



Achievement at Whose Expense? A Literature Review of Test-Based Grade Retention Policies in U.S. Schools

Andrew P. Huddleston
Abilene Christian University
USA

Citation: Huddleston, A. P. (2014). Achievement at whose expense? A literature review of test-based grade retention policies in U.S. school. *Education Policy Analysis Archives*, 22 (18).
<http://dx.doi.org/10.14507/epaa.v22n18.2014>

Abstract: The author uses Maxwell's method of literature reviews for educational research to focus on literature relevant to test-based grade retention policies to make the following argument: although some studies have documented average gains in academic achievement through test-based grade retention, there is increasing evidence that these gains have occurred by limiting the educational opportunities for the most vulnerable of students. The author begins by briefly synthesizing research on high-stakes testing policies and teacher-based retention in general and then examines studies that have evaluated specific test-based retention policies in Chicago, Florida, New York City, Georgia, Texas, Wisconsin, and Louisiana. Drawing on Bourdieu and Passeron's concept of reproduction in education, the author shows how testing policies have contributed to class selection and exclusion in U.S. schools. Short-term gains produced by test-based retention policies fade over time with students again falling behind but with a larger likelihood of dropping out of school. These unintended consequences are most prevalent among ethnic minority and impoverished students. The author concludes by providing alternatives for ending social promotion that do not include grade retention as well as suggestions for further researching the role such policies play in perpetuating class inequities.

Keywords: grade repetition; high-stakes testing; literacy; social promotion; test-based grade retention

¿Resultados a expensas de quién? Una revisión de la literatura de las políticas de retención de grado basado en exámenes en EE.UU.

Resumen: El autor utiliza el método de revisión de la literatura de Maxwell para la investigación educativa para centrarse en la literatura relevante sobre políticas de retención de grado basado en exámenes de alto riesgo para argumentar que: si bien algunos estudios han documentado aumentos promedio en el rendimiento académico a través de la repetición de curso, hay cada vez más pruebas de que estas ganancias se han producido mediante la limitación de las oportunidades educativas para los estudiantes más vulnerables. El autor comienza por sintetizar brevemente la investigación sobre exámenes de alto riesgo y retención basada en docentes en general y luego examina los estudios que han evaluado las políticas de retención en base a exámenes de alto riesgo específicos en Chicago, Florida, Nueva York, Georgia, Texas, Wisconsin y Louisiana. Sobre la base del concepto de reproducción en la educación de Bourdieu y Passeron, el autor muestra cómo las políticas de exámenes de alto riesgo han contribuido a la selección de clases y la exclusión en las escuelas estadounidenses. Los beneficios a corto plazo producidos por las políticas de retención a base de pruebas se desvanecen con el tiempo con los estudiantes siendo retrasados sistemáticamente, pero con una mayor probabilidad de abandonar la escuela. Estas consecuencias no intencionales son más frecuentes entre estudiantes de minorías étnicas y pobres. El autor llega a la conclusión, que proporcionando alternativas para poner fin a la promoción social que no incluyen la repetición de curso, así como sugerencias para investigar aún más el papel que tales políticas desempeñan en la perpetuación de las desigualdades de clase.

Palabras clave: repetición de curso; exámenes de alto riesgo; alfabetización; promoción social.

Resultados À custa de quem? A revisão da literatura sobre políticas de retenção de grau com base em testes nos EUA

Resumo: O autor usa o método de revisão de literatura de Maxwell para a pesquisa educacional para se concentrar na literatura relevante sobre as políticas de retenção de grau com base em testes de alto risco para argumentar que: embora alguns estudos têm documentado aumentos médios em desempenho acadêmico através da repetição, há cada vez mais evidências de que esses ganhos têm ocorrido, limitando as oportunidades educacionais para os alunos mais vulneráveis. O autor começa por resumir brevemente investigação sobre testes de alto risco e retenção com base nos exames de professores em geral e , em seguida, examina os estudos que avaliam as políticas de retenção baseadas em testes específicos de alto risco em Chicago, Flórida, Nova York, Georgia, Texas , Wisconsin e Louisiana. Baseado no conceito de reprodução na educação de Bourdieu e Passeron, o autor mostra como as políticas de testes de alto risco têm contribuído para a seleção de classes e exclusão nas escolas americanas. Os benefícios de curto prazo produzidos por políticas de retenção com base em testes desaparecer ao longo do tempo com os alunos sendo adiada sistematicamente , mas com uma maior probabilidade de abandonar a escola. Essas conseqüências não intencionais são mais comuns entre os estudantes de minorias pobres e étnicas. O autor conclui que o fornecimento de alternativas para acabar com a promoção social não incluem repetição e sugestões para investigar o papel dessas políticas jogar na perpetuação das desigualdades de classe.

Palavras-chave: repetição , teste de alto risco , de alfabetização , de promoção social.

Achievement at Whose Expense? A Literature Review of Test-Based Grade Retention Policies in U.S. Schools

Although test-based grade retention in U.S. schools occurred as early as the 1870s with the use of written examinations (White, 1886, 1888), its current use with standardized tests for large-scale accountability purposes is rather new, dating back to the minimum-competency testing of the late 1970s and early 1980s (Koretz, 2008). Decreasing SAT scores (Wirtz, 1977) and a perceived softening of grading and educational standards nationwide (Berliner & Biddle, 1995) fueled a growing concern that public schools were not making the grade. These fears culminated in the Reagan administration's publication of *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983) which called for additional testing designed to curb social promotion and increase student achievement.

Cities such as New York and Chicago (Millicent, 1997) and states like Florida (Morris, 2001) and Georgia (Orfield & Ashkinaze, 1991) soon adopted test-based retention policies. Many of these policies were cancelled by 1990 because of their high costs with few apparent gains (House, 1998, 2004). Despite these initial failures, by the late 1990s and early 2000s, the popularity of test-based retention policies was again increasing.

President Bill Clinton was largely responsible for regenerating interest in ending social promotion during the late 1990s. In 1996, at the National Education Summit in Palisades, New York, Clinton urged governors to administer exams students must pass to be promoted (Cannon, 1996). Chicago had again implemented a test-based retention policy in 1996, under the direction of Mayor Richard M. Daley, and Clinton showcased the Chicago policy as a model for what other cities and states could accomplish (Russo, 2005). In 1999, he again challenged the nation's governors: "Look dead in the eye some child who has been held back [and say], 'We'll be hurting you worse if we tell you you're learning something when you're not'" (as cited in Hurwitz & Hurwitz, 2000, p. 21). Clinton then issued a report for educators, state, and local leaders (U.S. Department of Education, 1999) that included guidelines for ending social promotion. Since the late 1990s, several states (e.g., Florida, Georgia, Louisiana, North Carolina, Texas) and cities (e.g., Chicago, New York City) have enacted promotional gates by requiring students to pass a standardized test to be promoted to the next grade (Marsh, Gershwin, Kirby, & Xia, 2009). According to the Educational Commission of the States, as of 2005, 12 states had established test-based retention policies (Educational Commission of the States, 2005). Since then, Arkansas, Oklahoma, and Tennessee have passed similar legislation (Educational Commission of the States, 2011).

Purpose

Test-based grade retention policies have elicited great debate, both in education circles and among the general public. Proponents of retention (e.g., Owen & Ranick, 1977; Winters & Greene, 2006) have argued that retention is necessary to ensure that students who are behind master the skills needed to succeed in the next grade level. Opponents (e.g., Shepard & Smith, 1989), however, have claimed that retention unfairly targets the most vulnerable students, rarely results in academic improvement, and increases the likelihood that students will drop out of school. So what do research findings suggest about the impact of test-based retention policies, especially on low-income and ethnic minority students?

Some researchers (e.g., Boote & Beile, 2005) claim that the key attributes of quality literature reviews are their thoroughness and comprehensiveness. Maxwell (2006), however, argued that rather than comprehensively summarizing or synthesizing research on a specific topic, effective literature reviews for educational research should instead focus on studies that are most relevant to a specific argument made evident in the literature, thus demonstrating new scholarly insights and areas needing further research. Drawing on Maxwell (2006), rather than providing a comprehensive review of the research on high-stakes testing and grade retention, I focus on studies relevant to test-based retention to make the following argument: although some studies have documented average gains in academic achievement through test-based grade retention, there is increasing evidence that these gains have occurred by limiting educational opportunities for the most vulnerable students.

Other researchers have made similar arguments when reviewing the literature on retention (e.g., Shepard & Smith, 1989). However, these reviews have only addressed teacher and not test-based retention. Researchers of test-based retention policies (e.g., Allensworth & Nagaoka, 2010; Greene & Winters, 2007) have argued that although similar, the two differ in terms of how the retention decision is made and thus warrant separate study.

I begin the review by introducing Bourdieu and Passeron's (1987/1990) theoretical concept of reproduction in education. Reproduction suggests testing policies produce social inequality, bolstering certain types of students while hindering others, and provides a useful lens for analyzing the effects of test-based retention.

Because test-based retention policies are a combination of high-stakes testing and grade retention, I briefly review research on high-stakes testing policies in general. This includes research conducted on testing policies under No Child Left Behind (NCLB) as well as research on minimum-competency and state testing policies prior to NCLB that have important implications for test-based retention policies. Although NCLB does not require test-based grade retention, some researchers have argued that the assessments resulting from NCLB provide a mechanism for using test results in retention decisions, thus indirectly proliferating test-based retention policies (Penfield, 2010). No studies have specifically examined the effects of NCLB on test-based retention; however, research conducted prior to NCLB does provide some interesting findings. For example, Heilig and Darling-Hammond (2008) found that some Texas schools preemptively retained ninth graders to provide an extra year of instruction prior to the tenth grade high school graduation exams. Others preemptively promoted students from ninth to eleventh grade to avoid tenth grade testing all together.

Next, because teacher and test-based retention are closely related and have produced similar findings, I provide an overview of research on teacher-based grade retention. I briefly discuss the findings of major meta-analyses and literature reviews, highlighting new contributions made by each study. Several literature reviews have been written synthesizing the research on high-stakes testing and teacher-based grade retention. However, no comprehensive reviews have been written on retention policies dictated by standardized test scores. To address this need I conclude by discussing in-depth the research on test-based grade retention policies, noting the similarities between its findings and those on teacher-based retention.

Methods and Definitions

A variety of search methods were used to locate the sources for this review. I first searched for relevant books, articles, and research reports by using numerous databases such as Educational Resources Information Center (ERIC), PsycINFO, Web of Science, Google Scholar, and the GIL Universal Catalog of the University System of Georgia. I used search terms such as "standardized

test*,” “high-stakes test*,” “grade repetition” and “social promotion.” I then reviewed the reference lists of each of those sources.

Throughout this review, I use terms such as social promotion, test-based retention, promotional gates, standardized testing, and high-stakes testing. For clarity, I provide the following descriptions. The U.S. Department of Education (1999) has defined social promotion as “allowing students who have failed to meet performance standards and academic requirements to pass on to the next grade with their peers instead of completing or satisfying requirements” (p. 5).

