Kept Back to Get Ahead? Kindergarten Retention and Academic Performance

Yingying Dong†

Department of Economics

California State University - Fullerton

Original: August 2006

Revised: April 2009

Abstract: While most existing research concludes that grade retention generates no benefits for the retainees' academic performance, holding low achieving children back has been a popular practice for decades. Drawing on a recently collected nationally representative dataset in the US, this paper estimates the causal effect of repeating kindergarten on the retained children's academic performance. Since we observe children being held back only when they enroll in schools that permit kindergarten retention, this paper jointly models the choice of enrolling in a school that allows kindergarten retention, the decision of repeating kindergarten, and children's academic performance in higher grades. A control function approach is developed to estimate the resulting double hurdle treatment model, which accounts for unobserved heterogeneity in the retention effect. A nearest-neighbor matching estimator is also implemented. Holding children back in kindergarten is found to have positive but diminishing effects on their academic performance up to third grade.

JEL Codes: C31, C35, I21

Keywords: Kindergarten Retention, Retention Policy, Average Treatment Effect on the Treated, Control Function, Nearest Neighbor Matching

---

† I would like to thank Arthur Lewbel, Peter Gottschalk, Shannon Seitz, and Zhijie Xiao for detailed and insightful comments and suggestions. I am also grateful to Donald Cox for helpful suggestions. All errors are my own.

1. Introduction

The practice of having low-performing students repeat a grade has been hotly disputed and heavily studied by educators, psychologists, and sociologists. This practice is usually referred to as grade retention. Most of the existing research shows either negative or insignificant effects of grade retention on children's social-emotional adjustment and academic outcomes (Holmes and Matthews 1984; Holmes 1989; Jackson 1975; Hong and Raudenbush 2005; Meisels and Liaw 1991; Smith and Shepard 1987). Meta-analyses conclude that the cumulative evidence does not support the use of grade retention as an academic intervention (Jimerson 2001; Holmes 1989; Holmes and Matthews 1984; Jackson 1975).

Although existing research does not support the use of grade retention as a way to remedy children's poor academic performance, it has been popular for decades in the US. The popularity of grade retention has even increased due to recent emphases on educational standards and accountability in schools (Hauser, Pager, and Simmons 2004; Jimerson and Kaufman 2003; McCoy and Reynolds 1999). By 1998, at least 10 states in the US had developed explicit policies for ending social promotion (American Federation of Teachers, 1998).[1] Social promotion ended in Chicago and New York City in 1999, and in numerous other cities including Baltimore and Philadelphia in the 1990s. In North Carolina, the retention rate in kindergarten through third grade almost doubled from 1992 to 2002 (Early et al., 2003).[2]

Using data from the US Current Population Survey (CPS), Hauser, Frederick, and

---

[1] Social promotion is the practice of promoting a student to the next grade despite their poor academic performance in order to keep them with their peers.

[2] In the US, kindergarten is a class that is organized to provide educational experiences for children (typically 5 - 6 years old) before they enter first grade. About 98% of children attend kindergarten, though kindergarten attendance is mandatory in some states and optional in others (Kauerz, 2005). Kindergarten is usually physically located within the same institution as elementary school. However, unlike first and higher grades, the purpose of kindergarten is not primarily education, but to introduce children to a school's social environment and acclimate them to all of the activities involved in attending and learning in an institutional setting. As a result, retention in kindergarten is likely to differ both qualitatively and quantitatively from retention in other grades.

Andrew (2007) show that since 1996 there is a clear increasing trend in retention rates, though the increase is only moderate. Based on data gathered from state educational agencies, they also show that the growth in retention was mainly concentrated in kindergarten and early primary grades. Frederick and Hauser (2008) identify increasing levels of retention beginning in the early 1970s and show that the rise in retention is driven in part by kindergarten retentions.

Arguably retention is more likely to have positive effects when applied early because repeating a year hurts less emotionally and may ultimately be more beneficial for young children (Shepard 1989). However, limited existing research on kindergarten retention also suggests that young children get no academic benefits from repeating the program (Niklason 1987, Shepard and Smith 1986a, 1986b, Mantzicopoulos 1989, Mantzicopoulos and Morrison 1992, Mantzicopoulos 1997). Shepard (1989) identifies only one well-controlled study where an academic advantage was found for kindergarten retainees at the end of first grade, though it is not known whether this advantage was maintained beyond first grade.

The discrepancy between educational practice and research findings makes it both interesting and of practical significance to study this issue using quality data and rigorous methods. Comparing directly the academic outcomes of retained children and their promoted peers does not give the causal effect of retention, because the two groups of children are not comparable in observed and possibly unobserved characteristics that can affect both their probabilities of being held back as well as academic performance. Differences in their academic performance might be attributable to any of these confounding factors instead of repeating a grade.

The existing literature on grade retention is mostly based on quasi-experimental designs. Some recent exceptions include Eide and Showalter (2000) and Jacob and Lefgren (2004). Eide and Showalter (2000) adopt linear instrumental variables (IV) estimation to examine the effect of grade retention on the probability of dropping out of high school and on labor market earnings. Their study finds insignificant beneficial effects of grade retention. Jacob and Lefgren (2004) explore a natural experiment, the implementation of an accountability policy in Chicago Public Schools. They adopt a regression-discontinuity approach and show that grade retention has a modest but positive net impact on third-grade students’ achievement scores.

Studies based on quasi-experimental designs feature a direct comparison of the academic performance or social-emotional adjustment between a retained group and a designated control group. A control group is usually constructed based on the similarity of demographic characteristics such as gender, race, age, and sometimes measures of pre-treatment cognitive levels. A control group may also draw on those children who were recommended to repeat kindergarten but whose parents chose not to do so.

As noticed by many researchers, in either case the two groups of children may not be comparable. In the former case, systematic differences may exist in some unobserved (to researchers) characteristics; in the latter, the sample is self-selected by parents, so potential differences may exist in family characteristics (Mantzicopoulos 1997). A retention decision is usually jointly made by schools and parents; as a result, child, school, and family characteristics may all affect a child’s probability of being held back as well as his academic outcomes. Failure to match the retained children with their promoted peers on any of these observed or unobserved dimensions would lead to violation of the unconfoundness assumption required for matching (Rubin 1978; Rosenbaum and Rubin 1984) and hence biased estimates.

A second problem with matching is that it cannot control for age effects, because retained and promoted children cannot be matched on both ages and their grade levels. When the outcome is measured at the same grade level, the estimated effect captures the retention effect and the effect of becoming one year older. As a remedy, some studies conduct same-grade comparisons as well as same-age comparisons (Mantzicopoulos and Morrison 1992). Same-grade comparisons compare retained children to promoted children at the same grade level, while the retained children are one year older than their promoted counterparts. Same-age comparisons compare retained children to promoted children at the same age, while the promoted children are one grade ahead of their retained peers.

In addition, these studies commonly use data from local school districts. Sample sizes are typically small, and school retention policies are entirely ignored. However, in the US not all schools allow children to be held back in kindergarten. Define schools that permit kindergarten retention as “retention schools", and schools that do not as non-retention schools.[3] Since we observe children being held back if and only if they attend retention schools and receive a retention treatment, it is interesting to examine if children are randomly assigned to the two types of schools. If yes, then school assignment would form a natural experiment and so we could compare children across school types to obtain more reliable causal inferences. Otherwise, one has to take into account the non-random selection into different types of schools to estimate the causal effect of kindergarten retention.

A closely related literature examines the effects of school start age—whether children benefit from delayed entry into kindergarten. Recent examples include Fredriksson and Öckert (2005), Fertig and Kluve (2005), Bedard and Dhuey (2006), Deming and Dynarski (2008), and Elder and Lubotsky (2009). A thorough survey of earlier literature on this topic is given by Stipek (2002). Both delayed entry and retention in kindergarten are intended to give children more time to mature. Together these practices shift up the age distribution in schools and cause increasing numbers of overage students in each grade over the past decades. They also share some interesting similarities. For example, boys and younger children are more frequently delayed enrollment or retained in kindergarten. In either case, any possible positive effects are documented to be short lived.[4] Here we focus on kindergarten retention.

Drawing on a nationally representative sample of kindergartners from the US, this paper estimates the causal effect of holding children back in kindergarten on their academic performance in later grades. The reason to focus on kindergarten retention is twofold. First, the decision-making of holding children back in kindergarten is different from that in higher grades. Kindergarten retention targets children who are socially immature or have difficulty acquiring basic academic skills (Mantzicopoulos and Morrison 1992), while upper grade retention is based solely or largely on academic performance. Second, kindergarten retention is administered early on, i.e., before any real failure occurs, and so is more likely to have positive effects.

[3] This definition is based on a school's kindergarten retention policy, not on its higher grade retention policy, which may be different.

[4] Existing studies generally examine these two issues separately; however, given the fact that they are closely related, it might be of interest to investigate the interplay of these two practices. We leave this for future studies.

This study adds to the existing literature in a number of ways. (1) It estimates the causal effect of kindergarten retention on academic performance, using recently collected nationally representative longitudinal data. (2) It considers potentially non-random selection of children into retention schools, and jointly models the school choice, retention decision, and children's academic performance in higher grades. This yields a double-hurdle treatment model, where the retention treatment is a binary choice with sample selection. (3) A control function estimator is derived and used to estimate the resulting model, which accounts for unobserved heterogeneity in the retention effect. (4) A nearest neighbor matching is also applied with different assumptions regarding school selection effects.

The rest of this paper proceeds as follows: Section 2 describes the data and variables. Section 3 sets up the econometric model. Section 4 develops the control function estimator, discusses instrumental variables, and describes the nearest neighbor matching estimator. Section 5 reports empirical results. Section 6 concludes.

