Chicago elementary schools with varying percentages of spring 1997 Iowa test failures. Chicago: Author.
- Designs for Change (1998b). What makes these schools stand out: Chicago elementary schools with a seven-year trend of improved reading achievement. Chicago: Author.
- Duffrin, E. (1998, September). 9,360 lesson plans on the way. Catalyst X (1), p. 6-7.
- Epstein, A.S., Schweinhart, L.J., and McAdoo L. (1996) Models of early childhood education. Ypsilanti: High/Scope Press.

- Fields, M. (1997, June). Students try to catch up in transition centers. Catalyst, VIII (9), pp. 23-24.
- Gould, S. J. (1981). The mismeasure of man. New York: Norton.
- Grissom, J. B. and Shepard, L. A. (1989). Repeating and dropping out of school. In Shepard, L. A., and Smith, M. L. Flunking grades: Research and policies on retention. London: Falmer Press., pp. 34-63.
- Hendrie, C. Chicago data show mixed summer gain. (1997, September 10). Education Week XVII (3), p.1.
- Herrnstein, R.J. and Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: The Free Press.
- Hess, G. A., and Lauber, D. (1985). Dropouts from the Chicago Public Schools. Chicago: Chicago Panel on Public School Finances.
- Holmes, C. T. (1989). Grade level retention effects: A metaanalysis of research studies. In Shepard and Smith, pp. 16-33.
- Hoover, H. D. et al. (1996a). Iowa Tests of Basic Skills: Interpretive guide for school administrators (Levels 5-14, Form M). Itasca, IL: Riverside Press.
- Hoover, H. D. et al. (1996b) ITBS: A message to parents. Itasca, IL: Riverside Press.
- House, E. R., Linn, R. and Raths, J.. (1981-1982). An Audit of the Evaluation of New York City's Promotional Gates Program. Four reports, October 1981, February 1982, April 1982, October 1982.
- Linn, R. L. (1998). Assessments and accountability. Center for Research on Evaluation, Standards, and Student Testing. Boulder: University of Colorado.
- Martinez, M. and Poe, J. (1997, November 7). Doors open to disabled students. Chicago Tribune, Sec. 2, p.1.
- Metsch, S. (1998, November). Chicago schools criticized: Unchallenging work and repetition seen. Daily Southtown, p. 1.

- New York City Schools, Office of Educational Evaluation (1982). The promotional gates program: Mid-year assessment and analysis of January, 1982, test results.
- New York City Schools, Office of Educational Evaluation (1983). The 1982-1983 promotional gates program: Mid-year assessment and analysis of August, 1982, and January, 1983, test results.
- Newstips (1998, September 3). Questions mount over CPS's use of standardized tests. Chicago: Community Media Workshop, p. 1.
- Oakes, J. (1985). Keeping track: How schools structure inequality. New Haven: Yale University Press.
- Pick, G. (1998, April). Strict retention comes, goes, LA, NYC wary. Catalyst IX (7), pp. 8-9.
- Riverside Press (no date). Test purchaser qualification form. Itasca, IL: Author.
- Ryndar, R. (1997, August 25). Suburban schools are leery of making kids repeat grade. Chicago Sun-Times, p. 1.
- Shepard, L. A., and Smith, M. L. (1989). Flunking grades: Research and policies on retention. London: Falmer Press.
- Shepard, L. A., Smith, M. L., Marion, S. F., (1996). Failed evidence on grade retention. Psychology in the Schools, 33(3), 251-261.
- Smith, M. L. (1989). Teachers' beliefs about retention. In Shepard and Smith, pp. 132-150.
- Smith, M. L. and Shepard, L. A. (1989). Flunking grades: A recapitulation. In Shepard and Smith, pp. 214-235.
- Wilson, W. J. (1987). The truly disadvantaged. Chicago: University of Chicago Press.