Numerous states and larger cities (e.g., Texas, Georgia, New York City, Chicago) have developed test-based grade retention policies in an effort to eliminate social promotion in schools (Marsh et al., 2009). These policies require that test scores be used, at least in part, to determine which students should be promoted and which should be retained. Rather than affecting all grades, these policies frequently contain promotional gates, which are specific grades in which test-based retention policies apply. For example, in Georgia, the test-based retention policy applies in grades 3, 5, and 8 (Georgia State Board of Education, 2001).

Most often, the tests involved in these policies are standardized tests, usually criterion-referenced, that are administered using standardized procedures for administration, completion, and scoring (Haney, 1984). What makes these tests “high-stakes” is that their results, in this case promotion or retention, are used to make important decisions that immediately and directly affect students, teachers, and schools (Madaus, 1988, p. 87).

Researchers of test-based retention policies use a variety of terms to describe students who are at-risk of retention as well as the social barriers they face (e.g., low-performing, low-achieving, low-scoring, low socio-economic (SES), impoverished, low-income, ethnic minority, marginalized, class inequities). When available, I provide the specific definitions researchers use for these terms. Drawing on Bourdieu and Passeron (1970/1990), as I discuss below, I also use the term vulnerable to describe students most at risk of the negative consequences of test-based retention. According to Bourdieu and Passeron (1970/1990) individuals are most vulnerable when they lack the capital to successfully compete for the resources in a given field.

Bourdieu and Passeron on Education and Testing

Teachers have often recognized that students with certain backgrounds tend to flourish in school while others do not. Researchers have frequently attributed such achievement gaps to opportunity gaps or unequal childhoods that occur among class lines in the U.S. (Lareau, 2003). French sociologists Bourdieu and Passeron (1970/1990) researched economic, social, and cultural class domination and, in so doing, developed a theory of reproduction in education that explains how social class inequities play out in terms of academic achievement in schools.

Bourdieu and Passeron (1970/1990) explained the process of reproduction with the theoretical concepts of field, capital, and habitus. Individuals are socialized by a variety of fields, what Lareau (2003) has described as “institutional arrangements” (p. 275). This socialization greatly influences what individuals recognize as feeling comfortable and natural and thus largely dictates the habitus, how they respond in specific situations. These background experiences also provide individuals with specific amounts and types of capital, which are resources they then use to compete for additional capital within the field.

According to Bourdieu and Passeron (1970/1990), education is a field that consists of its own rules for allocating and accruing resources (e.g., grades, promotions, diplomas) that ultimately determine winners and losers. Schools reward certain types of knowledge, resources, and ways of

speaking more than others. Students whose family backgrounds provide them with these skills do well in school while the rest often do not.

What terms such as “at-risk students,” “ethnic minorities,” “social inequalities,” and “class inequities” all have in common is recognition that within a given field, non-dominant groups lack the capital recognized as valuable within that setting. For Bourdieu and Passeron (1970/1990), it is not that students from these families lack knowledge, skills, and language. In fact, such students often bring a rich variety of resources to the classroom. However, what they do lack are resources that are valued within an educational system (field) that is built upon middle-class principles such as “standard” English.

Educational testing policies, Bourdieu and Passeron (1970/1990) argued, play a key role in making sure the rules of the field remain intact. Tests provide “objective” evidence that those who fail are not cut out for academics and proof of their merit and giftedness to those who pass. Reproduction, as Bourdieu and Passeron (1970/1990) called it, occurs when nondominant groups respond by accepting their failure as a natural or taken-for-granted part of the way life is and retreat from school experiences. In so doing, they unknowingly participate in their own oppression, what Bourdieu and Passeron (1970/1990) called misrecognition, ensuring that inequities will continue.

Bourdieu and Passeron (1970/1990) criticized much of the research on schooling and examinations because they believed it often helped hide the inequities these structures reproduce. Research, Bourdieu and Passeron (1970/1990) argued, must look beyond student outcomes (e.g., achievement gains) to determine what the examinations themselves are concealing. As I review the following research, in addition to discussing outcomes in educational achievement, I specifically examine the adverse impact on non-dominant students such policies are producing.

Research on High-Stakes Testing

Over the past twenty years, a significant amount of research has been conducted on the impact of tests used for accountability purposes. This research has consisted of a mixture of large-scale quantitative studies, surveys, and case studies on testing policies both prior to and under NCLB. This literature is relevant to test-based retention policies in that such policies are themselves a form of high-stakes testing. Moreover, like the research on test-based retention policies, there is evidence the academic gains that have been made occurred by limiting the educational opportunities for the most vulnerable of students.

Beneficial Outcomes of High-Stakes Testing

A few researchers have noted that some high-stakes testing policies have resulted in academic outcomes their supporters believe to be beneficial. For example, Koretz, Stecher, Klein, and McCaffrey (1994) showed that teachers in Vermont spent more time teaching newer curriculum elements such as problem-solving and mathematical representations to prepare their students for their state’s portfolio-based, high-stakes assessment. Stecher (2002) argued that some teachers have found tests useful for identifying students’ strengths and weaknesses and attaining additional resources for students. School districts have revised curriculum and testing programs to match state curricula and provided after-school and Saturday-school tutoring for low-performing students. Hamilton et al. (2007) found that in California, Georgia, and Pennsylvania, under NCLB, schools were aligning curricula with state standards and assessments, using data for decision making, and providing extra support to low-performing students.

Researchers have also suggested that high-stakes tests also play a role in teacher motivation. Hamilton et al. (2007) found that teachers in California, Georgia, and Pennsylvania have been encouraged by high-stakes testing to improve their own practice. Finnigan and Gross (2007) studied

ten elementary schools in Chicago that had been placed on probation for low test scores to determine if accountability sanctions influenced teacher motivation. Drawing on expectancy theory, they defined motivation as a function of a person's valuation of a goal and expectation that the goal could be achieved. Indeed the teachers were motivated to work harder, try new teaching approaches, and participate in professional development. However, Finnigan and Gross (2007) also noted that the teachers appeared to be more motivated to raise test scores because of their professional status and individual goals for students than by external threats. Moreover, the longer schools remained on probation, the more likely teacher morale declined and reversed any gains achieved via increased effort.

Unintended Consequences of High-Stakes Testing

Although some researchers have found positive effects of high-stakes testing, the research documenting unintended and negative effects is widespread. First, there is little evidence suggesting that these policies have actually resulted in academic achievement gains. Hout and Elliott (2011) recently conducted an extensive review of the research on high-stakes testing policies under NCLB. They found that small increases in test scores have occurred, but when similar low-stakes tests were administered, the academic gains were effectively zero for most programs. Hout and Elliott (2011) attributed the gain in test scores to score inflation, in which scores increased due to teaching to the test rather than actual gains in academic achievement.

Other researchers have documented such score inflation as well. Amrein and Berliner (2002, 2003) examined the test scores of 18 states that required that students pass a high-stakes test to graduate from high school during the 1990s. Although the high-stakes test scores increased in the 18 states, no apparent gains were made on the SAT, ACT, Advanced Placement (AP), or National Assessment of Educational Progress (NAEP) exams suggesting score inflation had occurred. Klein, Hamilton, McCaffrey, and Stecher (2000) found evidence of score inflation in Texas when they compared Texas Assessment of Academic Skill (TAAS) scores to NAEP scores from 1994-1998. They found that the gains on NAEP were much smaller than those on TAAS and were not present on the eighth-grade math or reading tests.

In addition to score inflation, researchers have found that negative effects occurred in the form of curriculum reallocation in which teachers provided more instruction towards those content areas and standards tested than those not tested (Au, 2007; Hamilton et al., 2007; Hargrove et al., 2000; Jones et al., 1999; Smith, 1991; Smith & Rottenberg, 1991; Stecher, 2002; K. W. White & Rosenbaum, 2008). Studies have also indicated that teachers adjust their teaching and assessment styles to match those found on high-stakes tests. In many cases this involved an increase in the use of multiple-choice questions (Au, 2007; Hamilton et al., 2007; Smith, 1991; Smith & Rottenberg, 1991; K. W. White & Rosenbaum, 2008). Stecher (2002) identified negative coaching as teachers spending large amounts of time coaching students on test-taking strategies and practice passages in lieu of time spent teaching content. Other researchers have likewise documented the need teachers feel to teach to the test (Herman & Golan, 1993; Hillocks Jr. & Wallace, 2002). Additional negative effects cited in the literature include cheating (e.g., teachers giving hints, changing answers) (Amrein-Beardsley, Berliner, & Rideau, 2010; Hoffman, Assaf, & Paris, 2001; Nichols & Berliner, 2007; Wilson, Bowers, & Hyde, 2011), emotional stress (Hargrove et al., 2000; Herman & Golan, 1993; Sheldon & Biddle, 1998; Smith, 1991; Smith & Rottenberg, 1991; Triplett & Barksdale, 2005), and the use of educational triage practices (Booher-Jennings, 2005; Neal & Schanzenbach, 2010; Reback, 2008) in which teachers focused on near passing students while providing less instructional time with the lowest performing students.

Adverse Impact in High-Stakes Testing

Numerous researchers have also found that these unintended consequences are most prevalent among schools with low-income and ethnic minority students, thus suggesting that any academic benefits of these policies likely occurred at the expense of the most vulnerable of learners (W.-P. Hong & Young, 2008). For example, Herman and Golan (1993) showed that schools serving low socio-economic-students (SES) spent more time teaching to the test than schools serving higher SES students.

Similarly, Hillocks and Wallace (2002) found increased teaching to the test in schools serving low-SES students. They conducted a case study contrasting the differences between an affluent school and a poor school in Texas preparing students for TAAS. The affluent school was a suburb school with only 5% economically disadvantaged students, while the poor school was an urban school with 96% classified as economically disadvantaged. Unlike teachers at the affluent school, who received progressive writing instruction through a National Writing Project, teachers at the low-SES school received training on test preparation, spent more time teaching to the test, and even postponed instruction in non-tested subjects.

Similarly, Diamond and Spillane (2004) compared instructional practices at four schools under a high-stakes testing policy in Chicago. Two of the schools were on probation for producing low test scores and two were not. The two probation schools consisted largely of low-income students (over 97%), 100% of whom were African American. The two schools that were not on probation had a smaller percentage of low-income students (69% and 85%, and one of those schools was majority White. They found that the probation schools targeted instruction in certain subjects and grades based on the subjects being tested and to whom, whereas the non-probation schools focused on all subjects equally and emphasized improvement for all students in every grade. Probation schools adopted interventions only for specific sub categories of students to raise key test scores whereas non-probation schools adopted interventions for all students. In addition, probation schools focused on strategic ways to raise overall test scores while non-probation schools used data to inform instruction.

Diamond and Spillane (2004) argued that a lack of resources and the extra pressure placed on non-probation schools resulted in the instructional differences. Such studies suggest that the positive gains in aggregate scores in districts and states produced through high-stakes testing policies occur most often in White, middle-class schools. However, those positive outcomes are eclipsed by the unintended, negative consequences that occur in low-income, ethnic minority schools.

Research on Teacher-Based Grade Retention

Teacher-based grade retention has been heavily studied over the last 60 years and has produced some of the most consistent findings in the research literature (House, 1989). Additionally, numerous meta-analyses (e.g., Holmes, 1989; Holmes & Matthews, 1984; Jimerson, 2001) and literature reviews (e.g., Jimerson, Anderson, & Whipple, 2002; Shepard & Smith, 1989; Xia & Kirby, 2009) have been published synthesizing this research. Below, I chronologically discuss these meta-analyses and literature reviews, focusing specifically on areas of agreement and disagreement and highlighting the various contributions each review makes.

