

Abstract Title Page

Title:

Examining Multidimensional Middle Grade Outcomes after Early Elementary School Grade Retention

Authors and Affiliations:

Sophia Hwang, MEd
Department of Applied Psychology, New York University
sophia.hwang@nyu.edu

Elise Cappella, PhD
Department of Applied Psychology, New York University
elise.cappella@nyu.edu

Kate Schwartz, MPP
Department of Applied Psychology, New York University
kjs440@nyu.edu

Abstract Body.

Background / Context:

Grade retention, the practice of requiring a student who has underperformed academically to remain at that same grade level the following year (Jackson, 1975), is a controversial and common educational practice. Retention has been characterized as a detrimental intervention in early empirical studies (see meta-analyses: Holmes, 1989; Jimerson, 2001); however, recent studies emphasizing study design and the construction of appropriate comparison groups (e.g., Hong & Raudenbush, 2005; 2006; Hong & Yu, 2007; 2008; Im, Hughes, Kwok, Puckett, & Cerda, 2013) have found mixed post-retention effects.

Two million children are retained annually in the United States (Eide & Goldhaber, 2005) and socio-demographically at-risk students facing academic challenges are disproportionately affected. Males, non-Hispanic Blacks, and students from lower income families are more likely to be retained (Warren, Hoffman, & Andrew, 2014; Xia & Kirby, 2009). Retention rates have fluctuated in the past century (Bali, Anagnostopoulous & Roberts, 2005), but retention became more regularly practiced in the last 25 years, coinciding with policy efforts to end social promotion (Clinton, 1998; Hursh, 2007). After the passage of No Child Left Behind (NCLB, 2002), administrators may have been compelled to retain low performing students so schools could meet adequate yearly progress as measured by high stakes tests (Hursh, 2007). From a pedagogical standpoint, retention increases classroom heterogeneity in age and decreases heterogeneity in achievement, theoretically easing the instructional demands on teachers (Shepard & Smith, 1988). Thus, with limited access to alternative remediation strategies, policy directives may have led to reliance on retention as an intervention for underperforming students (Lorence, 2009).

There is a policy and practice need to understand post-kindergarten retention because whereas short-term outcomes after kindergarten retention are well-studied (e.g., Hong & Yu 2006; Huang et al., 2014), little attention has been paid the effects of non-kindergarten grade retention on outcomes beyond childhood into early adolescence (e.g., 6th - 8th grade; for an exception, see Im, Hughes, Kwok, Puckett, & Cerda, 2013). Developmentally, this time period is particularly important because student attendance, test scores, grades, and behavior in the middle grades are critical for later academic achievement (Balfanz, 2009) and predict high school performance and graduation (Balfanz, 2007). Generally, retained students are more likely to struggle after they leave elementary school (Alexander et al., 2003) and with a lack of extant research examining longer term outcomes, there is a need for statistically rigorous, longitudinal research to better understand the academic and psychosocial experience of first and second grade retainees in comparison to their promoted peers beyond the elementary school years.

Purpose / Objective / Research Question / Focus of Study:

Recently, researchers have begun to employ rigorous statistical methods and developmentally informed theories to evaluate outcomes for students retained in non-kindergarten early elementary school (see Gleason, Kwok, & Hughes, 2007; Wu, West, & Hughes, 2008). However, the majority of this research focuses on academic outcomes. Gaps remain regarding retention's effects on psychosocial outcomes important to later school and life outcomes (Heckman, 2000). Additionally, some psychosocial and academic outcomes and changes may be evident in the short-term, but others may have a delayed presentation. These "sleeper effects" (Alexander et al., 2003; Gleason et al 2007; Pagani et al., 2001) are less well-understood, creating a need for new research focusing on longer-term outcomes.

In response, the current study uses causal methods and a national dataset to investigate both academic achievement and psychosocial outcomes in early adolescence after early elementary grade retention. Using propensity score methodological approaches (Rosenbaum, 2002) and informed by developmental systems theory (Ford & Lerner, 1992), this study estimates the effect of early (first and second) grade retention on academic achievement and psychosocial outcomes in eighth grade: six or seven years later. The goal is to determine whether retention is an effective intervention for students struggling to meet academic standards early in their elementary schooling, and to inform the education policy and practice debate regarding the costs and benefits associated with grade retention.

