A Framework for the Analysis of Dynamic Treatment Effects: Grade Retention and Test Scores*

Preliminary and Incomplete. Do not cite.

Jane Cooley[†], Salvador Navarro[‡] and Yuya Takahashi[§]

March 3, 2008

1 Introduction

In this paper we present and apply a simple framework for the analysis of models with dynamic treatment effects. We consider a model with multiple treatments in which selection into treatment occurs sequentially over time. In our framework getting one treatment is equivalent to getting treated at a given time. Because we assume each treatment time is as an absorbing state, treatment assignment can be thought of as a single spell time until treatment type of model. Associated with each treatment time there is a, possibly vector valued, outcome of interest used to measure the effectiveness of being selected into a particular treatment time. Since treatment time assignment is sequential, i.e. one can only be selected into treatment time t after not being assigned to time $t-1$, and because there is a potentially different effect of being treated at each time we say that there are dynamic treatment effects. The framework in this paper is

*First draft: December 26, 2007. Navarro's work was supported by the Institute for Research on Poverty at the University of Wisconsin-Madison. We thank participants at the Applied Microeconomics lunch at UW-Madison, Anirban Basu, David Meltzer and especially John Kennan and Chris Taber for providing helpful comments.

[†]Department of Economics, University of Wisconsin-Madison. email: jcooley@ssc.wisc.edu
[‡]Department of Economics, University of Wisconsin-Madison. email: snavarro@ssc.wisc.edu
[§]Department of Economics, University of Wisconsin-Madison. email: takahashi2@wisc.edu

closely related to the one in Heckman and Navarro (2007)[1].

We apply our framework to analyse the dynamic effects of grade retention on test scores using data from the 1998-99 kindergarten class of ECLS-K. Grade retention is a common practice in the U.S. For example, using data from the CPS, Hauser (1999) finds that at least 15 percent of pupils are retained between ages 6 to 8 and ages 15 to 17. The number is roughly 10% if one looks at retention between kindergarten and 4^{th} grade in ECLS-K. Most research on the effects on grade retention treats it as a single treatment. In general, the literature on the topic finds that retention at best has no effect and at worse has considerable negative effects. For example, using a regression discontinuity design to study test-based promotion in Chicago public schools, both Jacob and Lefgren (2004) and Nagaoka and Roderick (2005) find that retention has small short term gains on test scores that disappear over time. Other studies (Roderick (1994)) have found that being retained is associated with lower high school graduation rates.

In this paper we treat retention as a dynamic sequential selection process. That is, every period a child can either be selected into repeating the current grade he is in or not. Furthermore, we allow for the possibility that the effects of retention are dynamic and heterogeneous: being retained in kindergarten is not necessarily the same as being retained in first grade and these effects maybe different across the population in non-trivial ways. We develop a semiparametrically identified model of multidimensional ability (general, cognitive and behavioral) that allows us to not only look at average effects but also at the distributions. We find evidence of dynamic selection being important on grade retention. In general we find that, on average, repeating a grade reduces test scores at age 11 between one quarter and half a standard deviation depending on the grade that is repeated. However, we also find that high ability kids could benefit from being retained in kindergarten (although they would still be hurt if retained at higher grades).

The paper proceeds as follows. In the next section we lay down a general framework for the analysis of dynamic treatment effects. We then show how our framework nests the standard static treatment cases but with and without essential heterogeneity. In section 3 we apply our framework to analyze the effects of grade retention. We first clarify a point that has been missed in the literature: one can evaluate grade retention

[1]See also Robins (1989), Gill and Robins (2001), Abbring and Van den Berg (2003) and Murphy (2003).

by either keeping grade fixed or by keeping age fixed. Depending on what case one is working with this may lead to different conclusion. We then develop a model of multidimensional unobserved ability that we use when we analyze the ECLS-K data. We first present some preliminary regression analysis that supports the idea of dynamic selection being present in the retention decision as well as evidence of dynamic treatment effects. Then we estimate the full model and present our results. The last section concludes.

2 A Framework for the Analysis of Dynamic Treatments

Consider the generic problem of evaluating the efficacy of a treatment. Let $P \in \mathcal{P} = \{1, 2, ..., \bar{P}\}$ index calendar time and $i = 1, ..., I$ index the individual. Since we allow for the possibility of the treatment being taken at different times, we define a random variable T that indicates the time at which treatment is received. We assume that treatment is taken at most once[2]. We let $T \in \mathcal{T} \subseteq \mathcal{P}$, that is treatment may be received at any time or only in a subset if treatment selection is limited to a subperiod so $\mathcal{T} = \{\underline{T}, \underline{T}+1, ..., \bar{T}-1, \bar{T}\}$ for $\underline{T} \geq 1$ and $\bar{T} \leq \bar{P}$. We adopt the convention of letting $T = 0$ for the "never" treated state[3].

The (possibly vector valued) outcome of interest at time P for an individual i who takes treatment at time T is denoted by $Y_i(P, T)$. For notational simplicity, we keep all conditioning on covariates implicit. Finally, we define a random variable $D_i(T)$ that takes value 1 if an individual receives treatment at time T and 0 otherwise. For individual i the observed outcome in period P will be given by

$$Y_i(P) = \sum_{\tau=\underline{T}}^{\bar{T}} D_i(\tau) \left[Y_i(P, \tau) - Y_i(P, 0)\right] + Y_i(P, 0). \tag{1}$$

As opposed to the standard binary treatment case, we now have many possible potential outcomes. That is, while the standard case only has the treated and untreated

[2] Extending the framework to allow for the possibility of treatment being taken more than once is straightforward.

[3] Depending on the situation this case may be more accurately described as the "not treated yet" or "not treated in the sample period".

potential states we have the untreated, the treated at time $\underline{T}$, treated at time $\underline{T}+1$, etc. The potential states, however, are ordered sequentially over time. Because of the sequential nature of the problem, by letting $Y_i(P,T)$ depend on treatment time, we allow for the possibility that the effect of treatment is dynamic in the sense that it does not only depend on receipt and that there may be dynamic selection into treatment.

Following Abbring and Van den Berg (2003) we also impose that

A-1 *(No anticipations)* $Y_i(P,T) = Y_i(P,0) = Y_i(P)$ *for* $T \geq P$.

That is, we rule out that $Y_i(P, P+\kappa) \neq Y_i(P, P+\kappa+1)$ for $\kappa > 0$ and so we do not allow potential outcomes that ex-ante should be the same to differ because *in the future* treatment times will be different. While there is nothing wrong with this being the case and we can accommodate it, we keep the assumption for simplicity[4].

We further write the outcomes as

$$Y_i(P,T) = \mu(P) + \Phi(P,T) + \epsilon_i(P,T), \tag{2}$$

where, because of **A-1**, we impose $\Phi(P,T) = 0$ and $\epsilon_i(P,T) = \epsilon_i(P)$ if $T \geq P$. The additive separability in outcomes is not required, it can be relaxed using the analysis in Matzkin (2003).

We assume that selection into treatment, and treatment time, is determined by a single spell duration model that follows a sequential threshold crossing structure as in Heckman and Navarro (2007). If we define the treatment time specific index $V_i(T) = \lambda(T) + U_i(T)$, then treatment time is selected according to

$$\begin{aligned} D_i(T) &= \mathbf{1}\left(V_i(T) > 0 \mid \{V_i(\tau) < 0\}_{\tau=1}^{T-1}\right) \\ &= \mathbf{1}\left(V_i(T) > 0 \mid \{D_i(\tau) = 0\}_{\tau=1}^{T-1}\right) \end{aligned}$$

where $\mathbf{1}(a)$ is an indicator function that takes value 1 if a is true and 0 otherwise. The additive separability in the index is assumed for simplicity and can be relaxed using the analysis in Matzkin (1992). Notice that the selection process is dynamic in the sense that today's choice depends on yesterdays choice: treatment time T can only be selected if treatment has not been taken before. Generalizing the model for repeated treatments (i.e. multiple spells) is straightforward.

[4]See Abbring and Van den Berg (2003) and Heckman and Navarro (2007) for a discussion.

In general, we allow $(\epsilon_i(P,T), \epsilon_i(P'T''), U_i(T'''), U_i(T''''))$ to be all correlated. In particular, we impose the following assumption

A-2 *(Factor structure)* $\epsilon_i(P,T) = \theta_i\alpha(P,T) + \varepsilon(P)$ and $U_i(T) = \theta_i\rho(T) + \upsilon_i(T)$ where θ_i is a vector of mutually independent "factors" and we assume that $\varepsilon_i(P) \perp\!\!\!\perp \varepsilon_i(P')$ for all $P \neq P'$ and $\upsilon_i(T) \perp\!\!\!\perp \upsilon_i(T')$ for all $T \neq T'$ where $\perp\!\!\!\perp$ denotes statistical independence[5].

