Abstract Title Page
Not included in page count.

Title: The Causal Effects of Grade Retention on Behavioral Outcomes

Authors and Affiliations:

Lou Mariano, RAND Corporation
loum@rand.org

Paco Martorell, UC Davis, School of Education
pmartorell@ucdavis.edu

Abstract Body

Background / Context:

Requiring a failing student to repeat a grade is, of course, not a novel idea. However, the widespread availability of standardized assessment results for all students in grades 3-8, as mandated by the No Child Left Behind Act (and adopted earlier in some localities including NYC), makes it practical to establish uniform, objective policies for deciding whether and when to retain students.

The theory behind test-based promotion policies is that students who fail to demonstrate a sufficient understanding of their current grade's curriculum lack the prerequisite proficiency to fully engage in the following grade. This view recognizes the cumulative nature of the curriculum across grades. By providing an additional year of instruction in the current grade, the student is afforded the opportunity to improve their proficiency and more fully engage in the curriculum of the following grade; i.e., it ensures "that students are promoted only if they are prepared for higher level work" (New York City Department of Education [NYCDOE], 2009). That improved preparedness is expected to lead to higher achievement in the following grade, promulgating into subsequently improved preparedness and achievement in later grades—and ultimately a greater chance that the student will complete high school and be better prepared for long-term success.

Despite the possibility that grade retention could help struggling students, a large prior literature on the effect of grade retention, summarized in meta-analyses by Holmes (1989) and Jimerson (2001), finds that retention is negatively associated with academic performance and increases the potential for dropping out of high school. However, more recently this earlier literature has been criticized as having serious methodological flaws (cf. Alexander, Entwisle, & Dauber, 2003; Allen, Chen, Willson, & Hughes, 2009; Hong & Raudenbush, 2005; Lorence, Dworkin, Toenjes, & Hill, 2002). The primary empirical challenge facing these studies stems from the fact that "retention may reflect a subjective decision-making process based on a variety of factors" (Jimerson, Carlson, Rotert, Egeland, & Sroufe, 1997, p.4).

The use of standardized standards-based assessments as a uniform basis for retention eligibility allows for the implementation of quasi-experimental designs in evaluating the impact of retention. Studies using this framework and examining students in Chicago (Jacob & Lefgren, 2004; Roderick & Nagaoka, 2005) have found small benefits (at most) of grade retention, with the positive effects concentrated among younger students and dissipating quickly. On the other hand, studies using data from Florida (Greene & Winters, 2007; Schwerdt & West, 2013), Texas (Hughes, Chen, Thoemmes, & Kwok, 2010), and New York City (Mariano & Martorell, 2013) find evidence of much larger positive impacts that persist for at least several years.

While much of the prior literature has focused on short- to medium-run impacts on standardized test scores, critics of test-based grade retention decisions have focused on the negative consequences stemming from the punitive nature of grade retention. In particular, being a year older and a year behind one's peers may result in disengagement with school (Jackson, 1975; Roderick, 1994) that manifests itself in short- and longer-term behavioral problems (Byrd, Weitzman, & Auinger, 1997).

However, to our knowledge, very little causal evidence exists on whether grade retention generates negative behavioral outcomes, despite recent evidence linking behavioral problems in school to significantly lower wages and earnings in adulthood (Segal, 2013). Moreover, relatively little quasi-experimental evidence exists about how the impact of retention varies by the grade of retention, even though there are strong empirical and theoretical reasons to believe

that the impact of educational interventions such as this may vary with age. For instance, recent research has shown that interventions that occur earlier in life may be more effective than interventions that occur later in life (Cunha & Heckman, 2006).

Purpose / Objective / Research Question / Focus of Study:

This study examines the impact of grade retention under a comprehensive student promotion policy instituted by the NYCDOE. Building off earlier work that examined impacts on test scores, the current study aims to examine effects on various measures of behavioral outcomes (described below). We will also investigate how these effects vary by the retained grade, given that the punitive nature of grade retention may become more important the older a student is. To isolate the causal effect of grade retention, the study uses a quasi-experimental regression discontinuity research design (as described in detail below).

Setting:

The setting for the study will be NYCDOE public schools.