Retention Meta-Analyses and Reviews

One of the first meta-analyses conducted on teacher-based retention was Holmes and Matthews (1984). See Table I for further information about these reviews. Reviewing the effects of retention on academic achievement, student attitudes toward school, and personal adjustment,

Holmes and Matthews (1984) found that promoted students achieved higher academically than retained students in language arts, reading, mathematics, work study skills, social studies, and grade-point averages. Retained students did not have as favorable of attitudes toward school as did promoted students, and retained students also scored lower than promoted students on personal adjustment measures including three subareas: social adjustment, emotional adjustment, and behavior.

Table 1
Meta-Analyses and Literature Reviews on Teacher-Based Retention

Author/Year	Research Questions	Methods	Period Covered	Number of Studies
Holmes & Matthews, 1984	What is the effect of grade retention on academic achievement, student attitudes toward school, and personal adjustment in the elementary and junior high school grades?	meta-analysis	1929-1981	44
Holmes, 1989	What is the effect of grade retention on academic achievement and personal adjustment?	meta-analysis	1929-1987	63
Shepard & Smith, 1989	What are the effects of grade retention and school policies and practices regarding retention?	book	1989	8
Jimerson, 2001	What are the effects of grade retention on academic as well as socioemotional and behavioral outcomes?	meta-analysis	1990-1999	20
Jimerson, Anderson, & Whipple, 2002	What is the relationship between grade retention and dropping out of high school?	literature review	1970-2000	17
Xia & Kirby, 2009	What are the effects of grade retention on students' academic and socioemotional outcomes?	literature review	1980-2008	91
Allen, Chen, Willson, & Hughes, 2009	What is the effect of grade retention on academic outcomes?	meta-analysis	1990-2007	22

Holmes (1989) extended the Holmes and Matthews (1984) meta-analysis to include 63 retention studies and found negative effects occurring from retention in 54 of them. Retained students had lower achievement in language arts, reading, math, and social studies than promoted students. Retained students also scored lower on personal adjustment measures than promoted students though not statistically significant differences in the subcategories of social adjustment, emotional adjustment, and behavior. Of the nine studies that did show positive effects, the retention policies implemented in these settings also included early identification and special help for retained students through individual education plans, continuous evaluation, and low student-teacher ratios. The positive studies also included an unusual number of White and middle-class retainees with high IQs and did not follow up past the first year.

Shepard and Smith (1989) reviewed several studies on the effects of retention and school policies and practices regarding retention. Their review was among the first to address a variety of retention issues in addition to academic achievement (e.g., the relationship between retention and dropping out of high school, transition programs, and teacher, parent, and administrators' beliefs about retention). Shepard and Smith (1989) drew the following conclusions: (a) retention in grade does not benefit students academically or in personal adjustment in any way; (b) retention increases the likelihood of dropping out of school by 20-30%, even when controlling for achievement, socioeconomic status, and gender; (c) retaining students in kindergarten, even in a transition program, does not boost academic achievement or solve school readiness problems; (d) from the students' perspectives, retention is harmful; (e) and finally, despite the research findings listed above, teachers, parents, and school administrators often believe that retention is quite beneficial.

Jimerson (2001) conducted a meta-analysis examining the effects of retention on students' academic, socioemotional, and behavioral outcomes. Like the reviews mentioned above, Jimerson (2001) found that of the 20 studies he examined, 16 concluded that retention was not an effective strategy for boosting students' academic achievement and socioemotional adjustment. Consequently, Jimerson (2001) argued that since both social promotion and retention are ineffective strategies for helping low-performing students, more research should be devoted to learning about effective intervention strategies for helping these students.

Jimerson, Anderson, and Whipple (2002) conducted a systematic literature review focusing specifically on retention as a predictor of dropping out of high school. Consistent with the findings of Shepard and Smith (1989), the authors found that retention was one of the most powerful predictors of dropping out, and students who are retained more than once are at a considerably greater risk of dropping out.

By far, the largest, most comprehensive systematic literature review of the effects of retention on students' academic and socioemotional outcomes was recently conducted by the RAND Corporation (Xia & Kirby, 2009) in preparation for an evaluation of a test-based grade retention policy in New York City. Xia and Kirby (2009) examined 91 studies concerning retention published between 1980 and 2008 and produced findings that both complimented and challenged the conclusions of earlier studies. Like previous meta-analyses, Xia and Kirby (2009) found that retention alone is ineffective for increasing students' academic achievement. Although retained students may make significant gains during the retention year, they are usually not large enough to get retained students to the same level as promoted students. Xia and Kirby (2009) found that the vast majority of studies that demonstrated immediate gains from retention also showed that those effects began to dissipate two to three years after the retention, and completely disappeared after several years with retained students falling behind again.

Lorence and Dworkin's (2006) and Lorence, Dworkin, Toenjes, and Hill's (2002) studies in Texas and Alexander, Entwisle, and Dauber's (1994) longitudinal study in Baltimore found much longer-lasting positive effects of retention, but they too decreased over time. Xia and Kirby (2009) did note that in some of these studies where students showed longer-lasting gains (e.g., Lorence & Dworkin, 2006), an intervention was also provided; however, researchers were unable to determine if the improvement was linked to retention or intervention.

Unlike the previous meta-analyses, Xia and Kirby (2009) included newer studies that suggested the social, emotional, attitudinal, and behavioral effects on retained students were mixed and not solely negative. However, retention was associated with a higher likelihood of students' dropping out of school and working in low paying jobs, and a lower likelihood of pursuing postsecondary education. Retention was also found to occur most frequently among the most vulnerable of students (e.g., male, ethnic minority, low SES, youngest in their grade level, most

school transfers, living in single-parent households). All of these findings were consistent with previous reviews.

In addition to newer research, some of the findings of prior studies showing fewer negative consequences from retention could be linked to the quality of research selected. Xia and Kirby (2009) limited their review to empirical studies that used credible comparison groups or statistical methods to control for selection bias. They noted that a limitation of the meta-analyses by Holmes and Matthews (1984) and Holmes (1989) was that many of the studies they included were dissertations and master's theses that were dated and lacked methodological rigor. Similarly, Jimerson's (2001) meta-analysis included several studies that contained small sample sizes of less than 100 retained or promoted students. Only three studies had sample sizes of 1,000 students or more in the retained and comparison groups. Jimerson (2001) did not weight the study effects by sample size.

The most recent meta-analysis of the retention literature focused on the quality of the research design of the studies included in the sample. Allen, Chen, Willson, and Hughes (2009) examined 207 effect sizes across 22 studies using multilevel modeling to investigate the effect of retention on academic outcomes published between 1990 and 2007. They found that the use of more rigorous statistical controls was associated with fewer negative effects, challenging research that suggests retention has a negative effect on achievement (e.g., G. Hong & Yu, 2007). Although they did not find negative effects from retention, they did not find positive effects from it either and concluded that there is little justification for the claim that there are benefits of retention. The Allen et al. (2009) meta-analysis is consistent with other reviews in finding no benefits for retention. However, it differs by showing that quality of design is associated with fewer negative effects, thus suggesting that retention may not be as harmful to students as previously thought.

Adverse Impact in Teacher-Based Retention

As with high-stakes testing, the research on teacher-based retention points to a similar conclusion. Although there is some evidence that retained students experience academic gains that result from teacher-based retentions, those gains ultimately come at the expense of the most vulnerable of students. Low-income, ethnic minority students are most often targeted for retention (Xia & Kirby, 2009). Even in the cases where these students do receive an academic boost from repeating a grade, those gains fade over time. The children eventually fall behind and are at a much higher risk of dropping out of school.

Research on Test-Based Grade Retention Policies

The majority of researchers who have conducted studies on test-based retention policies have attempted to assess whether policies that combine retention with intervention improve student achievement and help low-performing students catch-up academically with their similarly aged peers. Despite a significant amount of research finding negative consequences of teacher-based grade retention, the popularity of test-based retention policies has continued to grow. Moreover, researchers have argued that studies need to be conducted assessing the outcomes of test-based grade retentions because they are qualitatively different from teacher-based retentions (Allensworth & Nagaoka, 2010; Greene & Winters, 2007). Test-based grade retention provides a different context and basis for retention decisions and thus produces different experiences with retention. Test-based retention policies also have a potential spillover effect on students who are not retained (Allensworth & Nagaoka, 2010; Greene & Winters, 2007). The threat of retention itself, along with extra supports like after-school tutoring and summer school are likely to motivate students who are

promoted to work harder as well, especially students who place a high value on passing such tests and believe a passing score is within their reach (Roderick & Engel, 2001).

Although the research on teacher-based retention is more conclusive in terms of the negative effects associated with it, the findings on test-based retention are mixed and based on a more limited pool of studies. Before discussing the findings, I describe the test-based retention policies that have been researched to date and provide an overview of the methods that have been used to understand their outcomes.

Test-Based Retention Policies

The bulk of the research on test-based retention policies has been conducted in Chicago, Florida, and New York City. A few studies have been conducted in Texas, Georgia, Wisconsin, and Louisiana. See Table II for a brief description of these policies, their implementation year(s), requirements, and exceptions as they are described in the research literature.

Table 2

Overview of Researched Policies

Location	Implementation Year(s)	Requirements	Exceptions	Research
Chicago	1996	required that third, sixth, and eighth graders reach a cutoff score on the ITBS for promotion	students in special education and students who were in bilingual education for three years or less; the policy was amended in 2000-2001 due to a civil rights complaint to use a range around the cutoff scores rather than a strict standard for promotion and allowed for consideration of students' grades, attendance, and teacher recommendations for retention decisions	Consortium on Chicago School Research studies (e.g., Jacob & Lefgren, 2004; Roderick & Engel, 2001; Roderick & Nagaoka, 2005)
Florida	2003	requires that third graders score a Level 2 (of five levels) on the reading portion of the FCAT for promotion	students with limited English proficiency or a severe disability, students who score above the 51 st percentile on the Stanford-9, students who demonstrate proficiency through a performance portfolio, or students who have been retained twice	Greene & Winters, 2007, 2009; Winters & Greene, 2006, 2012
New York City	2003-2004	requires that students in grades 3, 5, 7, and 8 score a Level 2 or higher (out of four levels) on the New York State English language arts and mathematics tests in order to be promoted	students who fail the tests in the spring are promoted if they demonstrate proficiency through a portfolio review of their spring work, pass the summer standardized tests, complete a portfolio review of their summer work, or appeal the process	McCombs, Kirby, & Mariano, 2009