We impose **A-2** for convenience even though it is much stronger than required. As shown in Heckman and Navarro (2007) one can still identify many parameters without the factor structure. Furthermore, following the analysis in Schennach (2004) we can relax the strong statistical independence assumptions.

The factor structure assumption is a convenient dimension reduction technique. It allows us to transform the enormous problem of identifying and estimating the joint distribution of $U_i(T'), \epsilon_i(P,T)\ \forall P,T$ into a simpler problem: that of recovering the factor loadings $\alpha(P,T)$ and $\rho(T)$ and the marginal distributions of the elements of θ_i and of $\varepsilon_i(P), \upsilon_i(T)\ \forall P,T$. It also aids in interpretation of results since we can now talk about a low dimensional set of common "causes"[6].

2.1 Defining Treatment Effects

Let's first consider the problem of defining what constitutes "the" effect of treatment at the individual level. Under our assumptions, only $\Phi(P,T)$ and $\alpha(P,T)$ are related to treatment (and treatment time). One would expect that the parameter of interest should only depend on them. We can define at least two different candidates for the individual effect of treatment. The first parameter

$$\Delta_i^1(P,T,T') = Y_i(P,T) - Y_i(P,T')$$

$$= \Phi(P,T) - \Phi(P,T') + \theta_i[\alpha(P,T) - \alpha(P,T')],$$

measures the effect at period P of receiving treatment at time T versus receiving treatment at time T'. If we let $T' = 0$, this parameter would measure the effect at P

[5] If **A-1** holds, $\alpha(P,T) = \alpha(P,0) = \alpha(P)$ for $T \geq P$.

[6] See Jöreskog and Goldberger (1975) for a discussion and Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) for recent developments.

of receiving treatment at time T versus not receiving treatment at all: $\Phi(P,T) + \theta_i [\alpha(P,T) - \alpha(P)]$. An example of this first parameter would be the difference in earnings at age 40 for an individual who graduates from college in 4 years versus if he graduates in 6 years.

The second individual parameter of interest

$$\begin{aligned}\Delta_i^2(\tau,T,T') &= Y_i(T+\tau,T) - Y_i(T+\tau,0) - [Y_i(T'+\tau,T') - Y_i(T'+\tau,0)] \\ &= \Phi(T+\tau,T) - \Phi(T'+\tau,T') + \theta_i[\alpha(T+\tau,T) - \alpha(T'+\tau,T')], \text{ for } \tau > 0\end{aligned}$$

measures the difference in the effect of receiving treatment versus not receiving treatment τ periods after treatment time for two different treatment times T and T'. By subtracting the no treatment outcome from both treatment times, we eliminate the outcome differences that occur simply because of the differences in calendar time P. An example of this parameter would be the difference in the 1 hour survival probability after taking treatment, say aspirin after a hearth attack, when treatment is taken 1 minute after the event versus 10 minutes after the event.

Depending on the type of analysis either Δ_i^1 or Δ_i^2 may be the parameter of interest for the analyst. Many other individual level parameters may also be defined. Regardless of the way one defines the effect of treatment, it is potentially individual specific even conditional on covariates. Furthermore, we can now define many different population average parameters, not only because of the potential heterogeneity of the effects but also because of the variety of possible treatments. For example, we can define the average effect of receiving treatment at time $T=t$ versus not receiving treatment

$$ATE(P,t) = E(Y(P,t) - Y(P)) = \Phi(P,t);$$

the average effect of treatment at time $T=t$ for people who receive treatment at time $T=t$

$$TT(P,t) = E(Y(P,t) - Y(P) | T=t)$$

and so on. We can also define many more mean treatment parameters like the average effect of receiving treatment at $T=t$ versus receiving treatment at $T=t'$

$$ATE(P,t,t') = E(Y(P,t) - Y(P,t'))$$

or the effect of treatment at $T = t$ versus treatment at $T = t'$ for people who are actually treated at time $T = t''$

$$TT(P, t, t', t'') = E(Y(P, t) - Y(P, t') | T = t''),$$

etc.

Depending on the assumptions one is willing to make, some of these parameters may equal each other. As shown below, four main cases can be distinguished. For simplicity we focus on Δ_i^1 but a parallel analysis can be made for Δ_i^2 or any other parameter of interest.

2.1.1 Static Homogeneous Treatment

The following assumption describes the case of static treatments.

S-1 *(No dynamics)* $\Phi(P, T) - \Phi(P, T') = \begin{cases} \Phi(P) & \text{if } T < P < T', \forall P \in \mathcal{P} \\ 0 & \text{otherwise} \end{cases}$ and

$$\alpha(P, T) - \alpha(P, T') = \begin{cases} \delta(P) & \text{if } T < P < T', \forall P \in \mathcal{P} \\ 0 & \text{otherwise} \end{cases}.$$

When assumption **S-1** holds we are back to the case in which only receipt of treatment and not time at treatment matters. Notice that, unless $\theta_i = \theta$ for all i and/or $\alpha(P, T) - \alpha(P, T') = 0$, the effect will be heterogeneous based on variables unobserved to the econometrician (the θ_i) and no single number (even for one definition of the effect, say Δ_i^1) will summarize the effectiveness of treatment at the individual level. If we further assume that

S-2 *(Homogeneity)* $\theta_i = \theta$ for all i.

we are back at the case in which there is only one effect of treatment which is the same for everyone regardless of treatment time. From equations (1) and (2) we can write

$$Y_i(P) = \mu(P) + D_i(P)[\Phi(P) + \theta\delta(P)] + \varepsilon_i(P).$$

In this case, under **A-2** , the effect of treatment $\Phi(P) + \theta\delta(P)$ is identified directly from the regression provided $\upsilon_i(T) \perp\!\!\!\perp \varepsilon_i(P)$. If this is not the case, one can instead use a standard instrumental variables estimator by finding a Z correlated with $D_i(P)$

but uncorrelated with $\varepsilon_i(P)$. Of course, experimental and quasi-experimental solutions in which people are randomized into treatment would also work.

2.1.2 Static Heterogeneous Treatment

If only **S-1** holds but **S-2** does not hold then the regression representation of outcomes will be given by

$$Y_i(P) = \mu(P) + D_i(P)[\Phi(P) + \theta_i \delta(P)] + \varepsilon_i(P)$$

and we now have a random coefficients model. Depending on whether $D_i(P)$ is correlated with θ_i , i.e. on whether selection into treatment is based on the unobserved (to the econometrician) treatment gains so $\rho(P) \neq 0$, we may have a correlated random coefficients model.

When there is no selection on unobservables ($\rho(P) = 0$) we have a standard uncorrelated random coefficients model, $\Phi(P)$ is identified from the regression above (or its instrumental variables counterpart) and we would only have to worry about heteroskedasticity.

When $\rho(P) \neq 0$ different methods like control functions, local instrumental variables, etc are required[7]. Depending on what the parameter of interest is, an experimental solution would require us to randomize people into treatment (to recover ATE) or to first condition on selection into treatment and then perform the randomization.

Notice that, as consequence of the sequential nature of the selection process, the heterogeneity of treatments can make it look as if the effect of treatment is dynamic. Since $\Pr(D_i(T) = 1) = \Pr(\theta_i \rho(T) + \upsilon_i(T) > -\lambda(T))$ the distribution of θ_i will be changing over time and it follows that

$$E(\theta_i | T = t) \neq E(\theta_i | T = t').$$

For example, if we were to simply take differences over the observed mean outcomes:

$$E(Y_i(P)|T = t) - E(Y_i(P)|T = t') = [E(\theta_i \mid T = t) - E(\theta_i \mid T = t')]\,\delta(P)$$

the end result looks as if the effect of treatment depends on treatment time even though

[7] See Heckman and Navarro (2004) for a discussion of different methods one can use in the presence of heterogeneous effects.

it in fact does not.

2.1.3 Dynamic Homogeneous Treatments

When assumption **S-1** does not hold we are back at the dynamic treatment effects case. When the effect of treatment (and treatment time) is homogeneous (i.e. if **S-2** holds) the observed outcome can be written as

$$Y_i(P) = \mu(P) + \sum_{\tau=1}^{P} D_i(\tau)\left[\Phi(P,\tau) + \theta\left[\alpha(P,\tau) - \alpha(P)\right]\right] + \theta\alpha(P) + \varepsilon_i(P)$$

and the unique effect of receiving treatment at treatment time T, $\Phi(P,T)+\theta\left[\alpha(P,T) - \alpha(P)\right]$ is identified from the regression above. If $D_i(T)$ is endogenous, then one would have to find as many instruments as treatment times in order to use standard instrumental variables methods. Notice that an experimental solution (i.e. randomization) would require people to be randomized not only into treatment but also into different treatment times. As shown in Heckman and Navarro (2007), one can estimate the joint selection-outcomes model and identify some, or all, treatment parameters even when no instruments are available.

