

Retaining Students in Grade:

Consequences for Florida

Policy Brief

Mary Lee Smith

Arizona State University

Education Policy Research Unit (EPRU)

Education Policy Studies Laboratory

College of Education

Division of Educational Leadership and Policy Studies

Box 872411

Arizona State University

Tempe, AZ 85287-2411

April 2004

EPSL | EDUCATION POLICY STUDIES LABORATORY
Education Policy Research Unit

EPSL-0401-114-EPRU

<http://edpolicylab.org>

Education Policy Studies Laboratory

Division of Educational Leadership and Policy Studies

College of Education, Arizona State University

P.O. Box 872411, Tempe, AZ 85287-2411

Telephone: (480) 965-1886

Fax: (480) 965-0303

E-mail: epsl@asu.edu

<http://edpolicylab.org>

Retaining Students in Grade:

Consequences for Florida

Mary Lee Smith

Arizona State University

Retaining Students in Grade:

Consequences for Florida

Mary Lee Smith
Arizona State University

Executive Summary

Retaining students in grade and using proficiency standards to determine students' progress through grades was intended to make schools accountable and increase academic achievement. Research evidence demonstrates that grade retention does not, in fact, improve achievement. Further, a student who has been retained stands a much greater chance of dropping out of school instead of persisting to graduation. Retaining students in grade is both expensive and ineffective, and there are effective alternatives to it. Abandoning this policy in favor of more productive and proven policies will save money and enhance the education and life chances of Florida's students.

Recommendations

This brief recommends that:

1. Florida legislators repeal the relevant section of Statute 1008 so that pupils' progression from grade to grade is disconnected from their FCAT scores.
2. Education policymakers abandon the policy of retaining students in grade because it fails the test of best evidence and cost-effectiveness.

Retaining Students in Grade:

Consequences for Florida

Mary Lee Smith
Arizona State University

Section 1: The Issue

In Chapter 1008 of Florida Statutes (2003) the state expressed its intent to end “social promotion” and determine students’ progression from grade to grade by their performance on tests of academic achievement.¹ The Florida Comprehensive Assessment Test (FCAT) is the principal instrument to measure proficiency: a “proficiency standard” is set for each grade level. Students scoring above the cutoff score advance to the next grade. The students scoring below the cutoff are targeted for intensive remediation of their low achievement in reading, math, and writing. Third-grade students scoring below the proficiency standard cutoff on the reading test must be retained. Schools must identify students whose results are below the proficiency standard and concentrate their resources on remediation. The state must monitor and enforce school compliance. Few exemptions or alternate methods of demonstrating proficiency are permitted. The FCAT proficiency standard makes no allowance for the standard error (the normal variation around any score on a test) of FCAT scores or for the cutoff score. This combination of one absolute standard, the strict monitoring, and the tight enforcement make the Florida policy one of the most stringent in the nation.

Section 2: Background

Eliminating “social promotion” has considerable intuitive appeal. It is just common sense that passing students along without their having attained grade-level skills condemns them to worse problems in subsequent grades. They are likely to be illiterate when they graduate and apply for jobs. Even the label “social promotion” is associated with a soft and fuzzy focus on students’ self esteem rather than a rigorous focus on academics. Surveys typically show that teachers and the public subscribe to this view.²

“Ending social promotion” makes good political sense, as well. Presidents of both parties since Ronald Reagan have endorsed it. Many states have adopted similar policies. To most voters, a politician who calls for retaining students in grade is signaling a concern for schools and tough standards. Policies such as these seem to have lower costs than policies to reduce class size or implement universal pre-school. The theory behind all such policies is that punishment and fear of punishment make people work harder.

All this common and political sense, however, runs counter to 50 years of research evidence, which indicates that retention is no better than promotion in students’ subsequent academic performances. In other words, the intent of the policy—to improve achievement and make schools and students more accountable by retaining students in grade—has little relationship with the actual outcomes of that policy. Even efforts to remediate failing achievement, such as the summer academic camps that Florida mandates, are often no more effective than promotion.

Section 3: Available Data

Evidence that pertains to this policy falls into three categories: a) historical records, b) comparative studies, and c) longitudinal studies. Research on the outcomes of summer programs and on the use of test scores to make decisions on pupils is also presented.

Historical records

New York City implemented a promotional gates policy in 1981. Nearly one-quarter of the students in the gates grades were retained and placed in remediation classes. Over one thousand teachers had to be added and class size soared to 43 students per teacher. The program cost over \$40 million. Evaluation results showed that students retained were no better off academically than those who were passed along. Eventually the gates program was abandoned. The New York policy had nearly the same result everywhere it was tried.³ Initial data from Florida indicate that almost one-quarter of third graders fell below the cutoff score. Most of those who attended summer camp failed the test again. The costs of this policy—the per-pupil expenditure multiplied by the number of students retained each year—are absorbed by local districts and therefore hidden. Retainees pile up in third-grade classes, and third-grade teachers are apt to request transfers to other grades. Children of color are much more likely than others to be retained.