Table 2 (Cont.'d)
Overview of Researched Policies

Location	Implementation Year(s)	Requirements	Exceptions	Research
Georgia	2003	requires that students in grades 3 pass one of two administrations of the CRCT in reading and students in grades 3 and 5 pass the CRCT in reading and math to be promoted	parents or teachers may appeal a retention through a Grade Placement Committee; the committee must unanimously agree for a child to be “placed” in the next grade	Henry, Rickman, Fortner, & Henrick, 2005; Livingston & Livingston, 2002; Mordica, 2006
Texas	2002-2003	requires that students pass one of three administrations of the TAKS/now STAAR (State of Texas Assessments of Academic Readiness) in math and reading in grades 3, 5, and 8 to be promoted to the next grade; the law was amended in 2009 to apply only to grades 5 and 8	parents only may appeal a retention through a Grade Placement Committee; the committee must unanimously agree for a child to be “placed” in the next grade	Booher-Jennings, 2008; Valencia & Villarreal, 2005
Wisconsin	2002-2003	required that students in grades 4 and 8 earn at least a basic score on the Wisconsin Knowledge and Concepts Exam (WKCE) to be promoted	the law was amended in 1999 to place retention decisions in the hands of local districts; school districts were to determine grade promotion in grades 4 and 8 by considering WKCE scores as well as other factors	Brown, 2007
Louisiana	2000-2001	required that students in grades 4 and 8 pass the LEAP 21 in language arts and math to be promoted to the next grade	this policy was suspended in 2009	Valencia & Villarreal, 2005

As can be seen from Table II, most of these policies were implemented in a close proximity, time wise, and are quite similar in the requirements and supports they provide. Many of these policies have been initiated by high-profile politicians. Mayor Richard M. Daley started the Chicago policy in 1996 when he was granted power to take over the Chicago Public Schools by the Illinois legislature. Similarly, in 2002, the New York state legislature granted Mayor Michael Bloomberg control of the New York City school system, and he implemented the Children First Initiative which included a series of new programs including a test-based retention policy. Some politicians acknowledged that they were inspired by the legislation of other states when deciding to implement test-based retention. Georgia Governor Roy Barnes, for example, referenced the policy in Texas promoted by then Governor George W. Bush when he urged the Georgia legislature to end social promotion in his 2001 State of the State address (Barnes, 2001). Governor Jeb Bush implemented the policy in Florida soon thereafter (Winters & Greene, 2012).

All of these policies are similar in that they require students to pass a standardized test for promotion. Likewise, they all combine components such as intervention, after-school tutoring, Saturday school and/or summer school in addition to retention. However, they do differ somewhat in their specific requirements for promotion. For example, Texas¹, Georgia (Livingston & Livingston, 2002), Louisiana (Valencia & Villarreal, 2005) and Chicago (Russo, 2005) have all required a passing score on a single standardized test at certain gateway grades for promotion, whereas New York City (McCombs, Kirby, & Mariano, 2009), Wisconsin (Brown, 2007) and Florida (Winters & Greene, 2006) have allowed for additional indicators such as an assessment portfolio or an alternative standardized test. These policies have also differed somewhat in the numbers of students who have been retained. Chicago initially retained between 7,000 and 10,000 students per year, roughly 20% of third graders and 10% of eighth graders (Roderick & Nagaoka, 2005; Russo, 2005). However, after receiving a civil rights complaint about the policy, Chicago schools softened the promotion requirements in 2000-2001 and retained a much smaller percentage of students thereafter. In Florida, the retention numbers have been high: 12-14% of third graders were retained during the initial years of the policy (Winters & Greene, 2012). In New York City, only 2 to 3% of fifth graders were retained in the first two policy cohorts and only 1% in the third cohort (McCombs et al., 2009). Similarly, in Georgia, 61 to 68% of third graders in 2003-2004 who failed both administrations of the Criterion-Referenced Competency Tests (CRCT) in third grade were “placed” in the next grade through an appeals procedure (Henry, Rickman, Fortner, & Henrick, 2005; Mordica, 2006). These variations in implementation and retention rates may explain some of the different outcomes these studies have documented (Greene & Winters, 2007).

Methodological Overview

Table II lists the research that has been conducted on the test-based retention policies in Chicago, Florida, New York City, Georgia, Texas, Wisconsin, and Louisiana. A wide range of methodologies has been used to understand these policies and assess their effects. For example, in Georgia, both Henry et al. (2005) and Mordica (2006) conducted an evaluation of the policy's first year in 2003-2004 using statistical data provided by the state. In Georgia, Texas, and Louisiana, two studies (Livingston & Livingston, 2002; Valencia & Villarreal, 2005) addressed the issue of adverse impact. When the Georgia legislature passed their test-based retention policy, Livingston and Livingston (2002) compared Georgia CRCT scores with student demographics to predict if adverse impact might occur among African American and impoverished students. Similarly, in Texas, Valencia and Villarreal (2005) correlated failing the Texas Assessment of Knowledge and Skills (TAKS) with being African American or Mexican American (or other Latino) to assess if there was

¹ See http://www.tea.state.tx.us/index3.aspx?id=3230&menu_id3=793

an adverse impact of the policy in Texas. In Louisiana, Valencia and Villarreal were able to confirm adverse impact among African Americans by comparing retention rates prior to policy implementation to retention rates after implementation.

Two qualitative studies have been conducted as well, one in Texas and one in Wisconsin. Booher-Jennings (2008) drew on the theoretical concepts of hidden curriculum, achievement ideology, and gender codes to understand how Texas students were responding to the retention policy, while Brown (2007) used a multiple streams approach to agenda setting to explain how key actors understood the need to construct and implement the policy in Wisconsin to improve students' academic achievement.

The majority of the research in Chicago has been conducted by the Consortium on Chicago School Research (Roderick, Nagaoka, & Allensworth, 2005). Their evaluation of test-based retention took place from the first years of the policy's implementation in 1996 to 2001 and used three primary sources of data. First, they used longitudinal school data that included both pre- and post-policy test scores, student demographics, school attendance, an indicator of dropping out, among others. A cohort of students in third and sixth grades were followed from 1993-1994 to 1998-1999 (Jacob & Lefgren, 2004). Second, they surveyed teachers, principals, and sixth- and eighth-grade students to assess their experiences with the policy. Finally, qualitative data were collected including observations of summer school and interviews with teachers.

The research on the New York City test-based retention policy was conducted at the request of the school system by the RAND Corporation. McCombs et al. (2009) analyzed demographic and achievement data for four cohorts of fifth graders, each with about 60,000 students, from 2003-2007. In addition, they conducted numerous case studies in which they interviewed and surveyed administrators, teachers, and students concerning their experiences with the policy. In Florida, Winters and Greene (Greene & Winters, 2007; Winters & Greene, 2006, 2012) have focused primarily on the demographic and achievement data of five cohorts of third graders beginning in the 2002-03 academic year.

Although consistent with the overall findings of the retention literature at large, the findings of the effects of the policies in Chicago, New York City, and Florida on students' academic achievement do differ to some degree. Roderick and Nagaoka (2005) and Allensworth and Nagaoka (2010) have argued that different findings among researchers concerning the short-term effects of retention tend to occur for three reasons: (a) the comparability of test scores across grades, (b) the ability of researchers to construct adequate comparison groups of retained and promoted students, and (c) the point at which researchers estimate achievement effects. The research teams in Chicago, New York City, and Florida all approached these issues differently. In terms of the comparability of test scores, in Chicago, to estimate achievement effects using the Iowa Test of Basic Skills (ITBS), researchers had to equate scores for comparisons of growth across grades as well as forms and levels of the test. To do this they converted ITBS scores to a logit metric using Rasch models (Roderick & Nagaoka, 2005). The process was much simpler in New York City and Florida. In Florida, scores on the Florida Comprehensive Assessment Test (FCAT) are reported as developmental scale scores which the Florida Department of Education designed to have the same meaning for proficiency across grades and years (Greene & Winters, 2007). Similarly, in New York City, the New York State English language arts and mathematics tests were vertically scaled through 2005, allowing for comparisons across grades and years.

For comparison groups, a method used in Chicago (Jacob & Lefgren, 2004; Roderick & Nagaoka, 2005), New York City (McCombs et al., 2009), and Florida (Greene & Winters, 2007; Winters & Greene, 2012) was regression discontinuity design in which researchers compared the performance of students just below the test-score cutoff for retention (most of whom were retained)

to those who scored just above (most of whom were promoted). Unlike sophisticated matching strategies, regression discontinuity accounts for both observed and unobserved characteristics to provide causal estimates. McCombs et al. (2009) combined a regression discontinuity design with propensity scores and doubly robust regression to estimate the effects of test-based retention on student achievement.

One of the drawbacks of comparing those who barely pass and are promoted to those who barely fail and are retained is the risk of mistaking positive effects from retention with actual regression to the mean (Allensworth & Nagaoka, 2010). Often students who are retained after a particularly bad year will perform better the next year, and students who are promoted after a better-than-average-year will perform more poorly. Roderick and Nagaoka (2005) addressed this possible problem by using growth curve modeling to estimate achievement effects on the basis of a student's entire prior test score history, thus correcting for regression to the mean.

Another possible problem with constructing comparison groups is the use of pre-policy comparisons such as those used by Greene and Winters (2007) who compared gains of students who were retained during the first year of the Florida policy with students who achieved the same low test scores but were not retained because they entered the third grade in the year prior and thus were not subject to retention. Roderick and Nagaoka (2005) have argued that it is likely that some students with low pre-policy test scores might have improved had they been exposed to the incentives of the policy. To address this concern they were able to make cross-cohort comparisons. When the Chicago policy was altered in 2000-2001 due to a civil rights complaint, most students who failed the ITBS were promoted whereas in 1998 and 1999 those with comparable failing scores were retained. This allowed the Chicago research team to compare students with similar scores across cohorts who all were exposed to the same incentives and instructional supports.

Finally, the point at which researchers estimate achievement effects can influence their findings on the short-term effects of retention. In Chicago, Roderick and Nagaoka (2005) compared the test scores of students at the end of their retained year to those of comparable students the same age and in the same cohort who had been promoted. Essentially, Roderick and Nagaoka were attempting to determine if two years of learning in the same grade provided greater academic growth than two years of learning in subsequent grades. They argued that only same-age comparisons should be used to study the effects of retention under high-stakes testing: "If the primary objective is to evaluate the effectiveness of having a student repeat a grade versus moving on to the next grade, then the evaluation should focus on estimating the counterfactual: what would have been the achievement of retained students in the absence of retention" (Roderick & Nagaoka, 2005, p. 311).

In New York City and Florida, McCombs et al. (2009) and Winters and Greene (2012) both used same-grade comparisons in which they compared retained students' test scores after two years in a specific grade to comparable students' scores after one year in the same grade. McCombs et al. (2009) explained that same-age comparisons were not feasible in their study because the New York City exams ceased to be vertically scaled in 2006. However, scores in each grade were equated across years and thus supported same-grade comparisons. McCombs et al. (2009) also argued that same-grade comparisons supported the theory of action behind the retention treatment. Same-grade comparisons focus on the interim grade improvement of an extra year of instruction provides and thus establishes the possibility that retained students may be better prepared at the end of their academic careers. Similarly, Winters and Greene (2012) used same-grade comparisons arguing that the Florida schools' interests concerned comparing students' performance to same grade peers. As I will discuss below, the use of same-age or same-grade comparisons can greatly influence the findings on academic achievement. In the next section I provide an overview of key findings from the studies described above.