Setting:

Data for this study are from the Early Childhood Longitudinal Study, Kindergarten Class 1998-1999 (ECLS-K; U.S. Department of Education, 2009), a nationally representative sample following a longitudinal cohort of 21,260 kindergarteners until the eighth grade. Data were collected at seven waves: fall and spring of kindergarten (1998-99), fall and spring of first grade (1999-2000), spring of third grade (2002), spring of fifth grade (2004), and spring of eighth grade (2007). During the fall of first grade, data were only collected from a subsample of participants and this wave will not be included in this analysis. The dataset includes multiple sources of data including administrator surveys, teacher assessments, direct child assessments, child questionnaires, and parent surveys.

Population / Participants / Subjects:

The sample is comprised of 5,586 students who (1) participated in the six data collection waves of interest; (2) attended public school through all data collection waves; (3) were first time kindergarteners at the study's start; (4) were enrolled in first grade in spring 2000; and (5) have valid grade level data in the spring of 2002 (i.e., students who were in ungraded classrooms, promoted ahead of time, retained in kindergarten in spring 1999, or whose grade level data were missing or not ascertained were excluded.)

Within this sample, 295 students who were retained once in either the first or second grade compose the treatment group. These students were identified through teacher report because they were in the second grade in spring 2002, when the majority of students in the sample are in the third grade. All remaining students in the sample who were reported as being in the third grade in spring of 2002 create the comparison group (n = 5,291). Table 1 presents the pre-treatment demographic characteristics for retained and promoted students. In summary, the retained group consists of a higher proportion of males (58% versus 48%) and lower proportion of white students (38% versus 61%) as compared to the promoted group. Additionally, 22% of retained students were classified as having a disability (a proxy for special education), while 13% of promoted students had this status.

Intervention / Program / Practice:

Grade retention supporters claim it is an effective remedial intervention that provides students "the gift of time" (Smith & Shepard, 1988; Tomchin & Impara, 1992), and some studies demonstrate student improvement in reading and math test scores (Roderick & Nagaoka, 2005) and social-emotional outcomes (Hong & Yu, 2008) after retention. Yet, scholars who oppose retention and advocate for social promotion (i.e., moving students to the next grade even if current performance standards are not met), argue that students would fare better academically if

they were not retained (Allen et al., 2009; Reschly & Christenson, 2013). Detrimental academic consequences (Jimerson & Ferguson, 2007; Jimerson, et al., 2002) and behavioral health outcomes provide empirical support for the negative effects of grade retention.

Research Design:

The ECLS-K followed a nationally representative sample of kindergarteners from the 1998-1999 school year through the 2006-2007 school year. The current study utilizes a quasi-experimental, longitudinal approach and propensity score methodology to estimate the effect of early grade retention on academic and psychosocial outcomes in the middle grades.

Data Collection and Analysis:

The ECLS-K collected data from multiple reporters to assess a range of child outcomes. Data utilized in the analysis are derived from the fall and spring of kindergarten (wave 1 and 2), spring of 1st grade (wave 4), and spring of 8th grade (wave 7). For students who were not retained (i.e. engaged in continuous progression) this translates to a data collection timeline occurring in kindergarten, first, third, and eighth grade; for retained students, they would be at a lower grade level. Covariates are drawn from data collected *during or before* spring 2000 (i.e., before retention occurs), and outcomes are measured in spring 2007. We use data from spring 2002 to identify treated versus comparison group students.

Academic outcomes. *Math and reading achievement* are measured by the direct assessment of students' math and reading skills on a 50 to 70 item reading test that measures latent ability in the subject. This study uses the IRT-derived theta scores, which are comparable across different time points and thus allow for the examination of growth over time (Najarian, Pollack, & Sorongon, 2009). Theta reliability estimates for reading IRT score is .87, and .92 for math in spring 2007 (Najarian et al., 2009). Student self-reported *reading competence* and *math competence* from the Self-Description Questionnaire (alpha = .76 and .89 respectively; Najarian et al., 2009), the indirect measure of teacher-report of reading competence from the Academic Rating Scale, and teacher rated oral and written skills are examined.