2. Data

We use data from the US Early Childhood Longitudinal Study — Kindergarten Cohort 1998-1999 (ECLS-K). It is an ongoing study conducted by the US National Center for Education Statistics (NCES). They began collecting data on a nationally representative sample of children in 1998 when these children were kindergartners.[5] So far, data have been collected on the full sample in the fall and spring of the kindergarten year, and the spring of grades 1, 3 and 5.[6] Data are gathered from direct assessments of children and from interviews with parents, teachers, and school administrators.

The primary advantage of this data set is that it provides test scores that are intended to reveal children's true academic levels and are comparable over time. The ECLS-K

[5] Since the ECLS-K follows a group of kindergarteners, our estimation and conclusion are conditional on attending kindergarten. Given the fact that only about 2% of children do not attend kindergarten in the US, considering selection into kindergarten may not substantially change our conclusions.

[6] This means the time when most students are in their first, third, or fifth grade, while some students may be in a different grade due to repeating or skipping a grade. These repeaters or skippers were assessed at the same time as the majority. No additional assessment was administered when they really were in their first, third, or fifth grade.

gives children two-stage adaptive tests, where a child's first-stage performance is used to determine a second test that is suitable for his ability. Test scores are computed based on the Item Response Theory (IRT), which places children on a continuous ability scale.[7] Compared with the traditional "one-test-fits-all" administrations, adaptive tests have the advantage of minimizing potential flooring and ceiling effects and so can reveal children's true cognitive levels. Moreover, the ECLS-K puts different waves of test scores on the same scale, so they are good for evaluating academic gains over time. Other benefits of this data set include having information on school retention policies and kindergarten enrollment age cutoff dates. As shown later, these variables are important for our modeling and identification.

This paper's analyses focus on children who are either first-time kindergarteners or kindergarten retainees in the 1998 – 1999 school year, and who were assessed in the spring and fall of kindergarten and the spring of their first and third grades. Removing observations with missing values yields a sample of size 8,672, including 8,391 promoted children and 281 retained children.

The outcome variables are reading and math IRT scale scores in first and third grades. For the retained children, these are test scores in their actual first or third grade, not scores when they would have been in first or third grade, had they not been retained in kindergarten. All the test scores are standardized to have mean zero and standard deviation one. The explanatory variables include a variety of child, family, and school characteristics as well as pre-retention (pre-treatment) test scores, i.e., the test scores at the end of the first year of kindergarten (K1).[8] Pre-treatment test scores summarize the

[7] Since not all students take the same second stage tests, IRT uses the pattern of right, wrong and omitted responses to the items actually administered in a test and the difficulty, discriminating ability, and "guess-ability" of each item to evaluate a child's cognitive level.

[8] In the sample, the retained children were assessed when they were already in their second year of kindergarten (K2). Therefore, for the retained children, we use their test scores at the beginning of K2, rather than the test scores at the end of K1 as their pre-treatment test scores. Due to the possibility of children attending summer school or just getting familiar with the test, checking the subsample of children (n=298) who were interviewed both at the end of K1 and at the beginning of K2 shows that their test scores on average increased over summer. Therefore, the pre-treatment test scores for the retainees should be lower than their K2 beginning scores, which means the estimated effect of kindergarten retention should be even larger than what we report.

cognitive or skill accumulation before the retention treatment, including those accumulated before entering kindergarten, which can affect later school achievement (Elder and Lubotsky 2009). A full description of these covariates is provided in the Appendix.

Child characteristics are measured in the base (treatment) year. Family characteristics, such as SES, may change over time and have missing values for some years. To save observations, we use the average of non-missing values. In each equation, school level variables are measured in the same year the dependent variable is measured.

Summary statistics for the key characteristics of retained and promoted children are listed in Table 2.1. Compared with promoted children, retained children are significantly disadvantaged in almost all the observed aspects. In particular, the retained tend to have lower than average math and reading test scores before the retention treatment. They are more likely to be boys and on average about 60 months old at kindergarten entrance, which makes them around 6 months younger than promoted children. Among retained children, 18.5% communicate less well; 5.7% have difficulty hearing speeches; 18.1% are less able in solving problems than their same-age peers; 17.8% are overactive; 28.5% are disabled, and 12.1% receive individualized education. These percentages range from two to over five times those of promoted children. Further, 11% of retained children's parents have less than high school education, and 24.2% of retained children's parents do not expect their children to attend college. These numbers are about twice those of promoted children. Finally, retained children are also more likely to live with a single parent or no parent and be from a low SES family.

Table 2.1 Summary statistics of children's key characteristics, by treatment

	Retained children (n=281) Mean (Std. Dev.)	Promoted children (n=8,391) Mean (Std. Dev.)	Difference
K1 math	-0.839 (0.873)	0.037 (0.990)	-1.194***
K1 reading	-0.756 (0.923)	0.031 (0.993)	-0.983***
White	0.655 (0.476)	0.660 (0.474)	-0.006
Female	0.324 (0.469)	0.514 (0.500)	-0.191***
Age-at-entry	60.43 (4.537)	66.69 (4.022)	-6.25***
Hearing difficulty	0.057 (0.232)	0.020 (0.139)	0.037***
Seeing difficulty	0.068 (0.252)	0.048 (0.214)	0.019
Communication ability: Less well than same-age children	0.185 (0.389)	0.085 (0.279)	0.100***
Overactive	0.178 (0.383)	0.109 (0.312)	0.069***
Problem solving ability: Less well than same-age children	0.181 (0.386)	0.048 (0.214)	0.133***
Disabled	0.285 (0.452)	0.126 (0.331)	0.159***
Individualized Education Plan (IEP)	0.121 (0.327)	0.023 (0.150)	0.098***
Parental educational expectation: High school or less	0.242 (0.429)	0.129 (0.335)	0.113***
SES	-0.009 (0.814)	0.129 (0.745)	-0.138***
Parents' highest education:			
Bachelor's degree or above	0.327 (0.470)	0.383 (0.486)	-0.056
Less than high school	0.110 (0.314)	0.049 (0.216)	0.061***
Family type: Single/no parent	0.206 (0.405)	0.159 (0.366)	0.047***

Note: *** Significant at the 1% level.

3. The Econometric Model

This section sets up an econometric model for the retention treatment and academic outcomes. We observe a child being held back in kindergarten if and only if the child attends a retention school and receives the treatment of repeating kindergarten. As shown by the data, these two decisions are jointly determined; i.e., observed and possibly unobserved family characteristics may affect both. This section therefore models the retention treatment as a binary choice (whether to repeat kindergarten) with sample selection (selection into a retention school). The outcome equation is a linear regression where the retention dummy is an endogenous regressor with a correlated random coefficient, which captures the heterogeneity of treatment effects.

3.1 The Retention Model

Among the 726 schools in our sample, there are 616 retention schools and 110 non-retention schools. Kindergarteners in the two types of schools have different risks of being held back, and the risk in non-retention schools is nearly zero.[9] If children are not randomly assigned to these two types of schools, then appropriate modeling of the retention treatment needs to take into account this non-random selection.

Let S be a binary variable indicating whether a child attends a retention school. Let X_S be a vector of observables and ε_S an unobservable that determine a child's propensity of attending a retention school. Assume that a child attends a retention school if and only if $X_S'\eta + \varepsilon_S \geq 0$, where η is the vector of coefficients. That is,

$$S = I(X_S'\eta + \varepsilon_S > 0), \tag{1}$$

where $I(\cdot)$ is an indicator function that is 1 if the bracketed expression is true, and 0 otherwise.

Let D^* be the potential kindergarten retention status for a child, regardless of the type of school he is in. By definition, D^* is observed only if the child attends a retention school. Let X_D be a vector of observables and ε_D an unobservable that determine a child's propensity of being held back in kindergarten. Assume that a child potentially belongs to the retained group if and only if $X_D'\delta + \varepsilon_D \geq 0$, where δ is the vector of coefficients. That is, we have

$$D^* = I(X_D'\delta + \varepsilon_D \geq 0). \tag{2}$$

Assume that ε_S and ε_D may be correlated with each other, but they are independent of X_s and X_D; i.e., X_s and X_D are exogenous in equations (1) and (2).

Further, denote the observed retention status as D. It follows that

$$D = S \cdot D^* = S \cdot I(X_D'\delta + \varepsilon_D \geq 0). \tag{3}$$

For clarity, we refer to equation (1) as the school selection equation, equation (2) as the

[9] In practice, children in non-retention school may repeat kindergarten if they change schools. Limited by our data, we do not consider this case in our analysis, i.e., we assume that children are at-risk only when they attend retention schools.

retention (treatment) equation, and both together along with equation (3) as the retention model.

If ε_S and ε_D are uncorrelated; i.e., children are not selected into retention schools due to their unobserved (to econometricians) characteristics that are correlated with D^*, kindergarten retention policies then generate exogenous variation in D. Equation (3) could therefore be estimated independently using data from retention school children. However, if ε_S and ε_D are correlated, estimation then needs to account for this joint determination of school selection and retention decisions. Ignoring non-random selection into retention schools would lead to biased estimates. Note that joint estimation of equations (1) and (3) does not rule out the case of random school selection. In fact, it nests random school selection as the special case in which the correlation between ε_S and ε_D is zero.

To investigate if joint estimation is necessary, we compare retention school children with non-retention school children. It appears that children are not randomly selected into retention schools. For example, Figures 3.1 and 3.2 show how the probabilities of attending non-retention schools are different for children from different family backgrounds. As one can see, children from low SES families, living with single parent or no parent or with less educated parents are more likely to attend non-retention schools. We also test the similarity of these family characteristics between retained and promoted children. The results are reported in Table 3.1, where all the mean values are percentages except that of SES. Table 3.1 shows that children in retention schools differ significantly from children in non-retention schools in terms of family characteristics, such as family SES, parents' education, and single or no parent family type.