2.1.4 Dynamic Heterogeneous Treatments

The final case arises when neither **S-1** nor **S-2** holds. The regression representation of the model in this case would be

$$Y_i(P) = \mu(P) + \sum_{\tau=1}^{P} D_i(\tau)\left[\Phi(\tau) + \theta_i\left[\alpha(P,\tau) - \alpha(P)\right]\right] + \theta_i\alpha(P) + \varepsilon_i(P),$$

and we have a model with many, potentially correlated, random coefficients. Furthermore, one has to worry not only about the endogeneity induced by selection into treatment but also the dynamic selection into time at treatment.

In this case, an experimental solution is still possible but, depending on the parameters of interest, may require a very complex design. In order to recover parameters like the effect of treatment at time t versus time t' conditional on $T = t''$ would require randomized groups to be re-randomized into different groups as treatment times progress. Again, one can estimate the joint system of selection equations and outcomes under the conditions in Heckman and Navarro (2007).

3 The Effect of Retention on Test Scores

In this section we apply the framework of section 2 to analyze the potentially dynamic effects that being retained at different grades may have on test scores. We first discuss the details associated with evaluating grade retention in general. We then develop a model for θ_i as multidimensional ability and show that it is semiparametrically identified. Finally we discuss the data and the results.

3.1 Evaluating grade retention

In order to fix ideas, we consider the following example for a model of the effect of retention on outcomes. To keep the discussion as focused as possible, we ignore the dynamic aspects and think of retention as an homogeneous treatment that happens just once between period $p-1$ and p. Let R be an indicator of retention. Let the outcome of interest (test scores in our case) in period p be written as

$$S = Age\gamma_{age} + Grade\gamma_{grade} + R\gamma_R + \varepsilon$$

so that the score an individual gets depends on age, grade and whether he has been retained.

Notice that, from a purely causal perspective γ_R would measure the effect of being retained on test scores. However, this effect holds both grade and age fixed, something that cannot be done in reality. To see why consider the following example. Take a 9 year old child who just finished 4^{th} grade and consider two potential situations. In the first case, he goes on to the next grade so that next period he is 10 years old and in 5^{th} grade. Alternatively he repeats 4^{th} grade so the next period he is 10 years old but he is in 4^{th} grade again. That is, going from $R=0$ to $R=1$ we can keep age fixed (at 10 years old) but we cannot keep the grade fixed. If $R=0$ then the kid would be in 5^{th} grade, while if $R=1$ he will be in 4^{th} grade. In this situation the effect of retention would be measured by

$$S\left(R=1, Age=11\right) - S\left(R=0, Age=11\right) = \gamma_{grade} + \gamma_R.$$

This is the right evaluation parameter when we think that retention is a policy that assumes that the kid will get more from repeating the 4^{th} grade than he would from skipping to 5^{th} grade even if he hasn't mastered completely the pre-fifth grade concepts[8].

Alternatively, one can keep grade fixed. In this case, the kid will be in 5^{th} grade at age 10 if he is not retained but at age 11 if he is. The effect of retention would be measured by

$$S\left(R=1, Grade=5\right) - S\left(R=0, Grade=5\right) = \gamma_{age} + \gamma_R.$$

This is the appropriate way to evaluate retention when the policy is put in place for the purpose of increasing the 5^{th} grade scores of a kid. This is the case if for example maturation effects are important and the kid is "too young" emotionally and would benefit from getting to 5^{th} grade a year older and so it includes the effect of age.

3.2 The ABC of ability

In order to put some structure on the unobservables determining selection (θ_i) in the framework of section 2, we think of the following model of individual ability as a multi-dimensional trait. In particular, we interpret θ_i as ability and propose that true ability consists of three independent components: 1) A trait associated purely with cognitive functions C, a purely behavioral trait B and general ability A that can be used for both cognitive and behavioral functions. That is, we assume that

$$\theta_i = \left(A_i, B_i, C_i\right).$$

Associated with ability is a set of tests or markers that measure these components of ability imperfectly. In particular, assume we have access to $N_c \geq 3$ measures (or tests) of cognitive functions $\zeta_{i,j}$, and $N_b \geq 3$ measures of behavioral functions, $\beta_{i,j}$, that are measured before any retention takes place (i.e. free of selection)[9]. As before, we keep all conditioning on covariates implicit to simplify notation[10]. We write the j^{th} demeaned cognitive test as

$$\zeta_{i,j} = A_i\alpha_{\zeta,j} + C_i\pi_{\zeta,j} + \varepsilon_{\zeta,j},$$

[8] In the empirical application we focus on the effect of retention keeping age fixed both because of the nature of the tests used and because of the structure of the data.

[9] Although this makes the identification argument cleaner, as shown in Heckman and Navarro (2007) it is not required.

[10] And we assume that these covariates are independent of θ_i.

and the j^{th} demeaned behavioral test as

$$\beta_{i,j} = A_i \alpha_{\beta,j} + B_i \phi_{\beta,j} + \varepsilon_{\beta,j}.$$

Under this interpretation, tests are noisy measures of the components of ability. Depending on the nature of the measure, some (like math and reading test scores) will be markers of cognitive ability C and general ability A and some (like measures of class disruptive behaviors or habits) will be noisy measures of the behavioral trait B and of general ability A. This is not to say that cognitive ability plays no role in behavioral aspects or viceversa but rather that whatever is common between these functions is captured by the general ability component A. The cognitive ability component C and the behavioral component B measure the part of ability that is used exclusively for the corresponding function.

Semiparametric identification follows from the following argument. Without loss of generality we impose the following normalizations $\alpha_{\zeta,1} = 1$, $\pi_{\zeta,1} = 1$ and $\phi_{\beta,1} = 1$ [11]. We first take cross moments between cognitive and behavioral measures

$$E\left((\zeta_j)^h (\beta_k)^n\right) = \alpha_{\zeta,j}^h \alpha_{\beta,k}^n E\left(A^{h+n}\right). \tag{3}$$

and form

$$\frac{E\left(\zeta_j (\beta_k)^n\right)}{E\left(\zeta_1 (\beta_k)^n\right)} = \frac{\alpha_{\zeta,j}\alpha_{\beta,k}^n E\left(A^{1+n}\right)}{\alpha_{\beta,k}^n E\left(A^{1+n}\right)} = \alpha_{\zeta,j}$$

to recover all of the the general ability loadings on cognitive tests, $\alpha_{\zeta,j}$, for $j = 2, \ldots, N_c$. We can then form

$$\alpha_{\zeta,j} \frac{E\left((\zeta_j)^h (\beta_k)^n\right)}{E\left((\zeta_j)^{h+1} (\beta_k)^{n-1}\right)} = \alpha_{\zeta,j} \frac{\alpha_{\zeta,j}^h \alpha_{\beta,k}^n E\left(A^{h+n}\right)}{\alpha_{\zeta,j}^{h+1} \alpha_{\beta,k}^{n-1} E\left(A^{h+n}\right)} = \alpha_{\beta,k}$$

and recover the general ability loadings on behavioral tests.

Without loss of generality, take any two tests, for example a cognitive and a behavioral one, and form

$$\frac{\zeta_j}{\alpha_{\zeta,j}} = \left[C \frac{\pi_{\zeta,j}}{\alpha_{\zeta,j}} + \frac{\varepsilon_{\zeta,j}}{\alpha_{\zeta,j}}\right] + A,$$

[11] Given that A, B, and C are all latent, these normalizations imply no restriction since $A\alpha_{\zeta,j} = A\kappa \frac{\alpha_{\zeta,j}}{\kappa}$ for any constant κ.

$$\frac{\beta_k}{\alpha_{\beta,k}} = \left[B\frac{\phi_{\beta,k}}{\alpha_{\beta,k}} + \frac{\varepsilon_{\beta,k}}{\alpha_{\beta,k}} \right] + A.$$

Then, it follows from a Theorem by Kotlarski (1967) that the distribution of A (and of $\left[C\frac{\pi_{\zeta,j}}{\alpha_{\zeta,j}} + \frac{\varepsilon_{\zeta,j}}{\alpha_{\zeta,j}}\right]$ and $\left[B\frac{\phi_{\beta,k}}{\alpha_{\beta,k}} + \frac{\varepsilon_{\beta,k}}{\alpha_{\beta,k}}\right]$) is nonparametrically identified[12].