Psychosocial outcomes. *Social self-concept* is a continuous composite measure comprised of five items related to peer acceptance and rated on a five-point scale (alpha = .89). In spring 2007, students rated eight items along a four-point scale about feeling lonely, frustrated, and worrying about school, which comprises the measure of *internalizing problem behaviors* (alpha = .75). *Self-esteem* (alpha = .79) is comprised of seven items derived from the Rosenberg Self-Esteem Scale (RSE, Rosenberg 1965), and *locus of control* (alpha = .68) is comprised of six items. For both self-esteem and locus of control, items were rated on a four-point scale (i.e., strongly agree to strongly disagree) and analyses utilized the scale score, which is the average of the standardized items with mean of zero and standard deviation of one.

Covariates. The selection of covariates is informed by previous empirical work to understand connections to both the likelihood of being retained, as well as outcomes post-retention. All covariates are from fall 1998, spring 1999, or spring 2000, ensuring collection prior to the treatment of grade retention. Selected covariates (Table 2) are organized by priority for achieving balance in the propensity score model, due to the strength of theorized connection to the treatment and/or outcome variables. The covariates include administrative (child and school level), demographic, teacher-report, and parent-report variables, in addition to child assessments. Pre-treatment indicators of dependent variables collected in fall 1998, spring 1999, and/or spring 2000 are also included as covariates. Informed by prior studies, 43 confounding

covariates are included in the propensity score model to address the ignorability assumption; of these, higher priority covariates with less than ideal balance will be included in the analytic model.

Data analysis. Prior to running analyses, we will conduct a single imputation using chained equation to address the issue of missing data. Then we use propensity score matching to construct a logical counterfactual group, and compare retained students to other students who would have been retained, based on similar observed characteristics, but were promoted. The utilization of propensity scores ensures that students who have been retained are only compared to similar individuals who have been promoted as measured by “causally related covariates” (Hill & Reiter, 2006), rather than *all* students who have been promoted. This allows us to estimate the *effect* of grade retention and disentangle it from other confounds because the only meaningful observed difference between the two groups is the treatment. By accounting for characteristics measured before retention that may act as potential confounds, the propensity score method adjusts for potential bias associated with the likelihood of a student experiencing grade retention versus being promoted on time.

Consistent with prior literature (e.g., Hong & Raudenbush 2005, 2006; Hong & Yu 2007, 2008), each student in the analytic sample will receive a propensity score fit using a logistic regression, which summarizes the likelihood of being retained based on all of the pre-retention covariates (Rosenbaum, 2002; Rosenbaum & Rubin, 1983). We use two different propensity score methodological approaches to estimate the average treatment effect on the treated: (1) nearest neighbor (one to one) matching without replacement and (2) caliper matching with replacement. The goal is to select the ideal control participants that serve as the best comparison for the students who have been retained. Retained students not in the area of common support are excluded, as they do not have comparable counterfactuals. We will employ two different matching approaches to address the issue of bias and determine if our findings are consistent and robust. Various propensity score models will be fit for each method to obtain balance between the retained sample and the comparison group of promoted students on as many covariates as possibility, but prioritizing the covariates on the left side of Table 2. To determine the treatment effect for each method and outcome, we will run a regression-adjusted matched estimate using probability weights and covariate adjustment (i.e., inclusion of the priority covariates related to the likelihood of being retained and/or outcomes post-retention, and variables that do not demonstrate sound balance).

Findings / Results:

Analyses are presently being conducted and formal results are forthcoming. As a preliminary step, descriptive results summarizing the outcomes of interest for the treatment group of students who have been retained are presented in Table 3.