Adverse Impact in Test-Based Retention

Studies in Georgia, Texas, Louisiana, and Florida have all examined the characteristics of the students most likely to be retained under test-based retention policies. As with the studies on high-stakes testing and teacher-based retention, researchers have specifically examined whether such policies have an adverse impact on ethnic minority and impoverished students. For example, Livingston and Livingston (2002) conducted a study after the Georgia law was passed but prior to its implementation. They examined CRCT scores from the State of Georgia's Office of Education Accountability and demographic data compiled by the University of Georgia Department of Housing and Consumer Economics for 39 southern counties with high numbers of African American and impoverished students. The demographic data consisted of the following: percentage of African Americans, per capita income, number of children living in poverty, number of African Americans living in poverty, number of female-headed families living in poverty, number of unwed births, and the percentage of the population without a high school diploma. They found that African American children and poor students are much more likely to fail the CRCT and consequently be retained. Livingston and Livingston (2002) argued that implementing test-based retention would have an adverse impact on these students and increase their likelihood of dropping out of school.

Likewise, Valencia and Villarreal (2005) examined the initial TAKS scores for Texas third graders in 2003. Although they were unable to analyze the scores of the second and third retakes², based on the initial scores, they predicted that more Mexican American (or other Latino) students would fail and thus be retained, increasing their likelihood of dropping out.

Valencia and Villarreal (2005) compared retention rates for Louisiana students over a four-year period from 1997 to 2001. The state's test-based retention policy was implemented in 2000-2001, and students began taking the Louisiana Educational Assessment Program for the 21st Century (LEAP 21) for promotion in grades 4 and 8. They found that prior to the policy, 1 in 15 African American fourth graders was retained compared to 1 in 29 Caucasian fourth graders. After the policy was implemented, 1 in 4 African American fourth graders was retained and 1 in 13 Caucasians. An even more disproportionate number of African American students were retained in eighth grade: 1 in 13 African Americans and 1 in 21 Caucasians were retained prior to the policy, and 1 in 3 African Americans and 1 in 10 Caucasians were retained afterwards. Valencia and Villarreal (2005) argued that the shift in retention rates before and after implementation provide evidence that such policies have a disproportionate impact on African American students.

Greene and Winters (2009) went a step further and examined the decisions of state-approved Grade Placement Committee meetings when retentions were appealed. They found that Florida educators discriminated against African American and Latino students when promoting or retaining students in these appeal meetings. African American and Latino students were more likely to be retained (4% and 9% respectively) than Caucasian students, even when controlling for academic achievement.

Short-Term Academic Achievement

The studies conducted in Chicago, New York City, and Florida focused largely on academic achievement and, at least initially, reported positive effects of retention on the initial cohorts they

² At this time, third graders had three opportunities to pass the reading TAKS before being retained. Valencia and Villarreal (2005) completed the chapter in which they presented these results in May 2003, about two months before the second and third administrations of the reading TAKS were given. Thus, because of this time lag, they were unable to report on these results.

investigated. For example, Roderick and Engel (2001) interviewed 102 low-achieving African American and Latino students in Chicago about their experiences with test-based retention. Students were considered low-achieving if based on prior test trajectory they needed average-to-above average learning gains to meet the test-score cutoff. Drawing on motivation theory, Roderick and Engel (2001) examined whether or not students would be motivated to work harder academically if they valued passing test-based retention exams and believed that passing them was possible. They found that 53% reported that the threat of retention motivated them to work harder and pay greater attention in class. The high stakes also appeared to increase the support these students received from teachers and the time students spent studying outside of school.

Roderick, Jacob, and Bryk (2002) found that the test scores in gateway grades in Chicago increased substantially when the high-stakes testing policies began, although the effects were larger for the sixth and eighth grades than for the third. Sixth- and eighth-grade gains were approximately one third to one half of a year's learning gain; results were less conclusive in third grade where policy effects were smaller in reading and less consistent across years. Similarly, in Florida (Winters & Greene, 2006) and New York City (McCombs et al., 2009), students who participated in the retention policy earned higher test scores than those who did not. Third graders in Florida scored 0.06 of a standard deviation higher in reading on both the FCAT and Stanford-9 and 0.14-0.15 of a standard deviation higher in math than equally performing third graders not subject to the policy the previous year. Fifth graders in New York City scored 0.10-0.20 of a standard deviation higher in reading and less than 0.10 of a standard deviation higher in math than those previously not subject to the policy.

In Florida (Winters & Greene, 2006), New York City (McCombs et al., 2009), and Chicago (Roderick & Nagaoka, 2005) students who were retained by these policies received an academic boost, at least in the short term. In Florida, students retained in 2003 scored 0.11-0.13 of a standard deviation (3.45-4.1- percentiles) higher on the FCAT and Standard-9 reading tests than comparable promoted students and 0.28-0.30 of a standard deviation (9.3-10.0 percentiles) higher on the FCAT and Standard-9 in math (Winters & Greene, 2006). In New York City, students who were retained in fifth grade scored moderately higher on seventh-grade assessments (0.40-0.60 of a standard deviation) than comparable students who were not retained (McCombs et al., 2009). In New York City, the retained students' gains persisted two years after retention (McCombs et al., 2009), and in Florida, retained students' gains increased the second year so that students had about 0.40 of a standard deviation higher academic proficiency than comparable promoted students (Greene & Winters, 2007). In Chicago, Roderick and Nagaoka (2005) found that retained third graders did receive a small academic boost the year after retention, but no substantial positive effects.

Research in Chicago, New York City, and Georgia also emphasized the effectiveness of the interventions that were implemented in tandem with test-based retention policies. In Chicago, Jacob and Lefgren (2004) found that summer school increased third graders' reading and math achievement two years later by about 12% of the average annual learning gains, although for sixth graders the effects were about half as much. In New York City (McCombs et al., 2009) and Georgia (Henry et al., 2005), students who attended summer school scored higher on subsequent tests than those who did not. In New York City, students who attended summer school in fifth grade through the test-based retention policy later scored 0.10-0.15 of a standard deviation higher on both sixth and seventh grade ELA and math tests than comparable students who did not attend summer school. Similarly, in Georgia, students who attended summer school scored on average over four points higher than non-enrolled students on the summer 2004 CRCT retake test in reading, and were more likely to pass it than comparable students who did not attend summer school.

Stone and Jacob (2005) found that the test-based retention policy was well-liked in Chicago schools by teachers, principals, and students. Teachers' time spent on test preparation increased substantially immediately after the policy was implemented but declined somewhat in later years. Teachers provided more time on grade-level materials in reading and math skills relevant to the test, and sixth and eighth graders received greater instructional support and reported being more academically engaged than they were prior to the policy. Finally, McCombs et al. (2009) reported that retained students in New York City did not exhibit negative emotional effects. Student surveys showed that retention did not harm their confidence in reading or math and that they reported a greater sense of connectedness to school than at-risk students who were promoted and students who were not-at risk even three years later.

Long-Term Academic Achievement

Although the short-term gains in academic achievement described above look promising, several studies have suggested that these positive outcomes may have occurred at the expense of the most vulnerable of students and fade over time. For example, Roderick and Engel (2001) found that 53% of low-achieving African American and Latino students in Chicago were working harder and paying greater attention in school than they reported they were prior to the implementation of the retention policy. However, 34% of the students reported that they were not motivated by the high-stakes test and consequently did poorly. This latter group of students included the most vulnerable students, who faced significantly larger skill gaps and barriers to learning. Roderick and Engel (2001) noted that test-based retention policies may benefit certain students while making "sacrificial lambs" out of those unable or unwilling to pass the required exams (p. 221). In other words, the increased motivation that the majority of students in Chicago experienced may have been produced by sacrificing the educational opportunities of the still sizeable group of students who failed.

Additionally, similar to the findings on teacher-based retention, the positive effects on academic achievement from test-based retention fade over time as cohorts of students are followed longitudinally. For example, in Chicago, Roderick and Nagaoka (2005) found that the small academic boost third-grade students received from retention dissipated to zero two years out, and retained sixth graders actually declined in academic growth. Retention in sixth grade was associated with negative growth in achievement one year after the retention year with the negative effects remaining two years later. Roderick and Nagaoka (2005) estimated that retained sixth graders experienced 6%-21% lower achievement growth than comparable non-retained students. The results also revealed that retained students in Chicago were much more likely to be placed in special education and thus be exempted from testing during their retained years. Teachers were given little guidance in working with retained students and thus usually gave them a second dose of the interventions from the previous year, a finding also documented by Stone and Engel (2007). The intervention provided during summer school was a scripted, test-preparation program that focused on skills needed for passing the ITBS (Roderick, Jacob, & Bryk, 2004).

In Florida, Greene and Winters (2007) initially found that retained students' academic gains persisted two years out, and even increased in the second year. However, an analysis conducted five years after retention indicated that although retained students were still benefiting from retention and intervention, the positive effects were dissipating (Winters & Greene, 2012). Winters and Greene (2012) found that the academic boost that students retained in third grade received from retention decreased from a high of 0.40 of a standard deviation to 0.183 of a standard deviation in reading and 0.174 of a standard deviation in math by the time they entered sixth grade. The McCombs et al. (2009) study in New York City only investigated two years beyond retention and found positive effects throughout those two years.

One possible explanation for the different findings on the long-term effects of test-based retention on achievement is that the Chicago researchers used same-age comparisons while the researchers in Florida and New York City used same-grade comparisons and the choice of comparison group may influence the findings generated by a study. Same-age comparisons typically produce negative effects in the short-term that plateau or become more positive with time, while same-grade comparisons are initially positive but decrease over time (Allen et al., 2009; Allensworth & Nagaoka, 2010). Another possible explanation for the differences in the short-term gains found in Chicago and the longer-persisting gains in Florida requires a closer look at what is being measured. In Chicago, the researchers were able to measure the impact of retention itself by disentangling it from the effects of summer school. As per the policy, Chicago students were able to take a second administration of the ITBS in summer school and avoid retention (Jacob & Lefgren, 2004). This enabled the researchers to compare the future growth of students who passed the retake exam at the end of summer school and were promoted to students who failed the retake and were retained. In Florida, however, retesting is less uniform, and thus the researchers were unable to separate the effects of retention from the effects of the intervention implemented as a component of the retention policy (Greene & Winters, 2007).

Increased Likelihood of Dropping Out

Thus far, only two studies on the Chicago retention policy have examined the long-term effects of test-based retention on school dropout rates. Allensworth (2005) linked retention in Chicago to dropout rates. She compared eighth-grade cohorts of students before and after implementation of the Chicago policy and showed that retention based on test scores did increase low-achieving students' likeliness of dropping out of school, although the relationship was smaller than the dropout rates associated with traditional teacher-based retention. Interestingly, Allensworth (2005) also found that small decreases in the dropout rates among those not retained counterbalanced the higher number of dropouts among those retained so that the overall retention rate slightly decreased. This finding suggested that although the overall drop-out rate slightly decreased under the policy, it may have done so by increasing the drop-out rate of those who failed the gateway exams and were retained, thus providing further evidence that positive effects from the policy came at the expense of the most vulnerable students.

Jacob and Lefgren (2009) also found a link between Chicago's test-based retention policy and high school dropout rates. They compared eighth graders to sixth graders during the first three years of the program (1997 to 1999) and found that retaining low-achieving eighth-grade students substantially increased the likelihood that they would drop out of high school.