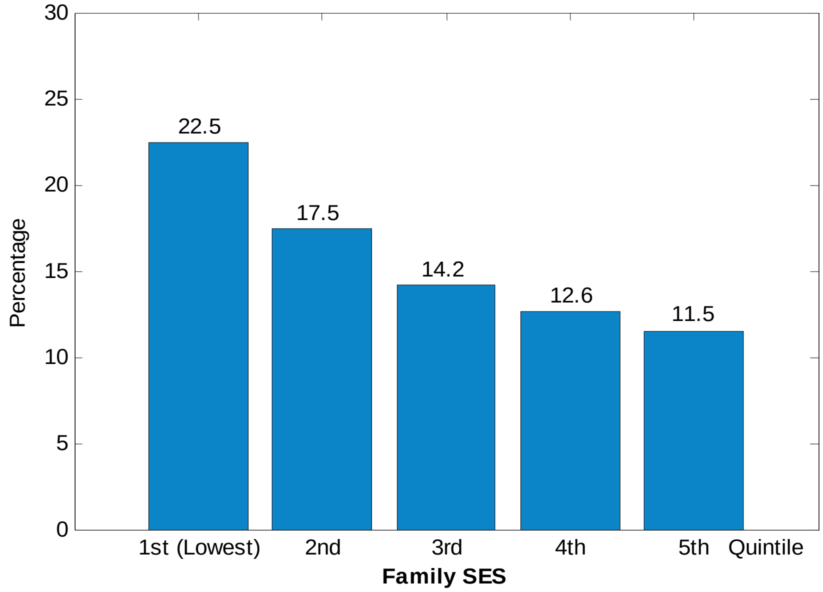

Figure 3.1: Percentage of kindergarteners attending non-retention schools, by family SES

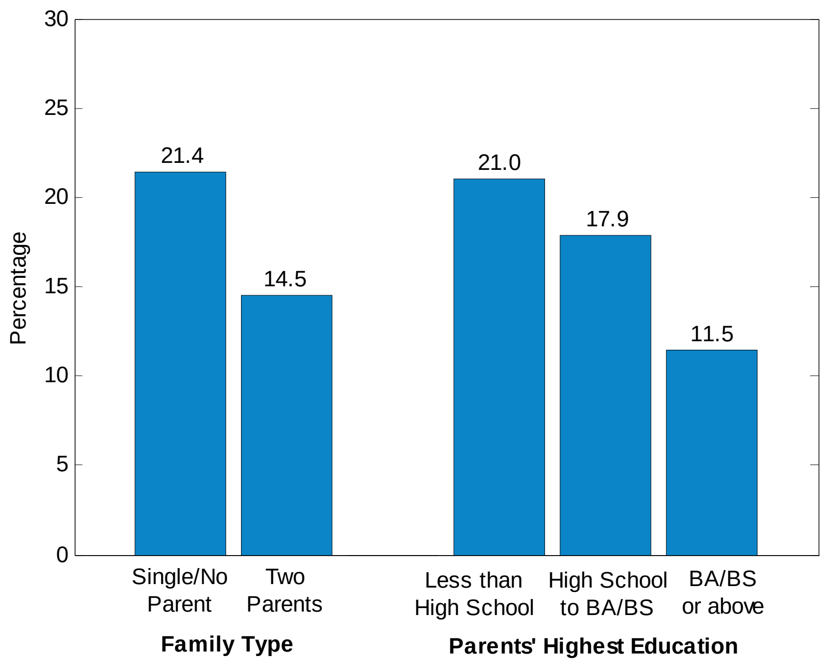

Figure 3.2: Percentage of kindergarteners attending non-retention schools, by family type and parents' education

Table 3.1 Summary statistics of family characteristics, by retention policy

	In non-retention school (n=1,357)	In retention school (n=7,315)	Difference
SES	-0.049 (0.764)	0.156 (0.740)	-0.202***
Parents' highest education: Less than high school	0.069 (0.213)	0.048 (0.253)	0.021***
High school or above, but less than a bachelor degree	0.552 (0.497)	0.650 (0.477)	0.098***
Bachelor degree or above	0.282 (0.450)	0.400 (0.490)	-0.118***
Family type: Single/no parent	0.220 (0.414)	0.149 (0.356)	0.071***

Note: *** Significant at the 1% level.

Given the differences in these observed family characteristics, it is likely that the two groups of children also differ in unobserved family characteristics that affect school selection and retention decisions as well as outcomes. For example, parents' preferences or willingness to invest in their children's education affects the children's chances of attending retention schools, probabilities of repeating kindergarten, and their academic performance. In particular, one may expect that parents who are more involved are more likely to enroll their children in retention schools, and high parental involvement may also reduce these children's risk of being held back. Therefore, equations (1) and (3) should be jointly estimated.

3.2 The Test Score Equation

If one could observe each child's test scores when he is retained in kindergarten and when he is not, then the average treatment effect (ATE) of kindergarten retention would be simply the mean difference of test scores for all children in the two states of the world. In this case the average treatment effect on the treated (ATT), i.e., the average retention effect of retained children, would be given by the mean difference of the retained children's test scores in the two states of the world. ATT is the effect of interest in the current context, so we focus on estimating this effect.

Denote a child's potential test scores with and without being held back as Y_1 and Y_0, respectively, regardless of his actual retention status. The ATE is then given by $E(Y_1) - E(Y_0)$, and the ATT is given by $E(Y_1 | D=1) - E(Y_0 | D=1)$. However, for each child, we can only observe his test score either when he is retained or when he is not. Given the potential outcomes Y_1 and Y_0, the observed test score Y can be written as $DY_1 + (1-D)Y_0$. Assuming that the conditional expectations of Y_1 and Y_0 are given by $X_Y{'}\beta_1$ and $X_Y{'}\beta_0$, respectively, where X_Y is a vector of observed child, family, and school characteristics, including a constant term, we then have

$$Y_1 = X_Y{'}\beta_1 + \varepsilon_{Y1}, \tag{4}$$

$$Y_0 = X_Y{'}\beta_0 + \varepsilon_{Y0}, \tag{5}$$

where ε_{Y1} and ε_{Y0} are mean zero error terms that are independent of X_Y. Further assume that the returns to observed characteristics are the same in the two states of the world; i.e., β_1 and β_0 differ only in constant terms. Denoting the difference as γ, we can then write[10]

$$Y = X_Y'\beta_0 + \gamma D + D\varepsilon_{Y1} + (1-D)\varepsilon_{Y0}, \tag{6}$$

which can be rewritten as

$$Y = X_Y'\beta_0 + (\gamma + \varepsilon_{Y1} - \varepsilon_{Y0})D + \varepsilon_{Y0}. \tag{7}$$

In the above equation γ represents the ATE, and $\varepsilon_{Y1} - \varepsilon_{Y0}$ represents the heterogeneity of treatment effects. $E(\varepsilon_{Y1} - \varepsilon_{Y0} \mid D=1)$ is non-zero as long as $E(\varepsilon_{Y1} \mid D=1) \neq E(\varepsilon_{Y0} \mid D=1)$; i.e., the returns to children's unobserved characteristics differ in the two states of the world, so children can be selected into retention due to higher gains from the treatment. The ATT, $\gamma + E(\varepsilon_{Y1} - \varepsilon_{Y0} \mid D=1),$ is different from the ATE in this case.

4. The Estimation of ATT

4.1 The Control Function Estimator

Given the retention model and the outcome equation developed in the previous section, the full model can be written as

$$S = I(X_S'\eta + \varepsilon_S > 0), \tag{8}$$

$$D^* = I(X_D'\delta + \varepsilon_D \geq 0), \tag{9}$$

[10] Note that this assumption is imposed for simplicity. It is not an identification assumption. In fact, our estimation applies to the general model without this assumption.

$$D = S \cdot D^*, \tag{10}$$

$$Y = X_Y'\beta_0 + \gamma D + D\varepsilon_{Y1} + (1-D)\varepsilon_{Y0}, \tag{11}$$

By this structure, we allow the kindergarten retention policy to have direct and indirect effects on retention probabilities as well as an indirect selection effect on test scores.

Since common unobservables may affect a child's school selection, probability of being held back, and academic performance, the four error terms in the above model are correlated with each other, which means that D is endogenous in equation (11). Therefore, OLS is biased, and linear IV can not consistently estimate equation (11) unless $\varepsilon_{Y1} = \varepsilon_{Y0}$ is assumed,[11] which would imply absence of heterogeneity in the treatment effect, and so rule out individuals' self-selection into the retention treatment based on higher returns, i.e., higher values of $E(\varepsilon_{Y1} - \varepsilon_{Y0} \mid D = 1)$. Details about this restrictive assumption and its implications can be found in Heckman (1997). We focus on the more general case without this restriction and allow the treatment effect to be heterogeneous across individuals.

The above model can be estimated by a control function (CF) approach. This involves plugging into the model of interest (the test score equation in our case) one or more bias correction terms. One way to construct a bias correction term is to use the conditional mean of the error term, conditional on all the covariates. This approach has been proposed and applied to the standard treatment model, i.e., a linear regression with an endogenous dummy regressor that is specified as a Probit (Vella and Verbeek, 1999). Here we derive a control function estimator for this paper's more general model.

---

[11] When $\varepsilon_{Y1} \neq \varepsilon_{Y0}$, under a monotonicity assumption about the instrumental variable, some form of IV estimation can yield the local average treatment effect (LATE) and the marginal treatment effect (MTE). The former is the treatment effect on compliers, i.e., those who are retained because of the exogenous variation caused by the instrument, while the latter is the limit of LATE for an infinitely small change in the value of the instrument. Both differ from ATT. Discussion about the necessary condition for identifying ATT can be found in Manning (2004) and the references therein.