Population / Participants / Subjects:

We will use data on students in all grades subject to the NYCDOE promotion policy from 2003–2004 through the 2011–2012 school year. We will have outcome data through 2012-13 incorporated into the analysis by the time of the SREE conference in March 2015. Table 1 identifies each cohort in the sample, along with the 2012–2013 grade for both promoted and retained students.

(please insert Table 1 here)

Each cohort contains approximately 54,000 to 63,000 general education students subject to the promotion policy. The percent retained ranges from one to six percent for each cohort. The NYCDOE has provided administrative data for each student in each cohort subject to the policy. We will focus on students who took the assessment given to students enrolled in the NYCDOE's mandatory summer school program, because, as we describe in the next section, these were the students who were "at risk" of being retained.

Intervention / Program / Practice:

The NYCDOE implemented a new assessment-based promotion policy for general education students in grade 3 in 2003-04. This policy was extended to grade 5 in the fall of 2004, to grade 7 in the 2006–2007, to grade 8 in 2008–2009, and finally to grades 4 and 6 in 2009–2010. Students in charter schools, special education students, and early English language learners are exempt from the policy.

The policy's central feature is its reliance on standardized test scores for grade promotion decisions. Students in the lowest assessment performance category (Level 1) on either the mathematics or English Language Arts (ELA) state spring assessments are at risk of being retained in grade under the policy, while those who score in the next highest category (Level 2: meets some of the standards or partially meets the standards) on both subjects are eligible for promotion. The policy provides students multiple attempts to demonstrate eligibility for promotion. Students with Level 1 scores on the mathematics or ELA spring assessment can demonstrate Level 2 proficiency through a portfolio review in June. Those who do not

demonstrate Level 2 proficiency attend the City's summer instructional program. At the SSA's conclusion, students take a city summer assessment in the subjects in which they scored at Level 1 on the spring assessment. Those not demonstrating Level 2 performance on the summer assessment have the opportunity to do so through an August portfolio review, and those not demonstrating Level 2 or better performance by the end of this process are retained in grade. An exemption to retention may be granted if the student's principal and community superintendent deem it appropriate.

Despite multiple opportunities to meet the promotion criteria, scoring below the Level 2 cutoff for the summer assessment in either math or ELA sharply increases the likelihood of being retained. This can be seen in Figure 1, which plots the fraction of students retained as a function of the standardized score on the summer assessment. In the next section, we describe how we will exploit this variation to isolate the causal effect of grade retention.

(please insert Figure 1 here)

Research Design:

We use a fuzzy regression discontinuity (RD) research design (Hahn, Todd, & Van Der Klaauw, 2001; Imbens & Lemieux, 2008) that exploits the fact that grade retention is largely determined by whether a student scores below the Level 2 cutoff on the summer assessment. While on the whole students scoring below Level 2 are likely to be different in many ways from those scoring higher, these differences are likely to be much smaller among those students scoring close to the Level 2 cutoff. In fact, under plausible and empirically supported assumptions (which we describe in detail below), students scoring just above or just below the Level 2 cutoff are likely to be similar in all other dimensions. This reasoning suggests that comparisons between students who score just above and below the Level 1 summer assessment cutoff can be used to identify the effect of grade retention.

More concretely, the research design we use will use Level 2 status on the summer assessment to be used as an instrumental variable (IV) for grade retention. The idea behind this approach is that the discontinuity seen in Figure 1 is a source of variation in grade retention that is not related to confounding factors (such as socioeconomic status or academic motivation), at least near the Level 2 cutoff. If this condition holds, any discontinuity in a given outcome at the Level 2 cutoff will reflect a causal impact of grade retention. The magnitude of this causal impact can then be estimated by relating the size of the discontinuity in the outcome at the Level 2 cutoff to the corresponding discontinuity in the likelihood of retention.

The estimation equations take the form:

$$(1) \qquad Y_i = \theta W_i + f_Y(X_i) + \varepsilon_i$$
$$W_i = \pi T_i + f_W(X_i) + v_i$$

where, Y_i represents an outcome of interest, W_i denotes grade retention status, X_i is the summer assessment score, $f_Y(X_i)$ and $f_W(X_i)$ are flexible functions that describe the relationship between the X_i and the outcome and retention, respectively, away from the Level 2 cutoff, and ε_i and are residuals. The variable T_i is a dummy variable for scoring below the Level 2 cutoff, and serves as the instrumental variable for W_i. It can be thought of as representing "eligibility" for the grade retention treatment. The parameter π represents the "first-stage" discontinuity seen

in Figure 1, and the parameter θ represents the effect of grade retention (which we assume to be constant for expositional purposes only; we describe the more realistic case of heterogeneous effects below).