Promotes and Demotes: Moral Boundary Work

Anagnostopoulos (2006) examined Chicago's test-based retention policy at the high school level in which ninth graders who failed end of year standardized math and reading tests were demoted. In Chicago, at the ninth-grade level, students were demoted by the test-based retention policy rather than retained. Demotion differed from retention in that demoted ninth graders were required to attend a homeroom class designated for demoted students and enroll in remedial math and reading courses at the ninth-grade level, but they still could take other tenth-grade courses. Anagnostopoulos (2006) found that high-school students and teachers used test-based retentions to create social boundaries that distinguished promoted students from demoted ones. Using a cultural sociological perspective, she showed that instead of encouraging teachers and ninth-grade students to achieve academically, the policy promoted a kind of moral boundary work in which teachers justified not providing demoted students, whom they considered undeserving, with enriching

learning opportunities. Success or failure on the test provided fodder for student identity constructions and social exclusion.

Masking Social Inequities

Finally, a few studies have suggested that test-based retention policies are reinforcing the ideology that success on high-stakes tests is solely the result of effort while masking the connection between educational achievement and social inequities within the U.S. Drawing on Bourdieu (1982/1991; Bourdieu & Passeron, 1970/1990), Anagnostopoulos (2006) showed that at the high school level, Chicago's test-based retention policy enacted symbolic violence on low-income, ethnic minority demoted students by obscuring the connection between test scores and class inequities while imposing the belief that educational achievement is largely based on moral decisions such as good behavior in school, self-discipline, and perseverance.

Similarly, Booher-Jennings (2008) found that under Texas' test-based retention policy, teachers exposed students to the hidden curriculum of achievement ideology. Through their day-to-day words and actions, teachers communicated to students that success on the state test was based on hard work and individual effort. However, Booher-Jennings (2008) also noticed the teachers differed how they communicated this message to boys and girls. Teachers blamed boys who failed for their poor behavior and bad attitudes. The teachers tended to tell girls that they just need more self-esteem to pass the test. Out of the 37 students Booher-Jennings (2008) interviewed, the vast majority believed that it was fair to promote students to the next grade based on their scores on a standardized test. Most boys accepted the teachers' reasons for their failure, and girls who failed worked hard to show others they were not like the boys who just did not try. Only three students, all boys, questioned the fairness of test-based retention and expressed doubt that working hard in school would benefit their futures.

Discussion: Achievement at Whose Expense?

It is evident that test-based retention has resulted in some of the intended benefits. For example, high-stakes tests can improve alignment between curriculum and instruction (Hamilton et al., 2007; Koretz et al., 1994; Stecher, 2002). Testing has been shown to help teachers identify students' needs and motivate teachers and students to work harder (Finnigan & Gross, 2007). In both teacher and test-based retention programs that incorporate interventions, students have shown to make short-term academic gains (Greene & Winters, 2007; Lorence et al., 2002; McCombs et al., 2009; Roderick et al., 2002; Winters & Greene, 2006; Xia & Kirby, 2009). These programs appear to be popular and motivate the majority of at-risk students to work harder (Roderick & Engel, 2001; Stone & Jacob, 2005).

On the other hand, negative, unintended consequences are often an outcome of these policies and adversely affect the most vulnerable of students. High-stakes testing policies have consistently resulted in negative curriculum reallocation, encouraging teachers to adapt their teaching styles to test formats, negative coaching, cheating, and educational triage practices (Booher-Jennings, 2005; Heilig & Darling-Hammond, 2008; McNeil, Coppola, Radigan, & Heilig, 2008). All of these practices produce score inflation (Koretz, 2008) and appear to be most prevalent in probationary schools with large numbers of low-income, ethnic minority students (W.-P. Hong & Young, 2008). These negative, unintended consequences are evident when retention is tied to high-stakes testing as well.

Researchers have consistently found that the academic boosts produced from teacher-based retention are short-lived, with the retained students falling behind again in later years (Xia & Kirby, 2009). Those retained are often the most vulnerable students, and retention increases the likelihood

that these students will later dropout of school (Xia & Kirby, 2009). Although many teachers, administrators, and the public at large (Bulla & Gooden, 2003; Byrnes, 1989; Smith, 1989; Tomchin & Impara, 1992) assume teacher-based retention will help low-performing students, instead it greatly increases the likelihood that many of these students will retreat from educational experiences.

With test-based retention policies, although the majority of students may be motivated to work harder, a significant number of low-achieving students appear unaffected by these policies (Roderick & Engel, 2001). In some cases, the curriculum and teaching methods the students experience during their retained year is not much different than those used in their classes prior to the retention (Stone & Engel, 2007), and retained students are at an increased risk for dropping-out (Allensworth, 2005; Jacob & Lefgren, 2009). Like teacher-based retention, academic gains through test-based retention fade over the long run (Winters & Greene, 2012). Gains in Chicago faded in the second year (Roderick & Nagaoka, 2005).

As seen above, a growing amount of evidence suggests that the academic gains that appear to result from test-based grade retention policies are likely occurring at the expense of the most vulnerable of students. Although these gains for some, but not all students are only short term gains, Linn (2000) has argued that for politicians, short-term gains may be all that is needed. In a case study on Wisconsin's test-based grade retention policy, Brown (2007) argued that Wisconsin policymakers implemented their policy "not to hold students back but rather to instill accountability into the educational system" (p. 4). Legislators were being pressed to raise achievement statewide. They saw the retention policy as a means to boost student achievement through increased accountability. Overall achievement, not the fate of those retained was their main concern. Retaining students was an unfortunate necessity to help foster the public perception that schools were maintaining high standards and that the majority of students were being encouraged to do better. Policymakers viewed retention as a byproduct of improving academic performance and not as an intervention itself. The harmful effects of retention were not a problem for Wisconsin policy makers unless they affected large numbers of students. Such findings echo the claims of proponents of test-based retention policies such as Russo (2005) who argued that "...student retention policies are not really about the students who are retained as much as they are about the way the rest of the school system operates when it knows there is not social promotion" (p. 47).

Implications for Policy Makers and Researchers

Several professional organizations have issued statements that have urged policy makers to abandon retention practices based on single, high-stakes test scores (AERA, APA, & NCME, 1999; American Educational Researchers Association, 2000; Dennis et al., 2012; Heubert & Hauser, 1999; National Association of School Psychologists, 2003). Standardized tests are only an estimate within a margin of error based on a small sample of questions in a certain area and should not be treated as an exact measure of student knowledge.

Penfield (2010) assessed if test-based grade retention is aligned with the National Research Council's (Heubert & Hauser, 1999) standards for fair and appropriate test use. He found that test-based grade retention violates standards related to attribution of cause and effectiveness of treatment. Penfield (2010) cited evidence that test scores for nondominant groups could be attributed to poor instruction or linguistic and cultural content of the assessment rather than their knowledge and skills. Second, research suggests that retention is a potentially harmful placement. If retention harms students' academic performance or increases the likelihood that students will drop out, this could be a violation of fair and appropriate test use.

Such consistent ethical concerns by professional educational organizations, along with the growing research documenting the harmful effects of test-based retention policies provide ample evidence that policy makers should strongly consider discontinuing these policies. However, ending test-based retention should not imply that social promotion is a beneficial alternative. Researchers have argued that simply retaining students without providing different instruction places the blame for low academic achievement solely on the student and offers little hope for improvement. However, simply promoting students to the next grade without additional support is a failed strategy as well (Darling-Hammond, 1998; Owings & Kaplan, 2001).

Nevertheless, retention and social promotion are not the only options available. Researchers have suggested numerous alternatives that include using classroom assessments that better inform teaching, and more effectively implementing differentiated and small group instruction (Dennis et al., 2012). Two practical alternatives suggested by researchers (e.g., Darling-Hammond, 1998; Smink, 2001; Smith & Shepard, 1989) and made evident by the studies examined in this review are increasing instructional effectiveness and increasing instructional time.

Darling-Hammond (1998) has advocated the need for improving skillful teaching as an alternative to retention, a point emphasized by the Chicago researchers in this review as well. Allensworth and Nagaoka (2010) noted that retention and not staff development was the focus of the Chicago policy. Few structures were established to improve teaching quality and thus retained students often received a second dose of the same instruction when they were retained. Interestingly, although they found that the city's summer school program (which was heavily scripted) did improve student achievement, they found that students whose teachers intentionally altered the script to meet students' needs performed higher than those who simply followed the script, leading the researchers to believe that teacher expertise made a difference (Roderick et al., 2005).

A second alternative to retention that is underscored by the findings of this review is increasing instructional time. The same-grade comparisons that were used to assess the New York City and Florida policies suggest that students who are given extra time to master material in a specific grade perform higher than those with less time to master the same material. Retention, however, is just one way of adding additional instruction. The studies in Chicago, Florida, and New York City all found that if students are provided additional instruction after school and in summer school, academic achievement increased. Additional instructional time has also been shown to be productive in the forms of universal pre-kindergarten (Lazarus & Ortega, 2007) or multi-grade instruction (Darling-Hammond, 1998; Smith & Shepard, 1989). Smith and Shepard (1989) described various approaches for reconceptualizing school organization to increase instructional time. One consists of having ungraded instruction in the primary grades. Another involves allowing a student who is behind in reading to go to a younger grade for instruction just in that subject. In schools where numerous students move among grades, students experience fewer stigmas related to being older than their peers. Finally, teachers can promote students who are still behind academically but work with their teachers in subsequent years to develop individualized intervention plans for the children.

In terms of implications for researchers, additional attention needs to be given to the short and long-term effects of test-base retention on student motivational processes. Studies on test-based retention that address motivation tend to simply look for evidence that students are working harder based on the degree they value passing the assessment and to what extent they believe passing is actually possible (e.g., Finnigan & Gross, 2007; Roderick & Engel, 2001). Additional attention needs to be paid to development of students' motivational attitudes toward school and learning.

Second, the vast majority of studies on test-based retention have been large-scale, quantitative studies seeking to determine if these policies improve students' academic achievement on standardized tests (e.g., McCombs et al., 2009; Roderick et al., 2002; Winters & Greene, 2006). Only a few qualitative case studies (e.g., Anagnostopoulos, 2006; Booher-Jennings, 2008; Brown, 2007) have attempted to understand how these policies are being negotiated by students, teachers, administrators, and policy makers. Few studies have followed students throughout these policies to better learn how they are actually being implemented in schools and the micro-level effects on students.

The lack of research in this latter area is one that needs to be addressed. As noted earlier, Bourdieu and Passeron (1970/1990) have argued that large-scale, quantitative studies solely focusing on achievement gains as measured by test scores mask the social inequities that produce such scores and the role schools and examinations play in these processes. Further exploration is needed to examine what these tests may be concealing and to flesh out the processes in which these policies obscure the connection between achievement scores and class inequities.