With all of the parameters associated with general ability A as well as its distribution identified, we can then take the system of cognitive tests and form

$$E\left((\zeta_j)^h (\zeta_k)^n \right) - \alpha_{\zeta,j}^h \alpha_{\zeta,k}^n E\left(A^{h+n}\right) = \pi_{\zeta,j}^h \pi_{\zeta,k}^n E\left(C^{h+n}\right),$$

for any $j \neq k$ with $j, k = 1, ..., N_c$. By forming

$$\frac{E\left((\zeta_1)^h (\zeta_k)^n \right) - \alpha_{\zeta,1}^h \alpha_{\zeta,k}^n E\left(A^{h+n}\right)}{E\left((\zeta_1)^{h+1} (\zeta_k)^{n-1} \right) - \alpha_{\zeta,1}^{h+1} \alpha_{\zeta,k}^{n-1} E\left(A^{h+n}\right)} = \frac{\pi_{\zeta,k}^n E\left(C^{h+n}\right)}{\pi_{\zeta,k}^{n-1} E\left(C^{h+n}\right)} = \pi_{\zeta,k}$$

we can recover $\pi_{\zeta,k}$ for all $k = 2, ..., N_c$. By iteratively applying the Kotlarski argument, we can nonparametrically recover the distributions of C and $\varepsilon_{\zeta,j}$ for all $j = 1, ..., N_c$. Finally, by applying the same argument to the system of behavioral tests we can recover $\phi_{\beta,j}$ and the nonparametric distributions of B and $\varepsilon_{\beta,j}$ for all $j = 1, ..., N_b$.

3.3 Empirical Evidence

We estimate the model described in the previous section on a cohort of children who started kindergarten in the 1998-99 school year from the ECLS-K survey. The ECLS-K is a panel survey of kids starting with the 1998-99 kindergarten cohort. The survey was applied again in the 1999-2000, 2001-02 and 2003-04 school years. It contains a rich set of covariates that include characteristics of the children, the family, the class and the school. More importantly, participants in the ECLS-K survey were given routine cognitive tests measuring general knowledge, reading, math and science skills as well as containing ratings on behavioral and social skills done by both parents and teachers. Finally, the ECLS-K survey contains information on the retention policies used by each school in all survey years. These variables are ideal to be used as exclusions in the sense that they do not determine the child's test score directly (conditional on the other X's) but they do affect the probability that a kid will be forced to repeat a grade. Table 1

[12] Intuitively, given the now known $\alpha_{\zeta,j}$ and $\alpha_{\beta,k}$, we can identify all of the moments of general ability A from equation (3). Since we can recover all moments of the random variable A we can, for all practical purposes (see Casella and Berger (2002) for conditions), recover its distribution.

shows descriptive statistics for the 28 covariates we include in all our equations as well as the scores and retention policy variables.

We take advantage of the structure of the ECLS-K and use the 1998-99 T-scores on the general knowledge, math and reading tests as our markers of cognitive ability. The T-scores in ECLS-K are standardized scores providing norm-referenced measurements of achievement, that is, estimates of achievement relative to the population as a whole, in this case the cohort the individual belongs to. As measurements for behavioral ability we use the teacher Social Rating Scale (SRS). The SRS measures consist of reports done by teachers on how often students exhibited certain social skills and behaviors. In particular we use the approaches to learning, self-control and interpersonal skills components of the scale.

Following our discussion of identification in section 3.2 we impose the following normalizations. We normalize the general ability loading on the general knowledge test to 1, so A can be interpreted as a trait that is associated positively with higher scores in the general knowledge test. The loading on cognitive ability is normalized to 1 on the math test, so C is associated with higher math scores. Finally, we normalize the behavioral loading on the self-control marker to 1. If we let $\zeta_{i,j,1}$ be our j^{th} cognitive measure for individual i in period 1 (kindergarten) and similarly for behavioral measures, our kindergarten measures are modeled as

$$\zeta_{i,j,1} = X_{i,1}\gamma_{\zeta,j,1} + A_i\alpha_{\zeta,j,1} + C_i\pi_{\zeta,j,1} + \varepsilon_{\zeta,j,1} \tag{4}$$

and

$$\beta_{i,j,1} = X_{i,1}\gamma_{\beta,j,1} + A_i\alpha_{\beta,j,1} + B_i\phi_{\beta,j,1} + \varepsilon_{\beta,j,1}. \tag{5}$$

We restrict the sample to kids who were retained only once and did not skip grades. Because of the nature of the survey, we are able to form 3 different retention indicators: kindergarten, early (1^{st} or 2^{nd} grades) and late (3^{rd} and 4^{th} grades)[13]. That is, our dynamic treatment time indicator takes values $T = 0, 1, 2, 3$ where $T = 0$ means the kid was not retained, $T = 1$ means he is retained in kindergarten, $T = 2$ means he is retained early and $T = 3$ he is retained late.

As our outcome of interest, we use the 2003-04 math and reading T-scores. That is, we look at the effect of retention on period 4 test scores. Since the survey is applied

[13]In principle we could form all separate early and late into the four grades. This, however, can only be done for less than half of the sample.

every two years, period 4 scores include 5^{th} graders (if they are not retained) and 4^{th} graders (if they were retained at some point). In terms of our discussion in section 3.1, we focus on the effects on retention keeping age fixed. Our model for period 4 scores is given by

$$\begin{aligned}\zeta_{i,j,4} &= X_{i,4}\gamma_{\zeta,j,4} + \textstyle\sum_{\tau=1}^{3} D(\tau)\left[\Phi_{\tau,4} + A_i\left[\alpha_{\zeta,j,\tau,4} - \alpha_{\zeta,j,4}\right] + C_i\left[\pi_{\zeta,j,\tau,4} - \pi_{\zeta,j,4}\right]\right] \\ &\quad + A_i\alpha_{\zeta,j,4} + C_i\pi_{\zeta,j,4} + \varepsilon_{\zeta,j,4}\end{aligned}. \quad (6)$$

This corresponds to the dynamic heterogeneous effects specification of section 2.1.4.

3.3.1 Preliminary Evidence

In order to better understand the stylized facts in the data, we first run a series of regressions as preliminary evidence of selection, dynamic selection and the potential for dynamic effects of retention. In table 2, we regress each of the kindergarten cognitive tests on indicators of whether the kid will be retained in the future. For all of our measures, children who will be retained are significantly different from kids who will not be retained. Furthermore, we reject the hypothesis that the effects of being retained at different grades are the same. We interpret these results not only as evidence that there is selection on grade retention but that this selection is dynamic: kids who are retained early are *ex-ante* different from kids who are retained late who are *ex-ante* different from kids who are not retained, even after conditioning on a rich set of observables[14].

In table 3 we present the regression of 4^{th} period (03-04 school year) math test scores on indicators of whether the kid was retained[15]. In general, being retained is associated with having worse outcomes. Furthermore, the effects of retention are consistent with either dynamic treatment effects, dynamic selection or both. That is, they are different depending on whether the child is retained in kindergarten, early or late. In columns 5 and 6 we use the kindergarten scores as proxies for ability in order to try to account for selection. Consistent with the idea that there is selection, the negative effects of retention become smaller but they are do not disappear. Furthermore, the effects are still different over time and we can reject the formal test of equality of the effects over time.

To the extent that one is willing to assume that controlling for the observed covari-

[14]Table A1 in the appendix shows a similar result is obtained for behavioral measures.

[15]Table A2 in the appendix performs a parallel analysis for period 4 reading scores.

ates and using kindergarten scores is enough to recover the coefficients on retention and interpret them as some interesting average treatment parameter, the evidence suggests that having kids repeat grades actually worsens their test scores. Furthermore, the effect of retention on math test scores worsens if the kid is retained after kindergarten.

While suggestive, the analysis in this section is not complete enough to distinguish between explanations or to allow us to distinguish whether the effects are heterogeneous unless we are willing to impose restrictive assumptions like homogeneous effects. In fact, under our interpretation of tests scores as noisy measures of true latent abilities, using the kindergarten measures as controls may actually worsen the bias[16]. In order to get a more complete picture, in the next section we estimate the full model of scores, multidimensional ability and retention selection.

3.3.2 Estimating a Multidimensional Model of Ability and Retention

In this section we present the results of jointly estimating the model described by equations (4)-(6) as well as a dynamic model of retention. The decision of whether to have a child repeat a grade is the solution to some complicated game being played between the parents, the teachers, the child and the school. While in principle we can think of modelling such a game, we choose to instead approximate it with a threshold crossing model as described in section 2. As shown in Heckman and Navarro (2007), this reduced form model is in fact non-parametrically identified.