The validity of this approach rests on two key conditions. The first is that barely falling below the Level 2 cutoff affects the probability of retention (i.e., $\pi \neq 0$). Figure 1 provides strong empirical support for this condition. The second is that barely falling above or below the Level 1 cutoff only affects Y_i by changing the probability of being retained (i.e., that T_i and ε_i are uncorrelated). This assumption would be violated if students can manipulate their exact score relative to the Level 2 cutoff. We argue against such manipulation on a priori grounds since the tests are machine scored, and neither the summer school instructors nor the students know the exact number of questions that must be answered correctly in order to meet the Level 2 cutoff. We also find empirical evidence consistent with the absence of such manipulation. First, the density of test scores is "smooth" through the Level 2 cutoff (Figure 2), suggesting the absence of explicit sorting at the cutoff. Second, we find no indication that pre-determined variables such as spring math assessment scores "jump" discontinuously at the Level 2 cutoff (Figure 3).

(Please insert Figures 1 and 2 here)

Data Collection and Analysis:

Data will include summer assessment scores in ELA and mathematics in the policy year (the retention assignment variables), retention status, spring and summer portfolio outcomes in the policy year, as well as spring assessment scores in third through eighth grade. In addition, the available data include background information on students and schools. The student-level measures will include factors such as gender, race/ethnicity, free- and reduced-lunch status, English Language learner status, and attendance (in non-outcome years). School-level variables, which are available for the entire school and for each grade individually, include enrollment, grades served, aggregates of each of the individual student measures, and a series of school-level teacher characteristics.

We examine impacts using the models described above for two types of measures of behavioral problems. The first is truancy, which is reflective of student engagement in school (Rohrman, 1993) and is calculated from daily attendance data. The second is based on disciplinary event data. We analyze whether a student was suspended, the number of suspensions, and the reason and severity of the reason for the disciplinary action.

Findings / Results:

Data analyses are ongoing. We have fit preliminary models for all behavioral outcomes and are in the process of refining the model selection to choose the most appropriate specification for these data. Sensitivity checks will then be implemented. Final results will be available well in advance of the March conference dates.

Conclusions:

Because we only have preliminary results, we cannot draw firm conclusions at this stage. However, by the time of the SREE conference in March 2015, we expect to have solid results that will shed light on whether or not grade retention leads to the behavioral problems that critics of the policy contend that it does.

Appendices

Not included in page count.

Appendix A. References

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary school grades* (2nd ed.) Cambridge, UK: Cambridge University Press.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). *Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multi-level analysis*. *Educational Evaluation and Policy Analysis, 31*(4), 480–499.

Byrd, R. S., Weitzman, M., & Auinger, P. (1997). Increased behavior problems associated with delayed school entry and delayed school progress. *Pediatrics*, *100*(4), 654–661.

Cunha, F., & Heckman, J. J. (2006). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, *43*(4), 738–782.

Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy, 2*(4), 319–340.

Hahn, J., Todd, P., & Van Der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201–209.

Holmes, C. T. (1989). Grade Level Retention Effects: Meta-Analysis of Research Studies. In Lorrie A. Shepard and Mary Lee Smith (Eds.), *Flunking Grades: Research and Policies on Retention* (pp. 16–33). London: The Falmer Press.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(3), 205–224.

Hughes, J. N., Chen, Q., Thoemmes, F., & Kwok, O. (2010). An investigation of the relationship between retention in first grade and performance on high stakes tests in 3rd grade. *Educational Evaluation and Policy Analysis, 32*(2), 166–182.

Imbens, G., & Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics, 142*(2), 615–635.

Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research, 45*(4), 613–635.

Jacob, B. A., & Lefgren, L. (2004). Remedial Education and Student Achievement: A Regression Discontinuity Analysis. *The Review of Economics and Statistics*, *86*(1): 226-244.

Jimerson, S. R. (2001). Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century. *School Psychology Review, 30*(3), 420–438.

Lorence, J., Dworkin, A. G., Toenjes, L. A., & Hill, A.N. (2002). Grade retention and social promotion in Texas 1994-99: Academic achievement among elementary school students. *Brookings Papers on Education Policy, 5,* 13-67.