Assume that the error terms in equations (8) - (11) have a joint normal distribution,[12] and denote the covariance for each pair of error terms $(\varepsilon_D, \varepsilon_S)$, $(\varepsilon_{Y1}, \varepsilon_S)$, $(\varepsilon_{Y0}, \varepsilon_S)$, $(\varepsilon_{Y1}, \varepsilon_D)$, and $(\varepsilon_{Y0}, \varepsilon_D)$ as σ_{SD}, σ_{1S}, σ_{0S} σ_{1D}, and σ_{0D}, respectively. Given the vector of all covariates $X = (X_Y', X_D', X_S')'$, the conditional mean of the structural error term in equation (11) can be written as

$$\begin{aligned} & E(D\varepsilon_{Y1} + (1-D)\varepsilon_{Y0} \mid S, D, X) \\ & = SD \cdot E(\varepsilon_{Y1} \mid S=1, D=1, X) \\ & + S(1-D) \cdot E(\varepsilon_{Y0} \mid S=1, D=0, X) \\ & + (1-S) \cdot E(\varepsilon_{Y0} \mid S=0, X). \end{aligned} \tag{12}$$

Under the normality assumption, the three conditional expectations on the right-hand side of equation (12) are given by

$$\begin{aligned} & E(\varepsilon_{Y1} \mid S=1, D=1, X) \\ & = E(\varepsilon_{Y1} \mid S=1, D^*=1, X) \\ & = E(\varepsilon_{Y1} \mid \varepsilon_S > -X_S'\eta, \varepsilon_D > -X_D'\delta, X) \\ & = \frac{\sigma_{1S} - \sigma_{1D}\sigma_{SD}}{1-\sigma_{SD}^2} \frac{\phi(X_S'\eta)}{\Phi(X_S'\eta)} + \frac{\sigma_{1D} - \sigma_{1S}\sigma_{SD}}{1-\sigma_{SD}^2} \frac{\phi(X_D'\delta)}{\Phi(X_D'\delta)}, \end{aligned} \tag{13}$$

$$\begin{aligned} & E(\varepsilon_{Y0} \mid S=1, D=0, X) \\ & = E(\varepsilon_{Y0} \mid S=1, D^*=0, X) \\ & = E(\varepsilon_{Y0} \mid \varepsilon_S > -X_S'\eta, \varepsilon_D < -X_D'\delta, X) \\ & = \frac{\sigma_{0S} - \sigma_{0D}\sigma_{SD}}{1-\sigma_{SD}^2} \frac{\phi(X_S'\eta)}{\Phi(X_S'\eta)} + \frac{\sigma_{0D} - \sigma_{0S}\sigma_{SD}}{1-\sigma_{SD}^2} \frac{-\phi(X_D'\delta)}{1-\Phi(X_D'\delta)}, \end{aligned} \tag{14}$$

and

$$\begin{aligned} & E(\varepsilon_{Y0} \mid S=0, X) \\ & = E(\varepsilon_{Y0} \mid \varepsilon_S < -X_S'\eta, X) = \sigma_{0S} \frac{-\phi(X_S'\eta)}{1-\Phi(X_S'\eta)}, \end{aligned} \tag{15}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the probability and cumulative density functions of the standard normal distribution, respectively. As usual, we have normalized the variance of ε_S and ε_D to one.

[12] Test scores are bounded, so formally the errors in the test score equations cannot be normal. However, the tails of the test score distributions resemble normal tails, for example, the kurtoses range from 2.7 to 3.1. Also, as will be discussed later, the CF estimator permits some forms of non-normality in the errors.

Denote $SD\dfrac{\phi(X_S'\eta)}{\Phi(X_S'\eta)}$, $SD\dfrac{\phi(X_D'\delta)}{\Phi(X_D'\delta)}$, $S(1-D)\dfrac{\phi(X_S'\eta)}{\Phi(X_S'\eta)}$, $S(1-D)\dfrac{-\phi(X_D'\delta)}{1-\Phi(X_D'\delta)}$, and $(1-S)\dfrac{-\phi(X_S'\eta)}{1-\Phi(X_S'\eta)}$ as $\lambda_k\,(k=1,2,...,5)$. Plugging these terms into equation (11) yields the augmented equation

$$Y = X_Y'\beta + \gamma D + \sum_{k=1}^{5}\rho_k\lambda_k + e, \tag{16}$$

where $\rho_k\,(k=1,2,...,5)$ are the coefficients of the bias correction terms, and e is the error term in the augmented equation. It is easy to show that e is uncorrelated with the covariates in equation (16), so OLS can be used to estimate this equation. That is, the full model can be estimated by a two-step procedure: First, jointly estimate equations (8) and (10) to get fitted values for $\lambda_k\,(k=1,2,...,5)$, and then plug these fitted values into equation (16) and estimate it by OLS.

In the previous section we show that the ATT is given by $\gamma + E(\varepsilon_{Y1}-\varepsilon_{Y0}\mid D=1)$. The second term can be rewritten as

$$\begin{aligned}
&E(\varepsilon_{Y1}-\varepsilon_{Y0}\mid D=1)\\
&= E_{X|D=1}\left[E(\varepsilon_{Y1}-\varepsilon_{Y0}\mid D=1, X)\right]\\
&= E_{X|D=1}\left[E(\varepsilon_{Y1}-\varepsilon_{Y0}\mid S=1, D^*=1, X)\right]\\
&= E_{X|D=1}\left[E\left(\varepsilon_{Y1}-\varepsilon_{Y0}\,\middle|\,\varepsilon_S > -X_S{}'\eta, \varepsilon_D > -X_D'\delta, X\right)\right]\\
&= E_{X|D=1}\left[\left(\frac{\sigma_{1S}-\sigma_{1D}\sigma_{SD}}{1-\sigma_{SD}{}^2}-\frac{\sigma_{0S}-\sigma_{0D}\sigma_{SD}}{1-\sigma_{SD}{}^2}\right)\frac{\phi(X_S'\eta)}{\Phi(X_S'\eta)}\right]\\
&\quad + E_{X|D=1}\left[\left(\frac{\sigma_{1D}-\sigma_{1S}\sigma_{SD}}{1-\sigma_{SD}{}^2}-\frac{\sigma_{0D}-\sigma_{0S}\sigma_{SD}}{1-\sigma_{SD}{}^2}\right)\frac{\phi(X_D'\delta)}{\Phi(X_D'\delta)}\right]\\
&= E_{X|D=1}\left[(\rho_1-\rho_3)\lambda_1+(\rho_2-\rho_4)\lambda_2\right],
\end{aligned}$$

where $E_{X|D=1}[\cdot]=E_X[\cdot\mid D=1]$, i.e., the conditional expectation over X, conditional on $D=1$. Since the first step gives estimates for λ_1 and λ_2, and the second step yields estimates for ρ_1, ρ_2, ρ_3, and ρ_4, the above conditional expectation can be estimated using the empirical expectation of these estimated terms over the retained children. The estimated ATT is then given by this conditional expectation plus the estimated γ. As with other two-step estimators, the correct standard error can be obtained by bootstrapping.

Given our distribution assumption, the full model could also be estimated using joint maximum likelihood estimation (MLE). The two methods are both consistent, but the two-step CF estimator is computationally more tractable. Moreover, the CF estimator remains valid where the joint MLE may not. For example, MLE requires the distribution of the outcome error to be fully specified, while the CF estimator permits the outcome error to equal a specified error (e.g. normal) plus another independent error that has an unknown distribution; the normal error may capture the common unobservables in the model, and the independent error may represent additional pure noise.

4.2 The Retention Model and Instrumental Variables

Although the CF approach described in the previous section is valid without any instrumental variables, in general one would like some instruments in the school selection and retention equations to avoid identification purely based on functional forms. In the current context, a potential instrument for the retention equation is the kindergarten enrollment age cutoff date, and possible instruments for the school selection equation are a school's policy regarding whether children can be retained multiple times in elementary years and a set of dummies indicating how children travel to school. Different school travel modes include walking or riding a bike (the default category), riding a bus, being dropped off by a parent, being dropped off by a day care provider, and other means. These instruments are discussed in detail below.

The kindergarten enrollment age cutoff date (the cutoff date hereafter) refers to the date by which children must reach the age of five to be eligible for kindergarten enrollment. Different states in the US have different cutoff dates. A full list is presented in the Appendix. These dates vary widely across states. They may also vary within a state for two reasons: first, private and charter schools can have cutoff dates that differ from the state requirement; second, some states allow the local education agencies (LEA's) to set their own cutoff dates. The ECLS-K dataset has information on cutoff dates at the school level, so we can exploit the exogenous variation in cutoff dates across schools.

As mentioned, child, family, and school characteristics can all affect a child's probability of being held back. Relevant child characteristics include a child's absolute age and relative age. Absolute age is a child's chronological age, and relative age refers to a child's age relative to the ages of his classmates. To a large extent, absolute age determines a child's maturity and readiness for learning, and hence also affects his social-emotional adjustment and academic performance.

Further, it is well documented that relatively young children are more likely to be held back, *ceteris paribus* (Mantzicopoulos et al., 1989; Collins and Brick 1993; McArthur and Bianchi 1993; Zill, Loomis and West 1997; McEwan and Shapiro 2006; and Elder and Lubotsky 2009). For a particular child, given his age at kindergarten entry (age-at-entry), the later the cutoff date is, the younger the reference group will be, and hence the smaller the child's risk of being held back. For example, assuming that birth dates are uniformly distributed over the year, the average age of kindergarteners in a school with Sept. 1 as the cutoff date will be three months older than those in a school with Dec. 1 as the cutoff date.[13] That is, the cutoff date can affect a child's probability of being held back by shifting the age distribution of his class.

Since relative age is a function of absolute age and the cutoff date, the reduced-form retention equation should include age-at-entry, the cutoff date and possibly their higher order and interacted terms as covariates. Whether higher order or interacted terms need to be included depends on the nonlinearity of the retention probability in absolute and relative ages. Based on tests and comparisons of alternative specifications,[14] we adopt a retention equation that includes absolute age, absolute age squared, and the cutoff date as covariates.