Conclusions:

Conclusions and recommendations will be developed after the results and findings have been finalized. However, if there is an absence of compelling evidence to support the extremely costly and time intensive intervention of grade retention (Allen et al., 2009), there is an opportunity for this research to contribute policy-relevant knowledge. Going beyond academic outcomes, this study aims to highlight the long-term, cumulative impacts of early elementary grade retention and the relationship between academic and psychosocial outcomes during this developmentally sensitive outcome period.

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*. 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, multilevel analysis. *Educational Evaluation and Policy Analysis*, 31(4), 480-499.
- Balfanz, R. (2009). *Putting middle grades students on the graduation path: A policy and practice brief*. Westerville, OH: National Middle School Association.
- Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235.
- Bali, V. A., Anagnostopoulos, D., & Roberts, R. (2005). Toward a political explanation of grade retention. *Educational Evaluation and Policy Analysis*, 27(2), 133-155.
- Cannon, J. S., & Lipscomb, S. (2011). *Early grade retention and student success: Evidence from Los Angeles*. Public Policy Institute of California
- Clinton, W. J. (1998). *State of the union address (January 1, 1998)*. Washington, DC: U.S. Government Printing Office.
- Eide, E. R., & Goldhaber, D. D. (2005). Grade retention: What are the costs and benefits?. *Journal of Education Finance*, 195-214.
- Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory: An integrative approach*. Newbury Park, CA: Sage.
- Gleason, K. A., Kwok, O. M., & Hughes, J. N. (2007). The short-term effect of grade retention on peer relations and academic performance of at-risk first graders. *The Elementary School Journal*, 107(4), 327.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54(1), 3-56.
- Hill, J. & Reiter, J.P. (2006). Interval estimation for treatment effects using propensity score matching. *Statistics in Medicine*, 25, 2230-2256.
- Hill, J., Weiss, C., & Zhai, F. (2011). Challenges with propensity score strategies in a high-dimensional setting and a potential alternative. *Multivariate Behavioral Research*, 46(3), 477-513.
- 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.
- 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), 205-24.
- Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multi-level observational data. *Journal of the American Statistical Association*, 101(475), 901-10.
- Hong, G., & Yu, B. (2007). Early-grade retention and Children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis*, 29(4), 239-61. doi:10.3102/0162373707309073

- Hong, G., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44*(2), 407-421. doi:10.1037/0012-1649.44.2.407
- Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American Educational Research Journal, 44*(3), 493-518.
- Im, M. H., Hughes, J. N., Kwok, O. M., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology, 51*(3), 349-65.
- Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research, 6*13-35.
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30*(3), 420-37.
- Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22*(3), 314-339. doi:10.1037/1045-3830.22.3.314
- Jimerson, S. R., Ferguson, P., Whipple, A. D., Anderson, G. E., & Dalton, M. J. (2002). Exploring the association between grade retention and dropout: A longitudinal study examining socio-emotional, behavioral, and achievement characteristics of retained students. *California School Psychologist, 7*, 51-62.
- Karweit, N. L. (1999). Grade retention: Prevalence, timing, and effects (CRESPAR Report No. 33). Baltimore: Johns Hopkins University, Center for Social Organization of Schools.
- Lorence, J., 2009. Grade retention redux: a dissenting perspective. In: Saha, L.J., Dworkin, A.G. (Eds.), *International Handbook of Research on Teachers and Teaching*, vol. 2. Springer, New York, pp. 1153–75.
- Najarian, M. Pollack, J.M., and Sorongon, A.G. (2009). Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Eighth Grade (NCES 2009–002). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- No Child Left Behind, N. C. L. (2002). Act of 2001, P. L. 107-110, § 115. *Stat, 1425*, 107-110.
- Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology, 13*, 297–315.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55.
- Shepard, L. A., & Smith, M. L. (1988). Escalating academic demand in kindergarten: Counterproductive policies. *The Elementary School Journal, 89*(2), 135-145.
- Warren, J. R., Hoffman, E., & Andrew, M. (2014). Patterns and trends in grade retention rates in the United States, 1995–2010. *Educational Researcher, 43*(9), 433-443. doi: 10.3102/0013189X14563599
- Wu, W., West, S. G., & Hughes, J. N. (2008). Effect of retention in first grade on children's achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology, 100*(4), 727-40. doi: <http://dx.doi.org/10.1037/a0013098>