References

- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1994). *On the success of school failure: A reassessment of the effects of retention in the primary grades*. New York, NY: 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*, 480-499. <http://dx.doi.org/10.3102/0162373709352239>
- Allensworth, E. M. (2005). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Educational Evaluation and Policy Analysis, 27*, 341-364. <http://dx.doi.org/10.3102/01623737027004341>
- Allensworth, E. M., & Nagaoka, J. (2010). Issues in studying the effects of retaining students with high-stakes promotion tests: Findings from Chicago. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schooling, and human development* (pp. 327-341). New York, NY: Routledge.
- American Educational Researchers Association. (2000). *AERA position statement concerning high-stakes testing in pre-k-12 education*. Retrieved from <http://www.aera.net/AboutAERA/AERARulesPolicies/AERAPolicyStatements/PositionStatementonHighStakesTesting/tabid/11083/Default.aspx>
- Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education Policy Analysis Archives, 18*(14). Retrieved from <http://epaa.asu.edu/ojs/article/view/714>
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing and student learning. *Education Policy Analysis Archives, 10* (18). Retrieved from <http://epaa.asu.edu/ojs/article/view/297>
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32-38.
- Anagnostopoulos, D. (2006). "Real students" and "true demotes": Ending social promotion and the moral ordering of urban high schools. *American Educational Research Journal, 43*, 5-42. <http://dx.doi.org/10.3102/00028312043001005>
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*, 258-267. <http://dx.doi.org/10.3102/0013189X07306523>

- Barnes, R. (2001). *Georgia state of the state address*. Retrieved from Stateline.org website: <http://www.stateline.org/live/details/speech?contentId=16091>
- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. New York, NY: Basic Books.
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42, 231-268. <http://dx.doi.org/10.3102/00028312042002231>
- Booher-Jennings, J. (2008). Learning to label: Socialization, gender, and the hidden curriculum of high-stakes testing. *British Journal of Sociology of Education*, 29, 149-160. <http://dx.doi.org/10.1080/01425690701837513>
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3-15. <http://dx.doi.org/10.3102/0013189X034006003>
- Bourdieu, P. (1990). *In other words: Essays towards a reflexive sociology* (M. Adamson, Trans.). Stanford, CA: Stanford University Press. (Original work published 1987)
- Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.). Cambridge, MA: Harvard University Press. (Original work published 1982)
- Bourdieu, P., & Passeron, J. C. (1990). *Reproduction in education, society and culture* (R. Nice, Trans. 2nd ed.). London, England: Sage. (Original work published 1970)
- Brown, C. P. (2007). Examining the streams of a retention policy to understand the politics of high-stakes reform. *Education Policy Analysis Archives*, 15(9). Retrieved from <http://epaa.asu.edu/epaa/v15n9/>
- Bulla, T., & Gooden, J. S. (2003). Retention and social promotion: Perspectives of North Carolina elementary school principals. *ERS Spectrum*, 21(3), 19-31.
- Byrnes, D. A. (1989). Attitudes of students, parents and educators toward repeating a grade. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 108-131). London, England: The Falmer Press.
- Cannon, A. (1996, March 28). Governors talk tough on school standards: Clinton calls for tests to earn promotions. *The Miami Herald*, p. 3A.
- Darling-Hammond, L. (1998). Alternatives to grade retention. *The School Administrator*, 55(7), 18-21.
- Dennis, D. V., Kroeger, D. C., O'Byrne, W. I., Meyer, C. K., Kletzein, S. B., Huddleston, A. P., & Gilrane, C. (2012). *Test-based grade retention*. (Policy Brief). Retrieved from Literacy Research Association website: <http://www.literacyresearchassociation.org/publications/LRATestBasedRetentionPBFinal.pdf>
- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106, 1145-1176. <http://dx.doi.org/10.1111/j.1467-9620.2004.00375.x>
- Educational Commission of the States. (2005). *Student promotion/retention policies*. Retrieved from <http://www.ecs.org/clearinghouse/65/51/6551.pdf>
- Educational Commission of the States. (2011). *Recent state policies/activities students--promotion/retention*. Retrieved from <http://www.ecs.org/ecs/ecscat.nsf/WebTopicView?OpenView&count=-1&RestrictToCategory=Students--Promotion/Retention>
- Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal*, 44, 594-629. <http://dx.doi.org/10.3102/0002831207306767>

- Georgia State Board of Education. (2001). *Official Code 20-2-282 through 20-2-285*. Retrieved from Georgia Department of Education website: <http://www.doe.k12.ga.us/External-Affairs-and-Policy/Policy/Pages/Promotion-and-Retention.aspx>
- Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, 2, 319-340. <http://dx.doi.org/10.1162/edfp.2007.2.4.319>
- Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review*, 28, 135-142. <http://dx.doi.org/10.1016/j.econedurev.2008.02.002>
- Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J. L., . . . Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND.
- Haney, W. (1984). Testing reasoning and reasoning about testing. *Review of Educational Research*, 54, 597-654. <http://dx.doi.org/10.3102/00346543054004597>
- Hargrove, T. Y., Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., & Davis, M. (2000). Unintended consequences of high-stakes testing in North Carolina: Teacher perceptions. *ERS Spectrum*, 18(4), 21-25.
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30, 75-110. <http://dx.doi.org/10.3102/0162373708317689>
- Henry, G. T., Rickman, D. K., Fortner, C. K., & Henrick, C. C. (2005). *Report of the findings from Georgia's third grade retention policy*. Atlanta, GA: Andrew Young School of Public Policy
- Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement, Issues, and Practice*, 12(4), 20-25, 41-42.
- Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academies Press.
- Hillocks Jr., G., & Wallace, V. L. (2002). Teachers and administrators under pressure: The case of Texas. In G. Hillocks Jr. (Ed.), *The testing trap: How state writing assessments control learning* (pp. 85-102). New York, NY: Teachers College Press.
- Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *Reading Teacher*, 54, 482-492.
- 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, England: Falmer Press.
- Holmes, C. T., & Matthews, K. M. (1984). The effects of nonpromotion on elementary and junior high school pupils. *Review of Educational Research*, 54, 225-236. <http://dx.doi.org/10.3102/00346543054002225>
- Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis*, 29, 239-261. <http://dx.doi.org/10.3102/0162373707309073>
- Hong, W.-P., & Young, P. (2008). Does high-stakes testing increase cultural capital among low-income and racial minority students? *Education Policy Analysis Archives*, 16(6). Retrieved from <http://epaa.asu.edu/epaa/v16n6/>
- House, E. R. (1989). Policy implications of retention research. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 202-213). London, England: Falmer Press.

- House, E. R. (1998). *The predictable failure of Chicago's student retention program*. Retrieved from Designs for Change website: <http://www.designsforchange.org/pdfs/house.pdf>
- House, E. R. (2004). Using multiple methodologies: The case of retention in Chicago. In K. B. deMarrais & S. D. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 353-367). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hout, M., & Elliott, S. W. (Eds.). (2011). *Incentives and test-based accountability in education*. Washington DC: National Academies Press.
- Hurwitz, N., & Hurwitz, S. (2000). Tests that count. *American School Board Journal*, 187(1), 20-25.
- Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics*, 86, 226-244.
<http://dx.doi.org/10.1162/003465304323023778>
- Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3), 33-58.
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30, 420-438.
- Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, 39, 441-457. <http://dx.doi.org/10.1002/pits.10046>
- Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81, 199-203.
- Klein, S. P., Hamilton, L. S., McCaffrey, D., & Stecher, B. M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives*, 8(49). Retrieved from <http://epaa.asu.edu/ojs/article/view/440>
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Koretz, D., Stecher, B. M., Klein, S. P., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement, Issues, and Practice*, 13(3), 5-16.
<http://dx.doi.org/10.1111/j.1745-3992.1994.tb00443.x>
- Lareau, A. (2003). *Unequal childhoods: Class, race, and family life*. Berkeley, CA: University of California Press.
- Lazarus, P. J., & Ortega, P. (2007). Universal pre-kindergarten in conjunction with universal screenings: An antidote to grade retention. *Journal of Educational Research and Policy Studies*, 7(1), 54-75.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
<http://dx.doi.org/10.3102/0013189X029002004>
- Livingston, D. R., & Livingston, S. M. (2002). Failing Georgia: The case against the ban on social promotion. *Education Policy Analysis Archives*, 10(49). Retrieved from <http://epaa.asu.edu/ojs/article/view/328>
- Lorence, J., & Dworkin, A. G. (2006). Elementary grade retention in Texas and reading achievement among racial groups. *Review of Policy Research*, 23, 999-1033.
<http://dx.doi.org/10.1111/j.1541-1338.2006.00247.x>
- Lorence, J., Dworkin, A. G., Toenjes, L. A., & Hill, A. N. (2002). Grade retention and social promotion in Texas 1994-1999: Academic achievement among elementary school students. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 13-67). Washington, DC: Brookings Institution Press.

- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh year-book of the National Society for the Study of Education* (pp. 83-121). Chicago, IL: University of Chicago Press.
- Marsh, J. A., Gershwin, D., Kirby, S. N., & Xia, N. (2009). *Retaining students in grade: Lessons learned regarding policy design and implementation*. (Technical Report No. 677). Retrieved from RAND Education website: http://www.rand.org/pubs/technical_reports/2009/RAND_TR677.pdf
- Maxwell, J. A. (2006). Literature reviews of, and for, educational research: A commentary on Boote and Beile's "Scholars Before Researchers". *Educational Researcher*, 35(9), 28-31. <http://dx.doi.org/10.3102/0013189X035009028>
- McCombs, J. S., Kirby, S. N., & Mariano, L. T. (Eds.). (2009). *Ending social promotion without leaving children behind: The case of New York City*. Santa Monica, CA, USA: RAND.
- McNeil, L. M., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives*, 16(3). Retrieved from <http://epaa.asu.edu/epaa/v16n3/>
- Millicent, L. (1997, June 11). Promote or retain? Pendulum for students swings back again. *Education Week*, 16(37). Retrieved from <http://www.edweek.org/ew/articles/1997/06/11/37social.h16.html?qs=promote>
- Mordica, J. (2006, April). *Third grade students not meeting standards in reading: A longitudinal study*. (Research Brief No. 02). Atlanta, GA: Georgia Department of Education.
- Morris, D. R. (2001). Assessing the implementation of high-stakes reform: Aggregate relationships between retention rates and test results. *NASSP Bulletin*, 85(629), 18-34. <http://dx.doi.org/10.1177/019263650108562903>
- National Association of School Psychologists. (2003). *Position statement on student grade retention and social promotion*. Retrieved from http://www.nasponline.org/about_nasp/positionpapers/StudentGradeRetention.pdf
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Author.
- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92, 263-283. <http://dx.doi.org/10.1162/rest.2010.12318>
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Educational Press.
- Orfield, G., & Ashkinaze, C. (1991). *The closing door: Conservative policy and black opportunity*. Chicago, IL: University of Chicago Press.
- Owen, S. A., & Ranick, D. L. (1977). The Greensville program: A commonsense approach to basics. *Phi Delta Kappan*, 58, 531-533, 539.
- Owings, W. A., & Kaplan, L. S. (2001). Standards, retention, and social promotion. *NASSP Bulletin*, 85(629), 57-66. <http://dx.doi.org/10.1177/019263650108562906>
- Penfield, R. D. (2010). Test-based grade retention: Does it stand up to professional standards for fair and appropriate use? *Educational Researcher*, 39, 110-119. <http://dx.doi.org/10.3102/0013189X10363007>
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92, 1394-1415. <http://dx.doi.org/10.1016/j.jpubeco.2007.05.003>
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23, 197-227. <http://dx.doi.org/10.3102/01623737023003197>