The cutoff date should not have a direct impact on test scores unless the curriculum taught in the classroom is affected by the age of one's peers in a particular grade. If this is true, the cutoff date may affect test scores by changing the age distribution of the class, and so may not be a valid instrument. To investigate this possibility, we tentatively include the cutoff date as a regressor in all four test score equations. None of its coefficients are significant at the 5% level. This provides evidence that conditional on the

[13] Given the typical school start date, Sept. 1, the average age of kindergarteners in the former case is about 66 months; whereas the average age of kindergarteners in the latter case is about 63 months. The enrollment cutoff date extracts the exogenous variation in the average age of kindergarten enrollees.

[14] Our evaluation criteria are the model's pseudo R^2, percentage predicted correctly and the significance of coefficients. When we add other higher order and interacted terms, the pseudo R^2 and the percentage predicted correctly do not change much and their coefficients are not statistically significant.

covariates, the cutoff date does not have additional explanatory power for test scores. Therefore, this variable can be appropriately excluded from those equations.

For the school selection equation, one of the proposed instruments is a dummy indicating multiple retention policies in elementary years.[15] We expect that schools that allow multiple retentions in elementary years are more likely to allow kindergarten retention, because retention policies more or less reflect a school's educational philosophy. This is supported by the following fact: Among schools that permit multiple retentions in elementary years, about 95% allow kindergarten retention, in contrast to 82.8% among schools that do not.

Further, instrument validity requires that this multiple retention policy dummy affects the probability of repeating kindergarten only through its effect on school's kindergarten retention policies. Our data show that among retention schools, those permitting multiple retentions in higher grades have a similar retention rate in kindergarten to those not permitting. The former cell mean is 3.9%, and the latter is 3.8%. The difference is not statistically significant at any conventional level. Therefore, conditional on kindergarten retention policy, the policy of multiple retentions in higher grades does not seem to have a significant impact on the probability of repeating kindergarten, and so we exclude it from the retention equation.

We also check if the multiple retention policy dummy itself has a direct effect on children's academic performance. When this variable is included as an additional covariate in all four test score equations, none of the coefficients are significant at the 5% level. Therefore, we assume that this multiple retention policy dummy is exogenous to the test score equation.

Another set of potential instruments for the school selection equation are the school travel mode dummies. We expect that different travel modes are correlated with the distance from home to school or how convenient it is to get to school, and so they may affect school choices exogenously. For example, parents that otherwise prefer retention schools may enroll their children in a non-retention school if it is conveniently located.

Admittedly, both the multiple retention policy and transportation mode instruments may be flawed, but for different reasons. Having multiple instruments allows us to do

[15] Information is missing on the multiple retention policy for about 7.5% of the full sample. To save observations, we use a dummy to indicate missing values instead of dropping them.

robustness checks: If either one is not a valid instrument, estimation with either one or both would not be expected to produce similar results. As will be shown, the estimation results are robust to the use of different instrumental variables.

4.3 Nearest Neighbor Matching

To compare with the CF approach, a bias-corrected nearest neighbor matching (NNM) as proposed by Abadie and Imbens (2002) is performed. Matching methods assume unconfoundedness; i.e., conditional on observed characteristics, individuals' potential outcomes are independent of the assignment of treatment. Therefore, matching estimators only balance observed, but not unobserved differences between treated and control units. Selection on unobservables is likely to be important here, so unconfoundedness is a potential issue for NNM. Comparing results from the CF and NNM may shed light on the role of unobservables in selection into different types of schools and the retention treatment.

Further, matching estimators rely on a common support assumption, which requires substantial overlap in the observed covariate distributions of the treated and control groups. Table 4.1 shows observed ranges and standard deviations of the continuous covariates for both. The common support assumption appears to hold well for pretreatment math and reading test scores and family SES, but not for age at kindergarten entry. Because age is likely to be an important determinant for academic performance, this possible violation of the common support assumption is another potential issue for NNM in this paper's setting.

Table 4.1 Observed ranges of the continuous covariates, by treatment

	Retained children (n=281)			Promoted children (n=8,391)		
	Min.	Max.	Std. Dev.	Min.	Max.	Std. Dev.
K1 math	-2.267	4.323	0.873	-2.255	5.555	0.991
K1 reading	-1.965	4.640	0.923	-1.887	6.470	0.993
Age-at-entry	46.27	69.03	4.537	57.03	80.90	4.022
SES	-2.800	2.573	0.814	-2.900	2.737	0.745

NNM imputes an individual's counterfactual outcomes using a weighted average of outcomes from individuals with similar characteristics (in the nearest neighbor sense) but opposite treatments. Since each individual can be characterized by a *k* vector of covariates, similarity is defined based on a measure of the distance between two vectors. When the dimension of the covariate vector is high, matching could be very inexact, and serious biases may result. To reduce such biases, NNM conducts linear smoothing.

The ECLS-K data has extensive information at the child, family, and school level, and so provides a rich set of control variables. If all the assumptions required for matching hold, NNM would asymptotically eliminate the biases resulting from inexact matching with a large set of covariates. In our case, retained children are matched to their promoted peers on the full set of covariates except for the instrumental variables.[16] In addition, although there is no rule of thumb for the optimal number of control units matched to one treated unit, generally the number should increase with the sample size. Given the large pool of controls (non-retainees), we set the number to 15.[17]

5. The Empirical Results

5.1 Risk factors for kindergarten retention

The full estimation results for the retention model are reported in Tables A3-(1) – A3-(3) in the Appendix. These three tables provide estimates with different instrumental variables in the school selection equation, namely, the multiple retention policy in elementary years, school travel modes, or both.

Typically, in discrete choice models when one of the alternatives is very overrepresented, identification becomes quite difficult. Given the small proportion of the retained and the relatively small proportion of non-retention schools, one may worry about the identification of the retention model, especially the retention equation. However, this equation appears to be well identified and most of the explanatory variables are significantly different from zero. One important source of identification is

[16] Age-at-entry is matched on instead of age-at-test., because it is more appropriate to control for children's pre-treatment characteristics, while age-at-test is a variable affected by the retention treatment. This is also because matching on age-at-test would result in a sample of promoted children being about one year older than the retained children at kindergarten entry. The average age of the retained children in our sample is 60.4 months, which means the average age for the promoted children would be 72.4.

[17] We tried different numbers between 10 and 20. The results remain stable.

the big difference in the distributions (particularly the means) of covariates between the promoted and retained and between the retention school and non-retention school children (See Tables 2.1 and 3.1).[18]

Compared with school travel modes, the multiple retention policy in higher grades seems to be a stronger instrument. In particular, it yields a higher Pseudo R^2 (0.121 vs. 0.043) and generally more significant coefficients for the school selection equation. However, using either instrument or both, the coefficients of the instrumental variables in the selection equation are statistically significant. All the estimated coefficients, especially those in the retention equation, are similar across specifications.

For the rest of the discussion, we focus on estimation results when both sets of instrumental variables are used. As we can see, the signs of all the estimated coefficients are plausible. In particular, the coefficient of the multiple retention policy dummy is significantly positive, which is consistent with our expectation. The coefficients of school travel modes are also jointly significant with a *P*-value less than 0.001. The estimated coefficients of all four travel mode dummies are positive; i.e., children who walk or ride a bike to school (the default category) are more likely to attend non-retention schools, which could be choices of convenience. In addition, younger children, children with seeing difficulty, being overactive, or children from single/no parent families are more likely to attend non-retention schools. Finally, all-day schools, private schools, or schools receiving federal Title I funds are more likely to be non-retention schools.

Also as expected, the cutoff date in the retention equation has a significantly negative coefficient; i.e., holding all the other covariates constant, the later the cutoff date, the smaller the chance of being held back in kindergarten. Negative coefficients of pretreatment test scores indicate that the lower the test scores in the first year of kindergarten, the higher the probability of being retained. Low math scores are more strongly correlated with repeating kindergarten than low reading scores. Some other child level risk factors, such as being young when entering kindergarten, being disabled, having lower problem solving ability than same-age children, receiving individualized education,[19] and low parental educational expectation can also increase children's

[18] The author would like to thank an anonymous referee for pointing this out.

[19] Receiving individualized education could be endogenous; however, treating this variable as endogenous does not affect our conclusions in this section. In particular, we re-estimate our model, excluding this

probability of being held back. However, hearing difficulty, seeing difficulty, low communication ability, and being overactive have no significant effect on children's probability of being held back in kindergarten. Family risk factors, such as living with a single parent or no parent and parents' highest education being less than high school, are also positively correlated with the probability of being held back.

Very interestingly, the coefficient of SES has a positive sign.[20] That is, conditional on the other covariates, including pretreatment test scores, higher family SES is likely to increase a child's probability of repeating kindergarten. One possible explanation could be that an extra year of kindergarten is more affordable, or the opportunity cost is lower for high SES families. Note that the relationship is negative without conditioning on pretreatment test scores, which implies that family SES is highly correlated with children's academic performance, and children from higher SES families tend to perform better and hence are less likely to be held back. Whereas when they perform equally poorly, children from the high SES families are more likely to accept an extra year of kindergarten.

As shown in the last row of Table A3-(1), the correlation of the errors in the school selection and retention equations is -0.275. A Wald test rejects the null hypothesis that the two errors are independent. Therefore, unobservables (e.g., low parents' involvement or willingness to invest in their children's education) that make children more likely to attend non-retention schools also make them more likely to be held back if there is an exogenous change in their school policies regarding kindergarten retention. Estimation with just the multiple retention policy instrument produces a similar correlation, while using just the travel mode instruments also yields a negative but insignificant correlation. Considering the travel mode instruments are relatively weak, the correlation of the two latent errors could be imprecisely estimated in this case.

variable. The estimates are close to the current ones. As will be shown, the ATT estimates also do not change much. This is probably because only about 2.6% of children have individualized education.

[20] Although this coefficient is not significant, the coefficients for SES and parents' highest education are jointly significant at the 1% level. Since parents' highest education is one component of SES, joint significance should be considered.