The actual form of the model for retention we use is the following. We write the latent index V as[17]

$$V_{i,t} = \lambda_{0,t} + X_{i,t}\lambda_x + Z_{i,t}\lambda_z + A_i\rho_A + B_i\rho_B + C_i\rho_C + \upsilon_{i,t} \text{ for } t = \underline{\mathrm{T}}, ..., \bar{T}.$$

$D_i(t)$ would then be defined as

$$D_i(t) = \mathbf{1}\left(V_i(t) > 0 | \{V_i(\tau) \leq 0\}_{\tau=1}^{t-1}\right).$$

Notice that, consistent with our data, we allow for exclusions in the index. That is

[16]See Heckman and Navarro (2004)

[17]Since we know the latent index is nonparametrically identified, we could instead write it as a polynomial on the variables instead of a linear function. Given that the number of parameters we are estimating is already 351, and the number of parameters would increase considerably, we stick with the linear form.

we allow for some variables (Z) to be included in the retention equations but not in the outcomes. In the data this correspond to 4 binary measures of whether the school has a policy that allows children to be retained in any grade, to be retained because of immaturity, to be retained at the parents request and to be retained without parental authorization.

As shown in section 3.2 and in Heckman and Navarro (2007), the distributions of the unobservables $(A, B, C, \varepsilon, \upsilon)$ in the model are nonparametrically identified. For estimation purposes, however, we specify all of the distributions and allow them to follow mixtures of normals with either two or three components. Estimation of the 351 parameters in the model is done by maximum likelihood. Tables A3-A7 in the appendix contain the parameter estimates as well as p-values.

3.3.3 Model Fit

In tables 4 and 5 we present evidence of how well the estimated model fits the data. As shown in table 4 the model replicates the probabilities of retention in the data remarkably well. In fact, we cannot reject the hypothesis of equality of predicted and actual probabilities. In table 5 we show that the same holds true for the cognitive and behavioral measures. The measures in the data are normalized to have mean zero and variance equal to one and we cannot reject that these are the values predicted by the model.

3.3.4 Selection on Unobservable Abilities

Figures 1 through 3 present evidence for selection on the components of ability. In figure 1 we present the distribution of general ability by retention status. The graph shows that general ability is ordered according to retention. That is, the lower the general ability of a kid the more likely he is to be retained in a higher grade. Behavioral ability does not follow such a clear pattern in figure 2. If anything, kids with lower behavioral ability are more likely to be retained in 1^{st} and 2^{nd} grade. The distribution of behavioral ability is very similar for kids who are not retained and for those who are retained in higher grades. That is, behavioral aspects seem to play more of a role in the decision of whether to retain a kid in early grades, but the pattern is not very clear. The same pattern is found much more clearly in terms of cognitive ability. Kids who are retained in kindergarten and on early grades have lower cognitive ability than kids who are not retained or those retained at higher grades.

3.3.5 Average Treatment Parameters

In tables 6 and 7 we present both treatment on the treated (and the untreated) as well as unconditional average treatment parameters for both reading and math test scores on the 2003-04 school year[18]. In all cases we condition on age being 11 at test date so, as explained in section 3.1, our measure of gains includes the effect of the grade differential between not being retained therefore taking the test in the 5^{th} grade and having repeated a grade therefore taking the test while in the 4^{th} grade.

Grade retention is uniformly bad for test scores. In almost all cases, whether conditioning on actual retention status or looking at the average gain for a random individual, retention reduces both reading and math test scores. If we look at the average treatment parameters we can see that the effect of being retained in kindergarten versus not being retained is not statistically different from zero in both cases. However, picking a kid at random from the population and having him repeat 1^{st} or 2^{nd} grade would, on average, reduce tests scores by about 0.34 standard deviations. The reduction would be even larger if he repeated a higher grade.

Focusing on the math test scores on table 7 we can see that all treatment on the treated parameters (i.e. those measuring gains conditional on being retained) are negative. In all cases the test is reduced the most when retention occurs early. Being retained, whether conditional on being retained or being picked at random, reduces average test scores anywhere between a quarter and half a standard deviation. In terms of the effect of treatment for those kids who are not retained (treatment on the untreated), having a kid repeat kindergarten has, on average, no effect on test scores. However, repeating either 3^{rd} or 4^{th} grade would reduce math test scores by half a standard deviation.

3.3.6 Distributional Effects

One of the advantages of estimating the whole model and identifying the distributions of the components of ability is that we can perform a more detailed analysis. In figure 4 we present average reading test scores conditional on different quantiles of the general

[18]Tables A8 and A9 in the appendix present the test score levels from which these gains are formed.

and cognitive ability distributions[19]. For children with either the 10^{th} or median level of general ability, grade retention reduces average reading test scores. The average loss gets smaller as the level of cognitive ability increases.

A different story arises when we look at high general ability individuals, i.e. children with A fixed at the 90^{th} percentile of the distribution. While retention in early or late grades reduces reading test scores for high general ability children, repeating kindergarten would increase them.

4 Conclusion

In this paper we develop and apply a framework for the analysis of dynamic treatment effects. Our analysis of grade retention shows the usefulness of extending the standard static framework used to estimate treatment effects. Our results show that while, in general, repeating a grade reduces tests scores at age 11, being retained in 1^{st} or 2^{nd} grade is worse for kids who are retained. We also show that, for the kids who are not retained, repeating a grade can go from having no effect if it occurs in kindergarten to reducing tests scores by half a standard deviation at higher grades.

The analysis in this paper could clearly be applied in other situations like treatments associated with health status indicators and/or costs of particular treatments. One could also analyze the effects of advertisement at different stages in the life of a product, the effect of attending a segregated school at different stages, etc.

The framework we develop can be thought of as a midpoint between the standard reduced form static treatment literature and a fully specified structural dynamic discrete choice model. In many situations it is not clear how to specify the selection process and our analysis provides a reduced form alternative (with all the advantages and problems associated with it). Furthermore, since extending it to the case in which treatment is not an absorbing state is straightforward (by letting treatment occur not only the first time a threshold is crossed but also the second, third, etc) it can be applied in more complicated situations.

[19] Figure A10 in the appendix presents the equivalent graphs for math test scores.

References

Abbring, J. H. and G. J. Van den Berg (2003, September). The nonparametric identification of treatment effects in duration models. *Econometrica 71*(5), 1491–1517.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422. 2001 Lawrence R. Klein Lecture.

Casella, G. and R. L. Berger (2002). *Statistical Inference* (2nd Edition ed.). Duxbury Press.

Cunha, F., J. J. Heckman, and S. Navarro (2005, April). Separating uncertainty from heterogeneity in life cycle earnings, the 2004 Hicks lecture. *Oxford Economic Papers 57*(2), 191–261.

Gill, R. D. and J. M. Robins (2001, December). Causal inference for complex longitudinal data: The continuous case. *The Annals of Statistics 29*(6), 1785–1811.

Hauser, R. (1999). Should we end social promotion? truth and consequences. 99-06 CDE Working Papers. Center for Demography and Ecology: The University of Wisconsin-Madison.

Heckman, J. J. and S. Navarro (2004, February). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics 86*(1), 30–57.

Heckman, J. J. and S. Navarro (2007, February). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Jacob, B. and L. Lefgren (2004, February). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics 86*(1), 226–244.

Jöreskog, K. G. and A. S. Goldberger (1975, September). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association 70*(351), 631–639.

Kotlarski, I. I. (1967). On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics 20*, 69–76.

Matzkin, R. L. (1992, March). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica 60*(2), 239–270.

Matzkin, R. L. (2003, September). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Murphy, S. A. (2003, May). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B 65*(2), 331–366.

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

Robins, J. M. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A. Mulley (Eds.), *Health Services Research Methodology: A Focus on AIDS*, pp. 113–159. Rockville, MD: U.S. Department of Health and Human Services, National Center for Health Services Research and Health Care Technology Assessment.

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

Schennach, S. M. (2004, January). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Table 1: Summary Statistics