- Roderick, M., Jacob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24, 333-357. <http://dx.doi.org/10.3102/01623737024004333>
- Roderick, M., Jacob, B. A., & Bryk, A. S. (2004). Summer in the city: Achievement gains in Chicago's Summer Bridge program. In G. D. Borman & M. Boulay (Eds.), *Summer learning: Research, policies, and programs* (pp. 73-102). Mahwah, NJ: Lawrence Erlbaum.
- Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, 27, 309-340. <http://dx.doi.org/10.3102/01623737027004309>
- Roderick, M., Nagaoka, J., & Allensworth, E. (2005). Is the glass half full or mostly empty? Ending social promotion in Chicago. *Yearbook of the National Society for the Study of Education*, 104, 223-259. <http://dx.doi.org/10.1111/j.1744-7984.2005.00032.x>
- Russo, A. (2005). Retaining retention. *Education Next*, 5(1), 42-48.
- Sheldon, K. M., & Biddle, B. J. (1998). Standards, accountability, and school reform: Perils and pitfalls. *Teachers College Record*, 100, 164-180.
- Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention*. London, England: Falmer Press.
- Smink, J. (2001). Alternatives to retention. *NASSP Bulletin*, 85(629), 3-17. <http://dx.doi.org/10.1177/019263650108562902>
- Smith, M. L. (1989). Teacher's beliefs about retention. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 132-150). London, England: Falmer Press.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11. <http://dx.doi.org/10.3102/0013189X020005008>
- Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement, Issues, and Practice*, 10(4), 7-11. <http://dx.doi.org/10.1111/j.1745-3992.1991.tb00210.x>
- Smith, M. L., & Shepard, L. A. (1989). Flunking grades: A recapitulation. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 214-236). London, England: Falmer Press.
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher & S. P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 79-100). Santa Monica, CA: RAND.
- Stone, S., & Engel, M. (2007). Same old, same old? Students' experiences of grade retention under Chicago's ending social promotion policy. *American Journal of Education*, 113, 605-634. <http://dx.doi.org/10.1086/518490>
- Stone, S., & Jacob, R. T. (2005). Teachers and students speak. *Education Next*, 5(1), 49-53.
- Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal*, 29, 199-223. <http://dx.doi.org/10.3102/00028312029001199>
- Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research*, 37, 237-260. http://dx.doi.org/10.1207/s15548430jlr3702_5
- U.S. Department of Education. (1999). *Taking responsibility for ending social promotion: A guide for educators and state and local leaders*. Retrieved from <http://www2.ed.gov/pubs/socialpromotion/index.html>
- Valencia, R. R., & Villarreal, B. J. (2005). Texas' second wave of high-stakes testing: Anti-social promotion legislation, grade retention, and adverse impact on minorities. In A. Valenzuela

- (Ed.), *Leaving children behind: How "Texas-style" accountability fails Latino youth* (pp. 113-152). Albany, NY: State University of New York Press.
- White, E. E. (1886). *The elements of pedagogy*. New York, NY: American Book.
- White, E. E. (1888). Examinations and promotions. *Education*, 8, 517-522.
- White, K. W., & Rosenbaum, J. E. (2008). Inside the black box of accountability: How high-stakes accountability alters school culture and the classification and treatment of students and teachers. In A. Sadnovik, J. A. O'Day, G. W. Bohrnstedt & K. M. Borman (Eds.), *No Child Left Behind and the reduction of the achievement gap: Sociological perspectives on federal educational policy* (pp. 97-114). New York, NY: Routledge.
- Wilson, R. E., Bowers, M. J., & Hyde, R. L. (2011, June 30). *Volume 1 of special investigation into CRCT cheating at APS: Overview, interviews, school summaries*. Retrieved from The Atlanta Journal-Constitution website: <http://www.ajc.com/news/volume-1-of-special-1000798.html>
- Winters, M. A., & Greene, J. P. (2006). Getting ahead by staying behind: An evaluation of Florida's program to end social promotion. *Education Next*, 6(2), 65-69.
- Winters, M. A., & Greene, J. P. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy*, 7, 305-330. http://dx.doi.org/10.1162/EDFP_a_00069
- Wirtz, W. (1977). *On further examination: Report of the advisory panel on the Scholastic Aptitude Test score decline*. New York, NY: College Entrance Examination Board.
- Xia, N., & Kirby, S. N. (2009). *Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes*. (Technical Report No. 678). Santa Monica, CA: Retrieved from RAND Education website: http://www.rand.org/pubs/technical_reports/TR678/

About the Author

Andrew P. Huddleston

Abilene Christian University
andrew.huddleston@acu.edu

Andrew P. Huddleston is an assistant professor of Teacher Education at Abilene Christian University. He received his Ph.D. in Language and Literacy Education from The University of Georgia in 2012. His research interests focus on high-stakes testing policies in reading.

education policy analysis archives

Volume 22 Number 18

April 7th, 2014

ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, Socolar (China).

Please contribute commentaries at <http://epaa.info/wordpress/> and send errata notes to Gustavo E. Fischman fischman@asu.edu

Join EPAA's Facebook community at <https://www.facebook.com/EPAAAPE> and Twitter feed @epaa_aape.

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Audrey Amrein-Beardsley** (Arizona State University), **Rick Mintrop**, (University of California, Berkeley)
Jeanne M. Powers (Arizona State University)

Jessica Allen University of Colorado, Boulder

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Angela Arzubiaga Arizona State University

David C. Berliner Arizona State University

Robert Bickel Marshall University

Henry Braun Boston College

Eric Camburn University of Wisconsin, Madison

Wendy C. Chi* University of Colorado, Boulder

Casey Cobb University of Connecticut

Arnold Danzig Arizona State University

Antonia Darder University of Illinois, Urbana-Champaign

Linda Darling-Hammond Stanford University

Chad d'Entremont Strategies for Children

John Diamond Harvard University

Tara Donahue Learning Point Associates

Sherman Dorn University of South Florida

Christopher Joseph Frey Bowling Green State University

Melissa Lynn Freeman* Adams State College

Amy Garrett Dikkers University of Minnesota

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Harvey Goldstein Bristol University

Jacob P. K. Gross Indiana University

Eric M. Haas WestEd

Kimberly Joy Howard* University of Southern California

Aimee Howley Ohio University

Craig Howley Ohio University

Steve Klees University of Maryland

Jackyung Lee SUNY Buffalo

Christopher Lubienski University of Illinois, Urbana-Champaign

Sarah Lubienski University of Illinois, Urbana-Champaign

Samuel R. Lucas University of California, Berkeley

Maria Martinez-Coslo University of Texas, Arlington

William Mathis University of Colorado, Boulder

Tristan McCowan Institute of Education, London

Heinrich Mintrop University of California, Berkeley

Michele S. Moses University of Colorado, Boulder

Julianne Moss University of Melbourne

Sharon Nichols University of Texas, San Antonio

Noga O'Connor University of Iowa

João Paraskveva University of Massachusetts, Dartmouth

Laurence Parker University of Illinois, Urbana-Champaign

Susan L. Robertson Bristol University

John Rogers University of California, Los Angeles

A. G. Rud Purdue University

Felicia C. Sanders The Pennsylvania State University

Janelle Scott University of California, Berkeley

Kimberly Scott Arizona State University

Dorothy Shipps Baruch College/CUNY

Maria Teresa Tatto Michigan State University

Larisa Warhol University of Connecticut

Cally Waite Social Science Research Council

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Ed Wiley University of Colorado, Boulder

Terrence G. Wiley Arizona State University

John Willinsky Stanford University

Kyo Yamashiro University of California, Los Angeles

* Members of the New Scholars Board

archivos analíticos de políticas educativas
consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

Armando Alcántara Santuario Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

Claudio Almonacid Universidad Metropolitana de Ciencias de la Educación, Chile

Pilar Arnaiz Sánchez Universidad de Murcia, España

Xavier Besalú Costa Universitat de Girona, España

Jose Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

María Caridad García Universidad Católica del Norte, Chile

Raimundo Cuesta Fernández IES Fray Luis de León, España

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Inés Dussel FLACSO, Argentina

Rafael Feito Alonso Universidad Complutense de Madrid, España

Pedro Flores Crespo Universidad Iberoamericana, México

Verónica García Martínez Universidad Juárez Autónoma de Tabasco, México

Francisco F. García Pérez Universidad de Sevilla, España

Edna Luna Serrano Universidad Autónoma de Baja California, México

Alma Maldonado Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

Alejandro Márquez Jiménez Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Felipe Martínez Fernández University of California Los Angeles, USA

Fanni Muñoz Pontificia Universidad Católica de Perú

Imanol Ordorika Instituto de Investigaciones Económicas – UNAM, México

María Cristina Parra Sandoval Universidad de Zulia, Venezuela

Miguel A. Pereyra Universidad de Granada, España

Monica Pini Universidad Nacional de San Martín, Argentina

Paula Razquin UNESCO, Francia

Ignacio Rivas Flores Universidad de Málaga, España

Daniel Schugurensky Universidad de Toronto-Ontario Institute of Studies in Education, Canadá

Orlando Pulido Chaves Universidad Pedagógica Nacional, Colombia

José Gregorio Rodríguez Universidad Nacional de Colombia

Miriam Rodríguez Vargas Universidad Autónoma de Tamaulipas, México

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Luis San Fabián Maroto Universidad de Oviedo, España

Yengny Marisol Silva Laya Universidad Iberoamericana, México

Aida Terrón Bañuelos Universidad de Oviedo, España

Jurjo Torres Santomé Universidad de la Coruña, España

Antoni Verger Planells University of Amsterdam, Holanda

Mario Yapu Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**
(Universidade Federal do Rio Grande do Sul)

Dalila Andrade de Oliveira Universidade Federal de Minas Gerais, Brasil
Paulo Carrano Universidade Federal Fluminense, Brasil
Alicia Maria Catalano de Bonamino Pontifícia Universidade Católica-Rio, Brasil
Fabiana de Amorim Marcello Universidade Luterana do Brasil, Canoas, Brasil
Alexandre Fernandez Vaz Universidade Federal de Santa Catarina, Brasil
Gaudêncio Frigotto Universidade do Estado do Rio de Janeiro, Brasil
Alfredo M Gomes Universidade Federal de Pernambuco, Brasil
Petronilha Beatriz Gonçalves e Silva Universidade Federal de São Carlos, Brasil
Nadja Herman Pontifícia Universidade Católica –Rio Grande do Sul, Brasil
José Machado Pais Instituto de Ciências Sociais da Universidade de Lisboa, Portugal
Wenceslao Machado de Oliveira Jr. Universidade Estadual de Campinas, Brasil

Jefferson Mainardes Universidade Estadual de Ponta Grossa, Brasil
Luciano Mendes de Faria Filho Universidade Federal de Minas Gerais, Brasil
Lia Raquel Moreira Oliveira Universidade do Minho, Portugal
Belmira Oliveira Bueno Universidade de São Paulo, Brasil
Antônio Teodoro Universidade Lusófona, Portugal
Pia L. Wong California State University Sacramento, U.S.A
Sandra Regina Sales Universidade Federal Rural do Rio de Janeiro, Brasil
Elba Siqueira Sá Barreto Fundação Carlos Chagas, Brasil
Manuela Terrasêca Universidade do Porto, Portugal
Robert Verhine Universidade Federal da Bahia, Brasil
Antônio A. S. Zuin Universidade Federal de São Carlos, Brasil