5.2 Estimated Treatment Effects

The estimated average retention effects on the retained kindergarteners' academic performance are reported in Tables 5.1 and 5.2. Table 5.1 summarizes the CF estimates. Table 5.2 presents the NNM estimates. The following discusses them both in detail.

The first column of Table 5.1 allows selection into retention schools to be non-random; i.e., the kindergarten retention policy can be endogenous, so the CF estimation proposed in the previous section is applied to the full model; the second column assumes selection into retention schools is random, and so the retention policy is exogenous. The corresponding CF estimation sets to zero σ_{SD}, σ_{0S}, and σ_{1S}, which are the covariances between the school selection equation error and the errors in both the retention equation and the test score equation. Also reported in Table 5.1 are two sets of OLS estimates. The first controls for pre-treatment test scores, while the second does not. OLS corresponds to the specification that sets to zero all the error correlations across equations.

Column 1 in Table 5.1 shows that retained children on average scored 0.552 standard deviations higher in the first-grade reading, and 0.547 standard deviations higher in the first-grade math than they would have scored, had they been socially promoted. These represent 14.5% and 13.9% increases, respectively. These gains in reading and math scores decrease to 0.176 standard deviations and 0.464 standard deviations, respectively, when the kindergarten retainees are in third grade, which corresponds to increases of 2.8% and 8.5%. The effect on the third-grade reading score is not statistically significant at any conventional level. The full estimation results for test score equations are provided in the Appendix.

The CF estimation in column 2, which assumes selection into retention schools is random, generally shows larger effects, particularly in the third grade test scores. Since retention school children may have advantages over non-retention school children in terms of unobserved parental characteristics, such as parents' involvement and preference, failure to balance these unobserved differences could result in over-estimates of the retention effect as seen in column 2. Further, the impact of unobserved parents' characteristics may accumulate over time, which may be the reason why the differences in third grade test scores are more significant.

Table 5.1 The Average Retention Effects on the Retained Kindergarteners' Academic Performance by CF and OLS

	CF (1)	CF (2)	OLS (1)	OLS (2)
	Retention policy endogenous	Retention policy exogenous	With pre-treatment score	Without pre-treatment score
1st grade reading	0.552***	0.610***	0.399***	-0.259***
	(0.129)	(0.130)	(0.043)	(0.058)
3rd grade reading	0.176	0.277	0.256***	-0.247***
	(0.178)	(0.181)	(0.056)	(-0.223)
1st grade math	0.547***	0.526***	0.487***	-0.223***
	(0.134)	(0.140)	(0.049)	(0.061)
3rd grade math	0.464***	0.555***	0.399***	-0.222***
	(0.196)	(0.200)	(0.052)	(0.060)

Note: Bootstrapped standard errors are in parentheses; *** Significant at the 1% level.

Table 5.2 The Average Retention Effects on the Retained Kindergarteners' Academic Performance by NNM

	NNM (1)	NNM (2)	NNM (3)
	Retention policy endogenous	Retention policy exogenous	Retention policy ignored
Source of matched controls	Retention school children	Non-retention school children	All school children
1st grade reading	0.264***	0.337***	0.265***
	(0.041)	(0.046)	(0.041)
3rd grade reading	0.060	0.176***	0.080
	(0.054)	(0.063)	(0.054)
1st grade math	0.473***	0.470***	0.472***
	(0.044)	(0.050)	(0.044)
3rd grade math	0.162***	0.236***	0.186***
	(0.052)	(0.059)	(0.052)

Note: Bootstrapped standard errors are in parentheses; *** Significant at the 1% level.

Columns 3 and 4 in Table 5.1 report OLS estimates. OLS without controlling for pre-treatment test scores produces the widely documented negative sign of the treatment effect. Adding pre-treatment test scores as regressors to the OLS switches the sign to positive. However, the estimated treatment effects are mostly still smaller than the CF estimates that account for selection effects.

Table 5.2 provides NNM estimates assuming selection into retention schools is either non-random or random (column 1 and 2). It also presents NNM estimates when the retention policy is ignored (column 3). When selection into retention schools is assumed non-random, NNM is conducted within retention schools; i.e., retained children are matched to their promoted peers only in retention schools. This within school type matching can eliminate the selection bias if the selection effect of the kindergarten retention policy is not heterogeneous; otherwise, it may still reduce the selection bias.

When selection into retention schools is assumed random, the exogenous change in the kindergarten retention policy serves as a natural experiment. NNM in this case is conducted across school types; i.e., the retained children in retention schools are matched to promoted children in non-retention schools.

When selection into different schools is completely ignored, the retained children are matched with promoted children from all schools. This corresponds to the case where school type information is not observed.

Compared with the CF estimates in Table 5.1, the NNM estimates in Table 5.2 show smaller effects. Recall we match children on their ages at kindergarten entry, not their ages at the time of tests, so retained children are about one year older than their promoted matches. Considering that age may have a nonnegative effect on test scores, if it were possible to account for this age effect, the estimated retention effects by matching should be even smaller than those reported in Table 5.2. This further supports our conjecture that matching fails to balance relevant unobserved differences between retained and promoted children and so tends to underestimate retention effects.

As discussed earlier with CF, comparing columns 1 and 2 in Table 5.2 also shows that when selection into different schools is assumed random NNM generates larger effects than when it is assumed non-random. Not surprisingly, in column 3 when selection into different schools is ignored, we observe that matching over the pooled sample yields retention effects that are larger than those from within school type matching but smaller than those from between school type matching.

Although the estimated retention effects from CF and NNM differ in size, both indicate positive but diminishing effects of kindergarten retention on the retained children's test scores.[21] In contrast with much of the existing literature, the positive effects we find may be due in part to our particular choice of models. Besides specifications, two other factors are worth emphasizing, namely, differences in samples and test scores.

[21] The estimated ATT for first and third grade reading and math scores when excluding individualized education plan are 0.535, 0.152, 0.527, and 0.409, respectively.

Unlike most existing studies, which draw samples from local school districts, this study uses a nationally representative sample. In addition, we adopt IRT test scores from a longitudinal study, which are specially designed to reveal children's true cognitive levels and are comparable across different waves of surveys. This is important because an accurate estimation of retention effects relies on the accuracy of test scores as a measure of children's true cognitive levels and their comparability before and after the retention treatment. In contrast, many existing studies either do not control for the pre-treatment cognitive level or use test scores from some other source as a proxy.

5. Conclusions and Policy implications

Motivated by the growing discrepancy between educational practice and research findings, this paper examines the effects of kindergarten retention on the retained children's later academic performance using a recently collected nationally representative sample from the US.

The primary research question investigated is whether the retained children actually did better than they would have done, had they been socially promoted. Other issues explored include non-random selection of the kindergarten retention policy, and the role of unobserved child, family, and school characteristics in selection into the retention treatment.

This paper models the retention treatment as a binary choice with sample selection. This retention model explicitly takes into account the non-random selection of children into different types of schools, i.e., retention vs. non-retention schools. The retention treatment dummy then shows up in the test score equation as an endogenous regressor with a correlated random coefficient, which captures the heterogeneity of treatment effects. A control function estimator is derived and applied to the resulting double-hurdle treatment model. As a comparison, a nearest neighbor matching analysis is also conducted. Both the parametric control function approach and the nonparametric matching method are implemented under a variety of assumptions regarding selection into retention schools.

Findings from this study show that repeating kindergarten has positive effects on the retained children's later academic performance; i.e., the retained children would do worse in terms of the first and third grade test scores, were they socially promoted. Our results also suggest that these effects diminish over time. For example, while the positive effect on the retainees' math test scores is still significant up to third grade, the effect on the reading test scores is not.

Comparison of the results from the control function and matching approaches shows that unobserved child, family, and school characteristics that affect a child's probability of repeating kindergarten also affect his academic performance. Even controlling for an extensive set of observables, matching fails to balance the unobserved differences between retained and promoted children, and so tends to underestimate the retention effect. This suggests a more cautious interpretation of the mainstream conclusions about retention effects in the existing research.

Results from this study should encourage researchers, education professionals, and legislators to take a more optimistic attitude regarding kindergarten retention. Specifically, it is shown that holding the low achieving kindergarteners back provides a boost in their academic performance, although the effect may wear off over time. That is, kindergarten retention may give lagging children a chance to make up, if not catch up.

Appendix

Table A1 Description of Explanatory Variables

Explanatory Variable	Description
D	A dummy indicating whether a child was retained in kindergarten.
K1 math	Math IRT test score at the end of the first year kindergarten (at the beginning of the treatment).
K1 reading	Reading IRT test score at the end of the first year kindergarten (at the beginning of the treatment).
White	A dummy indicating whether a child is non-Hispanic white.
Female	A dummy indicating whether a child is female.
Age-at-test	A child's chronological age at the time of test, in months.
Age-at-entry	A child's chronological age at kindergarten entrance, in months.
Hearing difficulty	A dummy indicating whether a child has difficulty in hearing speeches.
Seeing difficulty	A dummy indicating whether a child has difficulty in seeing far objects or letters on paper.
Communication ability	A dummy indicating whether a child pronounces words, communicates with and understands others less well than the same-age children.
Overactive	A dummy indicating whether a child is overactive.
Problem solving ability	A dummy indicating whether a child's ability to learn, to think, and to solve problems is below average among same-age children.
Disabled	A dummy indicating whether a child is disabled.
IEP	A dummy indicating whether a child received individualized education from school.
Parental educational expectation	A dummy indicating whether parents expect their children to have high school or lower education.
SES	Family social economic status. Details about how it is created can be found in ECLS-K user's manual.
Parents' highest education	Three dummies representing less than high school, high school or some college (the default), and bachelor's degree or above education.
Family type	A dummy indicating whether a child is from a single/no parent family.
All-day K	A dummy indicating whether the kindergarten is all-day.
Private K	A dummy indicating whether the kindergarten is private.
School received Title 1 Funds	A dummy indicating whether the school received federal Title I funds.
Survey regions	Four dummies representing the West, South, Midwest, and Northeast (the default) survey regions.