Mariano, L., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis*, *35*(1): 96–117.

New York City Department of Education. (2009, August 10). Mayor unveils plan to end social promotion in all remaining grades from three to eight [news release]. Retrieved from

http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2009-2010/20090810_social_promotion.htm

Rohrman, D. (1993, January). Combating truancy in our schools—A community effort.*National Association of Secondary School Principals Bulletin, 76*(549), 40-45.

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, *31*(4), 729–759.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis, 27*(4), 309–340.

Schwerdt, G., & West, M. R. (2013). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida (PEPG No. 12–09). Cambridge, MA: Harvard Kennedy School.

Segal, C. (2013). Misbehavior, education, and labor market outcomes. *Journal of the European Economic Association*, *11*(4), 743–779.

Appendix B. Tables and Figures

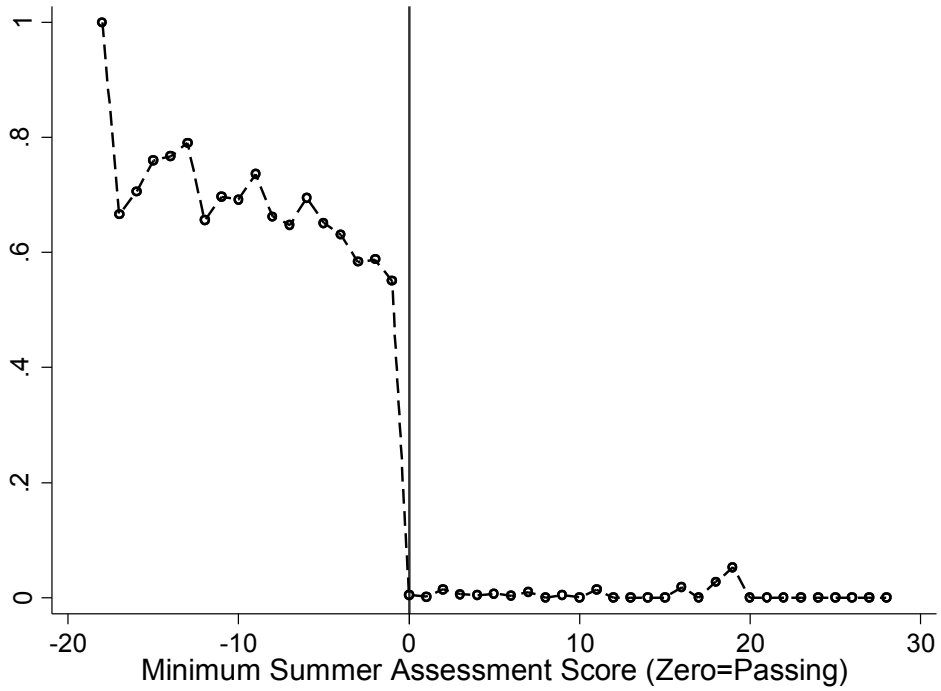
Not included in page count.

Table 1. Grade Completed in the 2012–2013 School Year for Each Cohort Subject to the NYCDOE Student Promotion Policy

Grade	Sub-Cohort	Cohort								
		2003–2004	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009	2009–2010	2010–2011	2011–2012
3	P	12	11	10	9	8	7	6	5	4
	R	11	10	9	8	7	6	5	4	3
4	P							7	6	5
	R							6	5	4
5	P		G	12	11	10	9	8	7	6
	R		12	11	10	9	8	7	6	5
6	P							9	8	7
	R							8	7	6
7	P				G	12	11	10	9	8
	R				12	11	10	9	8	7
8	P						12	11	10	9
	R						11	10	9	8