	1998-99 School Year					2003-04 School Year				
	# of Observations	Mean	Standard Deviation	Minimum	Maximum	# of Observations	Mean	Standard Deviation	Minimum	Maximum
Male	7822	0.50	0.50	0	1	2741	0.48	0.50	0	1
White	7822	0.64	0.48	0	1	2741	0.73	0.44	0	1
Black	7822	0.13	0.33	0	1	2741	0.08	0.27	0	1
Hispanic	7822	0.13	0.34	0	1	2741	0.11	0.31	0	1
Body Mass Index	7822	16.24	2.19	8.28	32.52	2741	20.28	4.62	10.78	42.2
Age	7822	5.64	0.35	4.83	6.92	2741	11.10	0.36	10.25	12.25
Age squared	7822	31.98	4.04	23.36	47.84	2741	123.38	7.92	105.06	150.06
Number of Siblings	7822	1.44	1.13	0	11	2741	1.50	1.07	0	12
Family Income	7822	6.98	3.15	0	12	2741	7.87	2.91	0	12
Nonenglish Spoken at Home	7822	0.10	0.30	0	1	2741	0.02	0.13	0	1
TV Rule at Home	7822	0.89	0.32	0	1	2741	0.90	0.30	0	1
Mother in Household	7822	0.02	0.13	0	1	2741	0.02	0.15	0	1
Father in Household	7822	0.19	0.39	0	1	2741	0.16	0.37	0	1
Mother's Education	7822	3.37	1.78	0	8	2741	3.73	1.78	0	8
Fatehr's Education	7822	2.90	2.27	0	8	2741	3.25	2.34	0	8
Number of Kids in Class	7822	20.53	5.32	1	52	2741	20.86	4.19	8	34
Number of Kids in Class squared	7822	449.68	249.78	1	2704	2741	452.64	180.15	64	1156
Teacher's Rating of Class Behavior	7822	1.81	1.94	0	14	2741	1.45	0.99	0	14
Percentage of Minority Students (categorical)	7822	1.44	1.48	0	4	2741	1.26	1.39	0	4
Public School	7822	0.78	0.41	0	1	2741	0.77	0.42	0	1
School's Average Daily Atendance	7822	3.29	1.20	0	5	2741	3.51	0.95	0	5
TT1 Funds Received by School	7822	0.63	0.48	0	1	2741	0.61	0.49	0	1
Crime a Problem	7822	0.45	0.57	0	2	2741	0.31	0.51	0	2
Students Bring Weapons	7822	0.18	0.38	0	1	2741	0.10	0.31	0	1
Children or Teachers Physically Attacked	7822	0.37	0.48	0	1	2741	0.26	0.44	0	1
Security Measures in School	7822	0.57	0.50	0	1	2741	0.79	0.41	0	1
Parents Involved in School Activities	7822	2.95	0.90	0	4	2741	3.09	0.92	0	4
Policy: Retained for any Reason	7738	0.93	0.25	0	1	2217	0.91	0.29	0	1
Policy: Retained for Immaturity	7186	0.76	0.43	0	1	2209	0.86	0.35	0	1
Policy: Retained by Parents' Request	7186	0.75	0.43	0	1	2180	0.82	0.39	0	1
Policy: Retained without Parents' Permission	7186	0.45	0.50	0	1	2167	0.50	0.50	0	1

Source: ECLS-K Longitudinal Kindergarten-Fifth Grade Public-Use Data File

Note: For our counter-factual analyses, we only use data on kids whose covariates and retention history are observable (i.e. not missing) for all time periods. Thus, we end up with much fewer observations at the 2003-04 school year.

Table 2: Regression of Kindergarten (1998-99 School Year) Cognitive Scores+

Dependent Variable	Kindergarten General Knowledge Score				Kindergarten Reading Score				Kindergarten Math Score			
Retained in Kindergarten	-0.53*	-0.47*	-0.49*	-0.33*	-0.61*	-0.54*	-0.56*	-0.44*	-0.69*	-0.63*	-0.66*	-0.52*
Retained Early (1st or 2nd grade)	-0.84*	-0.63*	-0.61*	-0.50*	-0.97*	-0.73*	-0.72*	-0.64*	-1.12*	-0.90*	-0.88*	-0.79*
Retained Late (3rd or 4th grade)	-0.71*	-0.44*	-0.42*	-0.37*	-0.72*	-0.41*	-0.41*	-0.38*	-0.88*	-0.60*	-0.59*	-0.55*
Child's Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Characteristics	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Age and Age Squared	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
No. of Observations	7427	7427	7427	7427	7441	7441	7441	7441	7723	7723	7723	7723
P-value for KI = EA = LA++	0.000	0.052	0.102	0.032	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
P-value for KI = EA	0.000	0.025	0.084	0.009	0.000	0.001	0.006	0.000	0.000	0.000	0.000	0.000
P-value for EA = LA	0.240	0.081	0.067	0.199	0.004	0.000	0.001	0.003	0.007	0.001	0.001	0.004
P-value for KI = LA	0.074	0.743	0.454	0.689	0.242	0.157	0.097	0.471	0.031	0.735	0.450	0.712
R squared	0.199	0.321	0.333	0.394	0.105	0.249	0.261	0.295	0.169	0.2957	0.3081	0.3578

+ Scores are standardized to have mean zero and variance equal to one

++ KI, EA, and LA stand for the coefficient of the dummy variable for "retained in kindergarten", "retained early", and "retained late", respectively.

* Statistically significant at 5% level

Note: If the p value is small compared to the critical value, we reject the hypothesis of equality of coefficients. P values less than 0.05 are colored with yellow. Yes/No tells us if each group of variables is included as controls.

Table 3: Regression of Math Score of 2003-04 School Year+

Dependent Variable	Math Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Retained in Kindergarten	-0.99*	-0.84*	-0.83*	-0.80*	-0.21*	-0.19*
Retained Early (1st or 2nd grade)	-1.12*	-0.92*	-0.92*	-0.91*	-0.51*	-0.50*
Retained Late (3rd or 4th grade)	-1.01*	-0.75*	-0.75*	-0.74*	-0.48*	-0.47*
Child's Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	No	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	No	Yes	Yes	Yes	Yes
Age and Age Squared	No	No	No	Yes	No	No
Kindergarten Cognitive Tests	No	No	No	No	Yes	Yes
Kindergarten Behavioral Measures	No	No	No	No	No	Yes
No. of Observations	7652	7652	7652	7652	3809	3782
P-value for KI = EA = LA	0.163	0.097	0.102	0.071	0.015	0.011
P-value for KI = EA	0.096	0.298	0.269	0.166	0.005	0.003
P-value for EA = LA	0.166	0.034	0.037	0.030	0.758	0.730
P-value for KI = LA	0.843	0.337	0.387	0.492	0.030	0.029
R squared	0.182	0.287	0.298	0.300	0.549	0.551

+ Scores are standardized to have mean zero and variance equal to one

++ KI, EA, and LA stand for the coefficient of the dummy variable for "retained in kindergarten", "retained early", and "retained late", respectively.

* Statistically significant at 5% level

Note: If the p value is small compared to the critical value, we reject the hypothesis of equality of coefficients. P values less than 0.05 are colored with yellow. Yes/No tells us if each group of variables is included as controls.

Table 4: Predicted and Actual Retention Probabilities (Conditional on Survival)*

	Data	Model	
		Predicted	Standard Error
Retained in Kindergarten	5.38%	5.35%	0.24%
Retained Early (1st or 2nd grade)	3.80%	4.07%	0.22%
Retained Late (3rd or 4th grade)	1.21%	1.23%	0.11%

Note: The table caclulates the probability of retention at t, conditional on not having been retained before t. Standard Errors obtained via 200 bootstrap replications. In the simulation, we keep age of a kid in the 1998-99 school year fixed at 6, age of a kid in the 1999-2000 school year fixed at 7, and age of a kid in the 2001-02 school year fixed at 9, respectively.

Table 5: Predicted Mean and Variance of Kindergarten (1998-99 School Year) Test Scores/Measures

All Measures Normalized to have mean zero and variance equal to one

Test / Measure	Mean	Variance
General Test	-0.003 (0.012)	1.001 (0.005)
Reading Test	-0.002 (0.015)	0.994 (0.007)
Math Test	-0.002 (0.015)	1.019 (0.008)
Approach to Learning	-0.013 (0.018)	1.007 (0.008)
Self-Control	-0.020 (0.018)	1.020 (0.008)
Interpersonal Skills	-0.021 (0.018)	1.020 (0.008)

Note: Bootstraped standard errors in parenthesis, 200 replications.

Table 6: Average Reading Test Score Gain (in standard deviations) by Retention Status: 2003-04 School Year

Average Gain	A kid who is actually (i.e. conditional on the retention status being:)				ATE (unconditional)
	Not Retained	Retained in Kindergarten	Retained Early	Retained Late	
Retained in Kindergarten vs Not Retained	-0.015	-0.279*	-0.372*	-0.365*	-0.041
Retained Early vs Not Retained	-0.331*	-0.491*	-0.535*	-0.519*	-0.346*
Retained Late vs Not Retained	-0.409*	-0.363*	-0.321*	-0.299*	-0.403*

* Statistically different from zero at 5% level

Note: Let T = 0,1,2, or 3 represent the actual retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively. Let S(i) be the potential test score if the kid were retained at time i=0,1,2,3. The row i, column j element of this table calculates E[S(i) - S(0) | D=j]. For example, the test score of a kid who was actually not retained would decrease by 0.331 standard deviations if he were retained at 1 or 2 grade instead. When calculating these figures, we keep kid's age fixed at 11.