Table A2 Kindergarten Enrollment Age Cutoff Dates by State[22]

State	Kindergarten enrollment cutoff date	State	Kindergarten enrollment cutoff date
Alabama	Sept. 1	Nebraska	Oct. 15
Alaska	Aug. 15	Nevada	Sept. 30
Arizona	Sept. 1	New Hampshire	LEA Option
Arkansas	Sept. 15	New Jersey	LEA Option
California	Dec. 2	New Mexico	before Sept. 1
Colorado	LEA Option	New York	LEA Option
Connecticut	Jan. 1	North Carolina	Oct. 16
Delaware	Aug. 31	North Dakota	Sept. 1
District of Columbia	Dec. 31	Ohio	Sept. 30 or Aug. 1
Florida	Sept. 1	Oklahoma	Sept. 1
Georgia	by Sept. 1	Oregon	Sept. 1
Hawaii	Dec. 31[23]	Pennsylvania	LEA Option
Idaho	Sept. 1	Puerto Rico	Aug. 31
Illinois	Sept. 1	Rhode Island	Sept. 1
Indiana	Jul. 1	South Carolina	Sept. 1
Iowa	Sept. 15	South Dakota	Sept. 1
Kansas	Aug. 31	Tennessee	Sept. 30
Kentucky	Oct. 1	Texas	Sept. 1
Louisiana	Sept. 30	Utah	Sept. 2
Maine	Oct. 15	Vermont	Jan. 1[24]
Maryland	Oct. 31[25]	Virgin Islands	Dec. 31
Massachusetts	LEA Option	Virginia	Sept. 30
Michigan	Dec. 1	Washington	Aug. 31
Minnesota	Sept. 1	West Virginia	Sept. 1
Mississippi	Sept. 1	Wisconsin	Sept. 1
Missouri	Aug. 1[26]	Wyoming	Sept. 15
Montana	Sept. 10		

[22] Source: State Note by the Education Commission of the States (ECS), 2005

[23] In 2006-07, the date changed to on or before Aug. 1

[24] In Vermont, districts may set the enrollment age cutoff date anywhere between Aug. 31 and Jan. 1 of the same school year.

[25] In 2005-06, this changed to Sept. 30. In 2006-07, it changed to Sept.1.

[26] LEA option between Aug. 1 and Oct. 1 for metropolitan districts.

Table A3-(1) The Retention Model – Using both the Multiple Retention Policy and School Travel Mode Instruments

	Retention Treatment	School Selection
Multiple retentions allowed in elementary school:		
Yes		1.11 (.050) ***
Unknown		-.139 (.055) ***
School travel modes:		
Riding a bus		.164 (.063) ***
Dropped off by a parent		.307 (.061) ***
Dropped off by a day care provider		.103(.120)
Other		.026 (.260)
K enrollment cutoff date	-.185(.031) ***	.037 (.013) ***
K1 math	-.370(.093) ***	-.012 (.026)
K1 reading	-.187(.105)	.047 (.025) **
White(non-Hispanic only)	.222(.091) **	.172 (.041) ***
Female	-.502 (.080) ***	-.022 (.036)
Age-at-entry	-1.37 (.171) ***	-.334 (.100) ***
Age-at-entry squared	.009 (.001) ***	.002 (.001) ***
Hearing difficulty	.232 (.204)	.152 (.127)
Seeing difficulty	.029 (.150)	-.170 (.084) **
Communication ability: Less well than same-age children	.004 (.116)	-.112 (.066)
Overactive	.109 (.105)	-.156 (.053) ***
Problem solving ability:		
Less well than same-age children	.350 (.118) ***	.091 (.082)
Disabled	.459 (.106) ***	.014 (.062)
IEP	.655 (.155) ***	.142 (.112)
Parental educational expectation:		
High school or less	.224 (.097) **	.078 (.054)
Family type: Single/no parent	.153 (.097) *	-.104 (.049) **
SES	.110 (.089)	.038 (.042)
Parents' highest education:		
Bachelor's degree or above	-.021 (.112)	.108 (.054) *
Less than high school	.508 (.146) ***	.059 (.083)
All-day K	.190 (.080) **	-.105 (.040) **
Private K	.557 (.086) ***	-.073 (.049) *
School received Title 1 Funds	-.43 (.085)	-.396 (.040) ***
West	-.008 (.121)	.132 (.060) **
South	.001 (.108)	-.253 (.058) ***
Midwest	-.014 (.117)	.155 (.055) ***
Constant	50.4 (5.77) ***	11.6 (3.29) ***
Correlation of the two Eq.s' errors	-.278 (.124)**	

Note: *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.

Table A3-(2) The Retention Model – Using the Multiple Retention Policy Instrument

	Retention Treatment	School Selection
Multiple retentions allowed in elementary school:		
Yes		1.09 (.050)***
Unknown		-.135 (.055)**
K enrollment cutoff date	-.185 (.031)***	.037 (.013)***
K1 math	-.371 (.094)***	-.012 (.026)
K1 reading	-.186 (.105)*	.047 (.025)*
White(Non-Hispanic only)	.224 (.091)**	.170 (.041)***
Female	-.503 (.080)***	-.023 (.036)
Age-at-entry	-1.37 (.171)***	-.335 (.100)***
Age-at-entry squared	.009* (.001)**	.003 (.001)***
Hearing difficulty	.232 (.204)	.147 (.126)
Seeing difficulty	.028 (.150)	-.166 (.083)**
Communication ability:		
Less well than same-age children	-.005 (.117)	-.113 (.066)*
Overactive	.108 (.105)	-.153 (.053)***
Problem solving ability:		
Less well than same-age children	.351 (.118)***	.084 (.081)
Disabled	.459 (.107)***	.012 (.062)
IEP	.655 (.155)***	.144 (.111)
Parental educational expectation:		
High school or less	.225 (.097)**	.067 (.054)
Family type: Single/no parent	.509 (.147)***	-.122 (.047)***
SES	.109 (.089)	.048 (.041)
Parents' highest education:		
Bachelor's degree or above	.153 (.097)	.106 (.054)*
Less than high school	-.021 (.112)	.045 (.083)
All-day K	.189 (.080)**	-.099 (.040)**
Private K	.558 (.086)***	-.017 (.047)
School received Title 1 Funds	-.144 (.085)*	-.404 (.040)***
West	-.007 (.122)	.171 (.057)***
South	.002 (.108)	-.211 (.057)***
Midwest	-.014 (.117)	.172 (.055)***
Constant	50.4 (5.78)*	11.8 (3.30)***
Correlation of the two Eq.s' errors	-.263 (.130)**	

Note: *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.

Table A3-(3) The Retention Model – Using the School Travel Mode Instruments

	Retention Treatment	School Selection
School travel modes:		
Riding a bus		.152 (.066)**
Dropped off by a parent		.287 (.065) ***
Dropped off by a day care provider		.058 (.120)
Other		.031 (.254)
K enrollment cutoff date	-.185 (.031) ***	.017 (.012)
K1 math	-.376 (.095) ***	-.011 (.025)
K1 reading	-.164 (.105)	.042 (.024) *
White(non-Hispanic only)	.235 (.095) **	.110 (.039) ***
Female	-.520 (.082) ***	-.025 (.035)
Age-at-entry	-1.38 (.176) ***	-.306 (.096) ***
Age-at-entry squared	.009 (.001) ***	.002 (.001) ***
Hearing difficulty	.250 (.208)	.165 (.122)
Seeing difficulty	.034 (.154)	-.195 (.081) **
Communication ability:		
Less well than same-age children	.037 (.118)	-.100 (.064)
Overactive	.086 (.109)	-.169 (.052) ***
Problem solving ability: Less well than same-age children	.366 (.120) ***	.057 (.079)
Disabled	.417 (.108) ***	.009 (.060)
IEP	.653 (.160) ***	.087 (.107)
Parental educational expectation:		
High school or less	.239 (.100) **	.079 (.052)
Family type: Single/no parent	.183 (.099) *	-.091 (.046) **
SES	.097 (.091)	.027 (.040)
Parents' highest education:		
Bachelor's degree or above	.009 (.114)	.107 (.052) **
Less than high school	.526 (.153) ***	.061 (.081)
All-day K	.188 (.082) **	-.106 (.039) ***
Private K	.557 (.088) ***	-.110 (.046) **
School received Title 1 Funds	-.130 (.089)	-.354 (.038) ***
West	-.037 (.124)	.018 (.058)
South	-.008 (.111)	-.079 (.054)
Midwest	-.023 (.119)	.119 (.053) **
Constant	50.8 (5.94) ***	11.1 (3.16) ***
Correlation of the two Eq.s' errors	-.168 (.177)	

Note: *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.