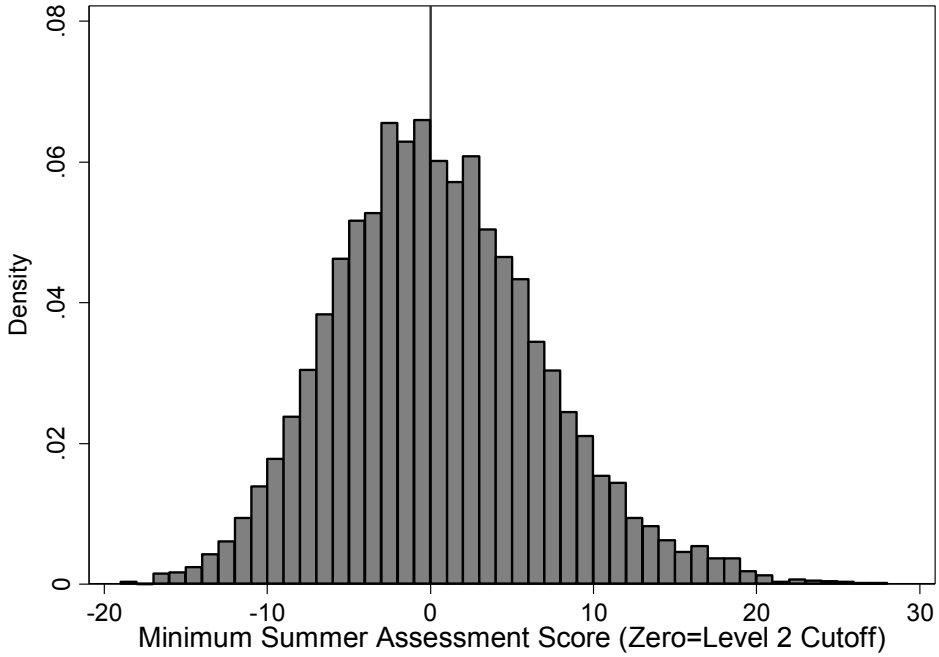
Notes: R=retained students; P=promoted students; G=high school completed.

Figure 1: Fraction of Students Retained in Grade by Summer Assessment Score



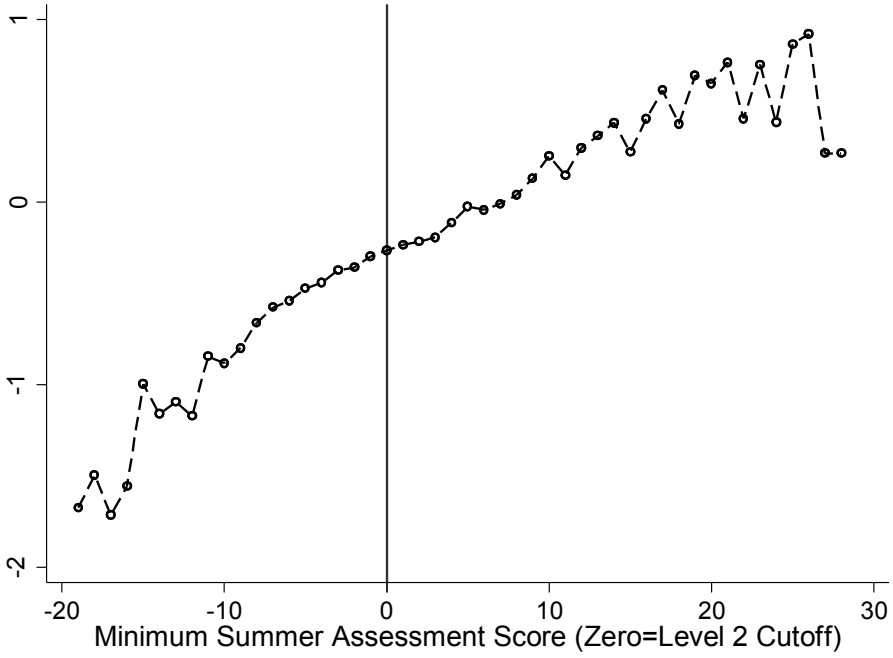
Note: The sample includes 10,222 students in the 2004-05 and 2005-06 5th grade cohorts who took the summer assessment. Results for other cohorts and grades subject to the promotion policy are similar. The vertical axis represents the number of students retained in 5th grade in the following year. The horizontal axis is defined as the minimum of the math and ELA summer assessment scores. For each subject, scale scores were converted to rank-ordered scores and centered to be equal to zero at the Level 2 cutoff. Thus all students were Level 1 on at least 1 subject if and only if they were below zero on the horizontal axis.

Figure 2: Histogram of the Summer Assessment Score



Note: Sample and minimum score variable same as in Figure 1.

Figure 3: Spring Assessment Math z-Score by Summer Assessment Score



Note: Sample and minimum score variable same as in Figure 1.