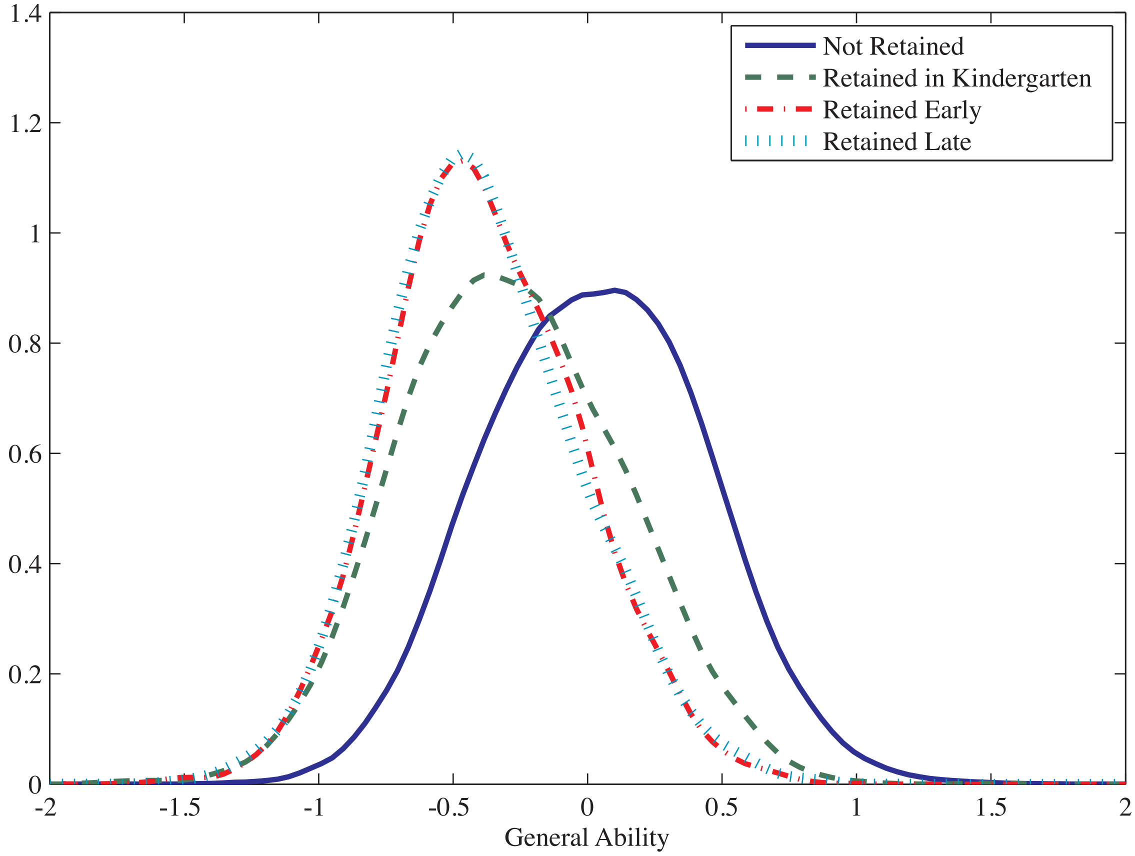
Table 7: Average Math Test Score Gain (in standard deviations) by Retention Status: 2003-04 School Year

Average Gain	A kid who is actually (i.e. conditional on the retention status being)				ATE (unconditional)
	Not Retained	Retained in Kindergarten	Retained Early	Retained Late	
Retained in Kindergarten vs Not Retained	0.013	-0.203*	-0.255*	-0.228*	-0.007
Retained Early vs Not Retained	-0.337*	-0.442*	-0.468*	-0.455*	-0.347*
Retained Late vs Not Retained	-0.517*	-0.371*	-0.318*	-0.322*	-0.503*

* Statistically different from zero at 5% level

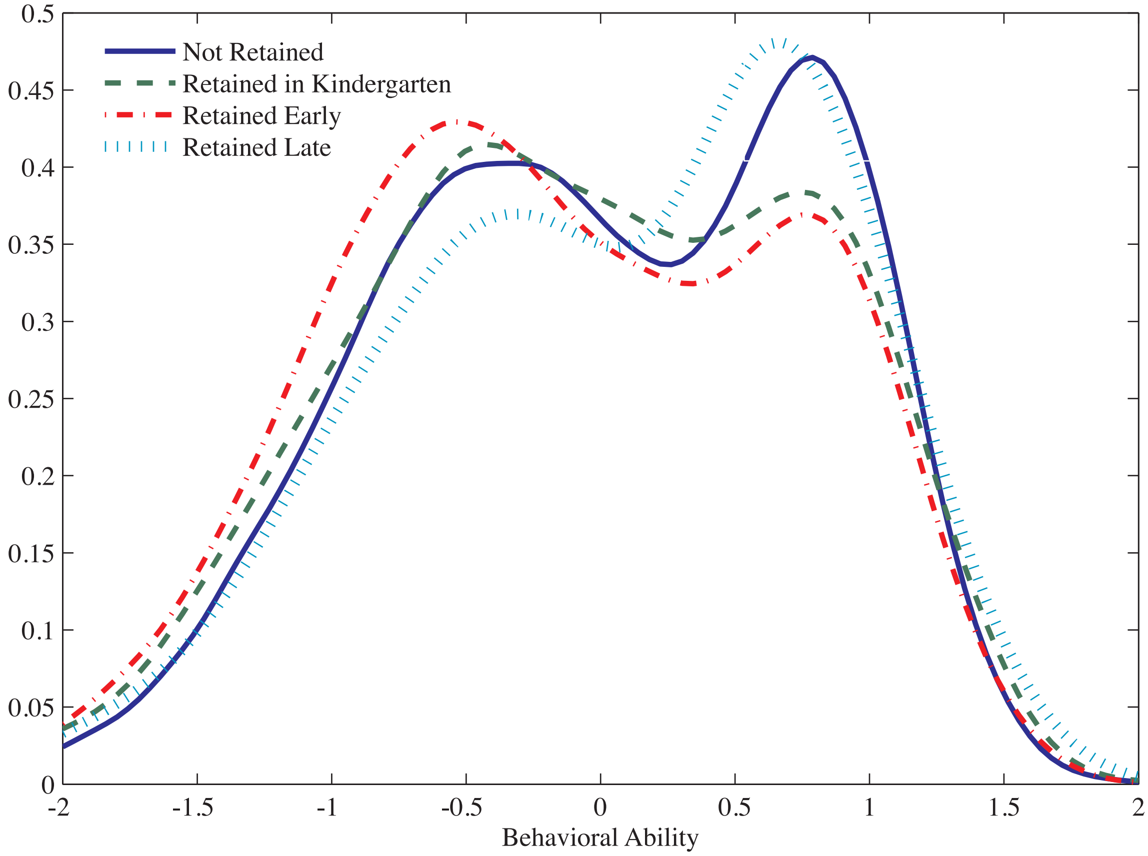
Note: Let T = 0,1,2, or 3 represent the actual retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively. Let S(i) be the potential test score if the kid were retained at time i=0,1,2,3. The row i, column j element of this table calculates E[S(i) - S(0) | D=j]. For example, the test score of a kid who was actually not retained would decrease by 0.337 standard deviations if he were retained at 1 or 2 grade instead. When calculating these figures, we keep kid's age fixed at 11.

Figure 1
Density of General Ability Conditional on Retention Status



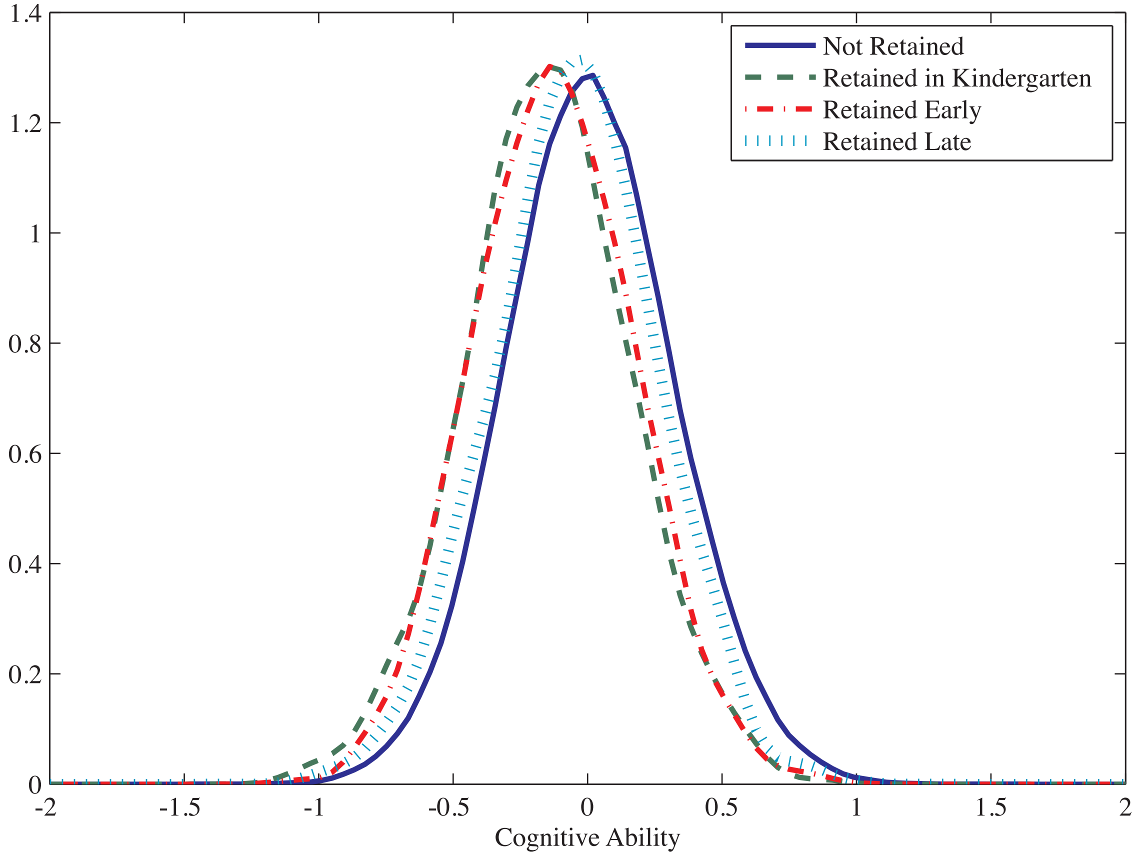
Let f(A) denote the probability density function of cognitive ability. We assume that f(A) is a mixture of normals. Let T=0,1,2,3 denote retention status: not retained, retained in kindergarten, retained early (1 or 2 grade) and retained late (3 or 4). The graph shows f(A|T=t) for each retention status.