Comparative studies

Studies on the effects of retention/promotion follow this design: Researchers identify a group of students who have been retained in grade. Then they find another group, matched in relevant characteristics to the retained group, which had instead been

promoted to the next grade. Usually, the groups are matched on achievement test scores so that they were equally low before the retention or promotion occurred. Then, the achievement test scores of the two groups are compared at the end of the following year or at the end of the next grade. Two meta-analyses have been conducted, one on the studies up to 1989 and the other on studies conducted between 1990 and 1999.

In the first meta-analysis,⁴ Holmes found that the overall difference in mean achievement across 63 studies was $-.31$. This means that, one or two years later, the group that had been eligible to be retained but was instead promoted had a better level of average achievement than the peer group which had been retained. Fifty-four of the 63 studies favored the promoted group. Such a one-sided division of studies is rarely observed in education research. Holmes also found that the better controlled the study was, the more likely its results would be favorable for the promoted sample. Any advantage for the retained group in the first year was wiped out further down the line.

In the second meta-analysis,⁵ Jimerson integrated 18 studies of the effects of retention conducted between 1990 and 1999. Only about 10 percent of the studies favored the retained group. The overall effect size was $-.39$. This means (roughly) that after retention or promotion the retained group members were about three school months behind their initially comparable peers. Any early advantage that had been observed for the retained group disappeared over time.⁶

Longitudinal studies

This research employs large representative samples of high school age students and links their academic outcomes to various events and conditions they encountered along the way. Grissom and Shepard, and separately, Rumberger,⁷ identified databases

of students who had been retained in particular grades or who were otherwise old for their grades; these researchers then statistically controlled for differences in academic achievement, gender and ethnicity, and compared their academic outcomes. These analyses showed that students who had been retained were much more likely to have dropped out of school than their peers who were of normal age for their grades or who had been promoted.

All of the research of this type shows the same pattern of results. Estimates of the effect of retention on dropout rates vary. Some calculate that retention increases the risk that a student will drop out by 20 to 50 percent. Retention is the single largest predictor of dropping out, across all categories of ethnicity, gender, and poverty.

Research on summer school remediation

Florida requires that third graders who fail to meet the proficiency standard must participate in the summer academy “camps” and then take the test again. Research on summer school is mixed. In Chicago, where this option was tried, the majority of students failed the test again after taking summer school. An important element in the effectiveness of summer school is the quality and type of program offered. If the curriculum is solely aimed at teaching to the test, any benefits are likely to be temporary and illusory.⁸ If the curriculum is focused on intensive training in phonics, the results are likely unpredictable, but certainly not guaranteed.⁹

Research on the unreliability of test scores

According to the consensus of experts in psychometrics (the science of measurement), any test score has a component of error. Any cutoff score also has a component of error. This means that scores of a particular pupil fluctuate day to day or

across different equivalent items. A pupil who scores above the cutoff score may score below it on another day, even though his or her reading skills have not changed. Some states acknowledge this inevitability and make allowances for this band of error; these states make decisions about passing or failing only those students outside the band of error. Florida does not make such allowances. Experts in psychometrics recommend against using a single test score as the criterion for any major decision about a pupil. This is the consensus in the field, as demonstrated by the standards published by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. The National Research Council, an arm of the National Academy of Sciences, concurred and recommended that single tests scores never be used to determine grade progression.¹⁰

Section 4: Quality of Available Data

Decisions about the need for “evidence-based” or “science-based” programs are based on the results of randomized field trials (such as those employed to test the effectiveness of drugs or medical procedures). In the comparative research described in Section 3, researchers did not randomly assign low-achieving students to be retained or promoted. Instead, researchers used samples of students whom the school had designated in each category. Researchers then used post-hoc matching to try to control pre-existing differences. Although this design falls short of the true experiment alluded to as “science-based,” it would be a mistake to judge these studies as deficient or of poor quality. First, to randomly assign students to a treatment well known to produce negative effects violates researchers’ standards of the ethical treatment of research subjects, particularly minors. Second, the available studies have the advantage of taking place in

real settings with real students; whereas randomized experiments necessarily involve a degree of artificiality. Third, the sheer weight of evidence, and the uniformity of its direction, goes some way toward offsetting the lack of randomization of groups. Fourth, this is the same design that allows most of us to conclude that smoking causes cancer. The longitudinal studies represent the best practices of survey research: carefully measured variables, nationally representative surveys, and statistical controls over relevant analyses.