- U.S. Department of Education, National Center for Education Statistics. (2009). Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Kindergarten through Eighth Grade Full Sample Public-Use Data and Documentation (DVD) (NCES 2009–005). Washington, DC.
- Xia, N., & Kirby, S. (2009). Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes. (Technical Report TR-678-NYCDOE). Retrieved from http://www.rand.org/pubs/technical_reports/2009/RAND_TR678.pdf.

Appendix B. Tables and Figures

Table 1
Pre-treatment Demographic Characteristics for Retained and Promoted Students
(N = 7,263)

	Retained (n = 295)		Promoted (n = 5,291)	
	n	%	n	%
Gender				
Male	170	57.63	2,551	48.21
Female	125	42.37	2,740	51.79
Race				
White	113	38.31	3,202	60.52
Black	81	27.46	540	10.21
Hispanic	72	24.41	939	17.75
Asian	8	2.71	310	5.86
Native Hawaiian or American Indian	14	4.75	168	3.18
Multiracial	7	2.37	128	2.42
Student has a disability at start of Kindergarten	61	22.34	641	12.79
Region				
Northeast	50	16.95	1,003	18.96
Midwest	53	17.97	1,434	27.10
South	140	47.46	1,666	31.49
West	50	16.95	1,185	22.40
Urbanicity				
Urban	122	41.36	1,710	32.32
Suburban	80	27.12	2,098	39.65
Rural	91	30.85	1,442	27.25
	Mean	SD	Mean	SD
Spring 2000 Reading IRT Test Score	-0.47	0.42	0.19	0.38
Spring 2000 Math IRT Test Score	-0.47	0.42	0.12	0.37

Table 2
Selected Covariates for Propensity Score Development (Organized by Priority)

Higher priority	Lower priority
Gender	Region
Race	Urbanicity
Parent marital status	Father employment
Student has disability at start of Kindergarten	Mother employment
SES	Family in poverty
Age in of child in months (spring of 1st grade)	Number of people in household
1st grade reading test score	Mother education
1st grade math test score	Student has disability in spring of first grade
1st grade general knowledge test score	Teacher report Interpersonal Skills
Teacher report externalizing problems	Teacher report Self-control
Teacher report approaches to learning	Teacher report Internalizing problems
Teacher report math competence	School average daily attendance
Teacher report literacy competence	% of students in school who are racially "other"
Number of students enrolled in first grade	% of limited English proficient students in school
% in school at or above grade level in reading	% of limited English proficient students in first grade
% in school at or above grade level in math	% eligible in school for free lunch
% of Hispanic students in school	If student can be retained in any grade ^a
% of Black students in school	If student can be retained by parent request ^a
School has formal retention policy	If student can be retained because of below grade level performance ^a
Number of retained first graders in school	If student can be retained in kindergarten ^a
	If student can be retained in any grade more than once ^a
	If student with disability can be retained ^a
	If student can be retained without parent approval ^a

^a Specific school retention policies.

Table 3
Spring 2007 Descriptive Outcomes for Students who have been Retained (n = 295)

	Mean	SD
Academic		
Reading IRT test score	0.90	0.27
Math IRT test score	0.97	0.36
Student self-reported reading competence ^a	2.34	0.73
Student self-reported math competence ^a	2.46	0.92
Teacher report of reading competence: oral ^b	2.54	0.82
Teacher report of reading competence: written ^b	2.11	0.82
Teacher report of math competence ^b	2.33	0.82
Psychosocial		
Social self-concept ^b	3.66	0.98
Student self-reported internalizing behaviors ^a	2.17	0.61
Locus of control ^c	-0.36	0.67
Self-esteem ^c	-0.23	0.70

^a Item response ranges from 1 – 4.

^b Item response ranges from 1 – 5.

^c Variable is a composite scale score with mean of zero and standard deviation of one.