Table A4 The Test Score Equations

	1st grade reading	3rd grade reading	1st grade math	3rd grade math
D	.727 (.092)***	.551 (.107)***	.746 (.094)***	.686 (.099)***
K1 math	.182 (.010)***	.324 (.012)***	.643 (.010)***	.565 (.011)***
K1 reading	.596 (.009)***	.230 (.011)***	.081 (.010)***	.079 (.010)***
White	.035 (.017)**	.224 (.019)***	.156 (.017)***	.171 (.018)***
Female	.043 (.014)***	.115 (.017)***	-.124 (.015)***	-.235 (.016)***
Age-at-test	.095 (.068)	-.034 (.032)	.138 (.070)**	-.003 (.030)
Age-at-test squared	-.001 (.000)	.000 (.004)	-.001 (.000)**	-.006 (.004)
Hearing difficulty	.024 (.049)	-.043 (.058)	-.093 (.051)*	-.062 (.054)
Seeing difficulty	-.026 (.036)	-.008 (.042)	-.006 (.037)	.005 (.039)
Communication ability: Less well than same-age children	-.112 (.027)***	-.188 (.032)***	.037 (.028)	-.040 (.029)
Overactive	.007 (.023)	-.042 (.027)	.000 (.024)	-.043 (.025)*
Problem solving ability: Less well than same-age children	-.099 (.034)***	-.197 (.039)***	-.185 (.035)***	-.243 (.036)***
Disabled	-.042 (.025)*	-.028 (.030)	-.047 (.026)**	-.013 (.028)
IEP	-.111 (.046)**	-.304 (.054)***	-.207 (.048)***	-.299 (.050)***
Parental educational expectation: High school or less	-.107 (.022)***	-.199 (.026)***	-.072 (.023)***	-.178(.024)***
SES	.050 (.016)***	.166 (.018)***	.078 (.016)***	.129 (.017)***
Parents' highest education: Bachelor's degree or above	.038 (.021)*	-.056 (.041)	.031 (.022)	.040 (.023)*
Less than high school	-.087 (.035)**	.036 (.025)	.055 (.036)	.038 (.038)
Family type: Single/no parent	-.033 (.020)	-.067 (.024)***	-.007 (.021)	-.069 (.022)***
All-day K	-.075 (.016)***	-.085 (.018)***	-.053 (.016)***	-.095 (.017)***
Private K	.026 (.018)	-.013 (.021)	-.058 (.019)***	-.197 (.020)***
School received Title 1 Funds	-.045(.015)***	-.007 (.017)	-.044 (.015)***	-.025 (.016)
West	-.083(.023)***	-.041 (.026)	-.029 (.023)	.021 (.025)
South	-.001 (.021)	.029 (.025)	.073 (.022)***	.059 (.023)**
Midwest	-.070 (.021)***	.005 (.024)	.006 (.021)	.009 (.022)
λ_1	-.257 (.051)***	-.248 (.059)***	-.169 (.052)***	-.179 (.054)***
λ_2	-.254 (.243)	-.110 (.284)	-.229 (.251)	-.353 (.264)
λ_3	-.153 (.092)*	.056 (.106)	-.045 (.095)	-.042 (.098)
λ_4	-.169 (.046)***	-.040 (.054)	-.003 (.048)	.010 (.050)
λ_5	.045 (.017)***	.044 (.019)**	-.014 (.017)	.025 (.018)
Constant	-3.77 (2.96)	.038 (.064)	-6.05 (3.06)**	.212 (.060)***

Note: 1. *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.

2. Using the estimated coefficients of $\lambda_1, \lambda_2, ...,$ and λ_5, one could construct 5 equations in 5 unknowns $\sigma_{SD}, \sigma_{1S}, \sigma_{0S}, \sigma_{1D}$, and σ_{0D}, so in theory one could then combine these covariances with estimated error variances to obtain correlations between errors in any two equations. However, these nuisance parameters are likely to be poorly estimated partly because these errors are either latent ($\varepsilon_S, \varepsilon_D$) or partially latent ($\varepsilon_{Y1}, \varepsilon_{Y0}$), so we do not calculate these correlations.

Reference

Glidden, H., 1998. Making Standards Matter 1998. An Annual Fifty-State Report on Efforts to Raise Academic Standards. American Federation of Teachers, Washington, DC

Abadie A., Drukker D., Herr J. L., Imbens G. W., 2001. Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal* 1, 1–18.

Abadie A. and Imbens G. W., 2001. Simple and Bias-Corrected Matching Estimators for Average Treatment Effects, no. 283, *NBER Technical Working Papers*.

Bedard K. and Dhuey E. 2006. The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics* 121(4), 1437-1472.

Collins, M. A., Brick, J. M., 1993. Parent reports on children's academic progress and school adjustment in the National Household Education Survey. *Proceedings of the Social Statistics Section*, American Statistical Association, Alexandria, VA.

Deming, D. and Dynarski, S., 2008. The Lengthening of Childhood. *Journal of Economic Perspectives* 22, 71-92

Early, D., Bushnell M., Clifford R., Konanc E., Maxwell K., Palsha S., Roberts L., 2003. North Carolina early grade retention in the age of accountability. FPG Publications: The Partners in Research.

Eide, E. R. and Showalter, M. H. 2001. The Effect of Grade Retention on Educational and Labor Market Outcomes. *Economics of Education Review* 20, 563-576.

Elder E. T. and Lubotsky H. D., 2009. Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers. *forthcoming, Journal of Human Resources*

Fertig, M. and Kluve, J., 2005. The Effect of Age at School Entry on Educational Attainment in Germany. *IZA Discussion Paper* No. 1507.

Fredriksson, Peter, and Bjorn Öckert, 2005. Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labour Market Performance. *IZA Discussion Paper* No. 1659.

Frederick, C., and Hauser, R.M. 2008. Have we put an end to social promotion? Changes in grade retention rates among children aged 6 to 17 from 1972 to 2005. Demography, 45(3), 719-740.

Hauser, R. M., Frederick, C. and Andrew M., 2007. Grade Retention in the Age of Accountability. Pp. 122-155 in Adam Gamoran, Ed. *No Child Left Behind and Poverty* Washington, DC: Brookings Institution.

Hauser, R. M., Pager, D. I. and Simmons, S. J., 2004. Race-Ethnicity, Social Background, and Grade Retention. In: H. J. Walberg, A.C. Reynolds, and M. C. Wang (Eds.), *Can Unlike Students Learn Together? Grade Retention, Tracking and Grouping*: pp. 97-114. Greenwich, CT: Information Age Publishing.

Heckman J., 1997. Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources* 32, 441- 462.

Holmes, C. T. and Matthews, K. M., 1984. The Effects of Nonpromotion on the Elementary and Junior High School Pupils: A Meta-analysis. *Review of Educational Research* 54, 225–236.

Holmes, C. T., 1989. Grade level retention effects: A Meta-analysis of Research Studies. In: L. A. Shepard, and M. L. Smith (Eds.), *Flunking grades: Research and policies on retention*: pp. 16–33. Philadelphia, The Falmer Press.

Hong, G. and Raudenbush W. S., 2005. Effects of Kindergarten Retention Policy on Children’s Cognitive Growth in Reading and Mathematics. *Educational Evaluation and Policy Analysis* 27, 205–224.

Jackson, L. B., 1975. The Research Evidence on the Effect of Grade Retention. *Review of Educational Research* 45, 438–460.

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

Jimerson, S. R. and Kaufman, A. M. 2003. Reading, Writing, and Retention: A Primer on Grade Retention Research. *The Reading Teacher*, 56, 622-635.

Leckrone, M. J. and Griffith, B. G. 2006. Retention realities and educational standards. Children and Schools 28, 53-58

Manning, A., 2004. Instrumental Variables for Binary Treatments with Heterogenous Treatment Effects: A Simple Exposition. *Contributions to Economic Analysis and Policy*, Vol. 3, Issue 1, Article 9.

Mantzicopoulos, P., Morrison, D. C., Hinshow, S. P. and Carte, E. T., 1989.

Nonpromotion in kindergarten: The role of cognitive, perceptual, visual–motor, behavioral, achievement, socioeconomic, and demographic characteristics. *American Educational Research Journal* 26, 107–121.

Mantzicopoulos, P. and Morrison, D., 1992. Kindergarten retention: Academic and behavioral outcomes through the end of second grade. *American Educational Research Journal* 29, 182–198.

Mantzicopoulos, P., 1997. Do certain groups of children profit from early retention? A follow-up study of kindergartners with attention problems. *Psychology in the Schools* 34, 115 – 127.

McArthur, E. K., Bianchi, S. M., 1993. Characteristics of children who are 'behind' in school', *Proceedings of the Social Statistics Section*. American Statistical Association, Alexandria, VA.

McCoy A. R., Reynolds A. J., 1999. Grade retention and school performance: An extended investigation. *Journal of School Psychology* 37, 273 – 298.

McEwan J. P. and Shapiro S. J., 2007. The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. (*Forthcoming) Journal of Human Resources.*

Mantzicopoulos P. Y., 1997. Do certain groups of children profit from early retention? A follow-up study of kindergartners with attention problems. *Psychology in the Schools* 34, 115 – 127.

Meisels, S. J. and Liaw, F. R., 1993. Failure in grade: Do retained students catch up? *Journal of Educational Research*, Vol. 87, pp. 69–77.

Niklason, L. B., 1987. Do certain groups of children profit from a grade retention?. *Psychology in the Schools* 24, 339–345.

Rosenbaum, P. and Rubin, D., 1984. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association* 79, 516–524.

Rubin, D. B., 1978. Bayesian Inference for Causal Effects: The Role of Randomization', *Annals of Statistics* 6, 34–58

Shepard, L. A., 1989. A review of research on kindergarten retention. In: L. A. Shepard, and M. L. Smith, (Eds.), *Flunking grades: Research and policies on retention*: 64–78. Philadelphia: Falmer Press.

Smith, M. L. and Shepard, L. A., 1987. What doesn’t work: Explaining policies of retention in the early grades. *Phi Delta Kappan* 69, 129–134.

Shepard, L. A. and Smith, M. L., 1986a. Synthesis of research on school readiness and kindergarten retention. *Educational Leadership* 44, 78–86.

Shepard, L. A. and Smith, M. L., 1986b. Effects of kindergarten retention at the end of first grade. *Psychology in the Schools* 24, 346–357.

Stipek, D., 2002. At What Age Should Children Enter Kindergarten? A Question for Policy Makers and Parents. *Social Policy Report* 16, 3–16.

Vella F. and Verbeek M., 1999. Estimating and Interpreting Models with Endogenous Treatment Effects. *Journal of Business and Economic Statistics* 17, 473–478.

Zill, N., Loomis, L.S. and West, J., 1997. *The Elementary School Performance and Adjustment of Children Who Enter Kindergarten Late or Repeat Kindergarten: Findings from National Surveys* (Statistical analysis report NCES 98-097). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.