Section 5: Findings

Evidence from controlled and longitudinal studies shows conclusively that retaining students in grade has little chance of improving achievement. Students who have been retained have no better academic achievement than initially comparable students who have been promoted. Although the comparative studies show no immediate academic benefits of retention over promotion, the longitudinal research shows substantial risks of retaining students. Students who have been retained are much more likely to drop out of school rather than persist to graduation. The evidence is remarkably one-sided: no short-term benefits, substantial long-term risks, substantial costs to taxpayers, and, because students of poverty and color are disproportionately represented among retainees, a failure of social justice. Programs to remediate academic failures, such as Florida's summer camps, have a poor track record because they focus too narrowly on minimum academic competencies and test-taking skills.

Section 6: Recommendations

The overwhelming weight of the evidence makes prompt and decisive action necessary. Florida is advised to move from policies that punish students for failing to policies that prevent students from failing in the first place. Policies supported by research that would increase the achievement of students most likely to suffer academic failure include reducing class size and ensuring a rich literacy experience by encouraging parent literacy and providing tutoring, universal preschool and all-day kindergarten.

It is recommended that:

1. Florida legislators repeal the relevant section of Statute 1008 so that pupils' progression from grade to grade is disconnected from their FCAT scores.
2. Education policymakers abandon the policy of retaining students in grade because it fails the test of best evidence and cost-effectiveness.

Notes and References

¹ Florida Statutes (2004). Title XLVIII K-20 Education Code, Chapter 1008.25(6). Retrieved from http://www.flsenate.gov/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=Ch1008/SEC25.HTM&Title=->2003->Ch1008->Section%2025

The term “social promotion” is controversial because it seems to suggest that schools promote failing students solely for social reasons. This brief makes clear that promotion is no worse than retention *on academic grounds alone*. The statute reads as follows:

- (a) No student may be assigned to a grade level based solely on age or other factors that constitute social promotion.
- (b) The district school board may only exempt students from mandatory retention, as provided in paragraph (5)(b), for good cause.

² Tomchin, E. M., & Impara, J. C. (1992, Spring). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal*, 29, 199-223. Tomchin and Impara's survey revealed that most teachers believed retention prevented a student's inevitable, subsequent academic difficulties.

³ House, E. R., et. al (1982). Reports on New York City's Promotional Gates program. New York, Office of Program Evaluation, New York City Schools, Unpublished Report. The failure of the policy in 1980s did not prevent New York City from again trying to end social promotion in 2001.

⁴ Holmes, C. T. (1989). Grade-level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16-33). London: The Falmer Press.

⁵ Jimerson, S. R. (2001). Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century. *School Psychology Review*, 30(3), 420-37. Note that I am ignoring the social and behavioral outcomes computed in the meta-analysis, not because they are unimportant but because the Florida policy concentrates on academic achievement.

⁶ Alexander, K., Entwisle, D., & Dauber, S. (1994). *On the success of failure: A reassessment of the effects of retention in the primary grades*. New York: Cambridge University Press. This study is often mentioned as an example of studies that contradict the trends in the meta-analysis. In that study, the retained students were measured against themselves. That is, their subsequent achievement was compared against the trajectory of what their achievement would have been if they had not been retained. The authors conclude that retention halted the downward trajectory.

For several alternative hypotheses for this result see:

Dawson, P. (1998). A review of “On the success of failure: A reassessment of the effects of retention in the primary grades.” *Communique*, 26, 20-21.

⁷ Grissom, J., & Shepard, L. (1989). Repeating and dropping out of school. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16-33). London: The Falmer Press.

Rumberger, R. W. (1995). High school dropouts: A review of issues and evidence. *Review of Educational Research*, 57(2), 101-121.

- ⁸ Koretz, D., Linn, R. L., Shepard, L. A., & Graue, E. (1991). The Effects of High-stakes testing on Achievement: Preliminary Findings about Generalizations Across Tests. Annual Meeting of American Educational Research Association, Chicago. They showed that gains on the taught-to test did not generalize to another test of the same content domain. In other words, reading test scores might go up, but reading ability does not.
- ⁹ Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching Children to Read: The Fragile Link Between Science and Federal Education Policy. Education Policy Analysis Archives. Retrieved from <http://epaa.asu.edu/epaa/v11n15> They showed that the claim of phonics researchers and advocates, based on the National Reading Panel's meta-analysis, was not supported. Nor were the studies particularly scientific, according to the standards of scholarship.
- ¹⁰ American Educational Research Association, American Psychological Association, & National Council of Measurement in Education (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.
- Heubert, J. & Hauser, R. (Eds.) (1999). *High stakes: Testing for tracking, promotion, and graduation*. Committee on Appropriate Test Use. Washington, DC: National Academy Press.