Figure 2
Density of Behavioral Ability Conditional on Retention Status



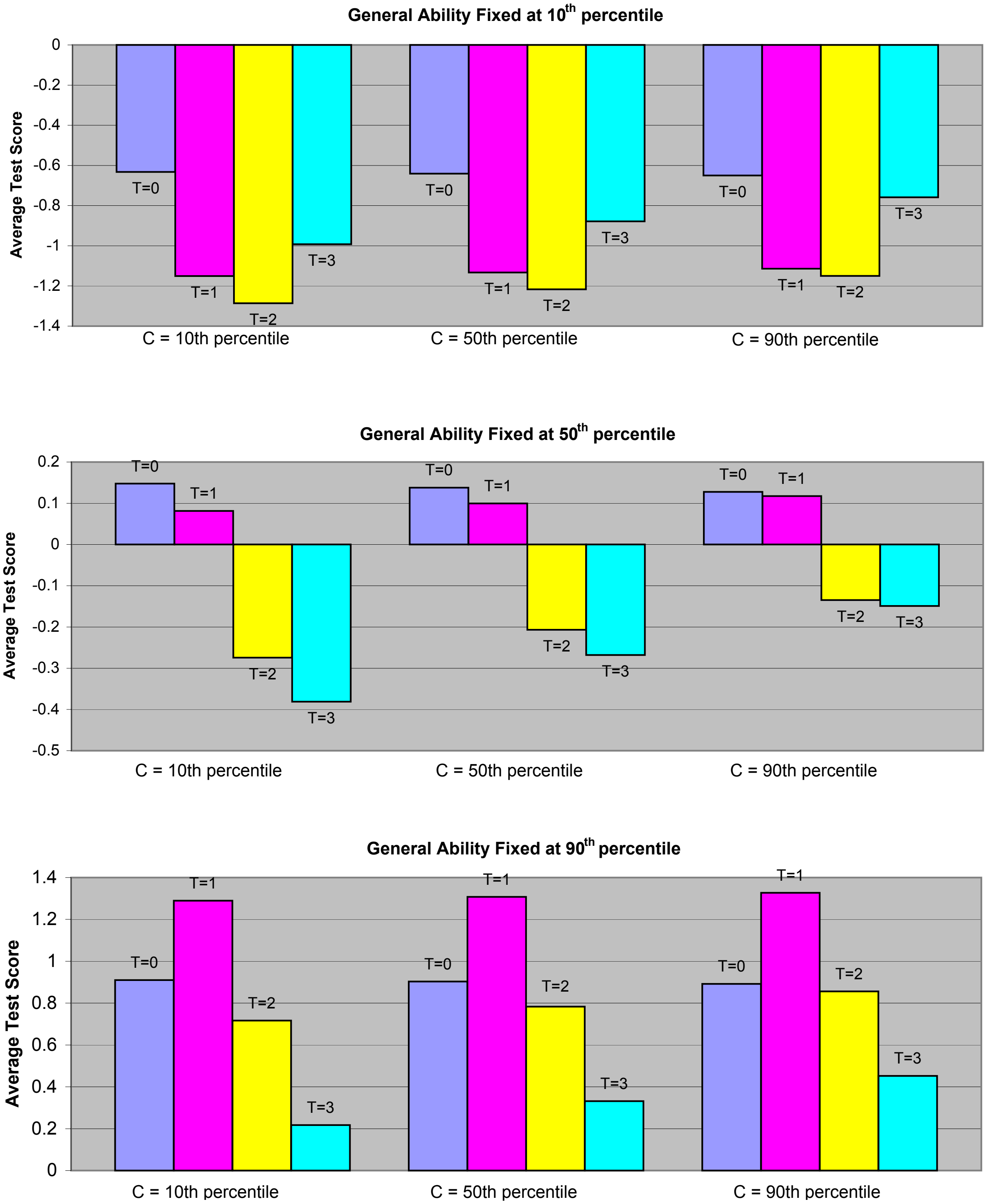
Let f(B) denote the probability density function of cognitive ability. We assume that f(B) is a mixture of normals. Let T=0,1,2,3 denote retention status: not retained, retained in kindergarten, retained early (1 or 2 grade) and retained late (3 or 4). The graaph shows f(B|T=t) for each retention status.

Figure 3
Density of Cognitive Ability Conditional on Retention Status



Let f(C) denote the probability density function of cognitive ability. We assume that f(C) is a mixture of normals. Let T=0,1,2,3 denote retention status: not retained, retained in kindergarten, retained early (1 or 2 grade) and retained late (3 or 4). The graph shows f(C|T=t) for each retention status.

Figure 4: Average Reading Score at 2003-04 School Year for Different Percentiles of the Cognitive Ability Distribution, Conditonal on Retention Status and General Ability.



T=0,1,2 and 3 are "not retained", "retained in kindergarten", "retained early (at 1 or 2 grade)", and "retained late (at 3 or 4 grade)", respectively. When calculating these values, we keep age of children fixed at 11. Scores are standardized.

Table A1: Regression of Kindergarten (1998-99 School Year) Behavioral Measures[+]

Dependent Variable	Approach to Learning				Self-Control				Interpersonal Skills			
Retained in Kindergarten	-0.54*	-0.51*	-0.51*	-0.43*	-0.24*	-0.23*	-0.22*	-0.20*	-0.28*	-0.26*	-0.25*	-0.22*
Retained Early (1st or 2nd grade)	-0.86*	-0.78*	-0.78*	-0.73*	-0.36*	-0.32*	-0.33*	-0.31*	-0.46*	-0.40*	-0.41*	-0.39*
Retained Late (3rd or 4th grade)	-0.40*	-0.28*	-0.29*	-0.27*	-0.15	-0.08	-0.08	-0.08	-0.23*	-0.14	-0.16	-0.15
Child's Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Characteristics	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Age and Age squared	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
No. of Observations	7818	7818	7818	7818	7802	7802	7802	7441	7769	7769	7769	7769
P-value for KI = EA = LA	0.000	0.000	0.000	0.000	0.163	0.127	0.118	0.123	0.027	0.050	0.046	0.041
P-value for KI = EA	0.000	0.000	0.000	0.000	0.142	0.227	0.178	0.148	0.013	0.049	0.036	0.025
P-value for EA = LA	0.000	0.000	0.000	0.000	0.094	0.049	0.050	0.060	0.068	0.041	0.050	0.061
P-value for KI = LA	0.249	0.050	0.066	0.165	0.435	0.211	0.249	0.316	0.693	0.348	0.434	0.553
R squared	0.115	0.136	0.140	0.155	0.057	0.068	0.073	0.075	0.062	0.0739	0.0796	0.0829

[+] Scores are standardized to have mean zero and variance equal to one

[++] KI, EA, and LA stand for the coefficient of the dummy variable for "retained in kindergarten", "retained early", and "retained late", respectively.

* Statistically significant at 5% level

Note: If the p value is small compared to the critical value, we reject the hypothesis of equality of coefficients. P values less than 0.05 are colored with yellow. Yes/No tells us if each group of variables is included as controls.

Table A2: Regression of Reading Score for 2003-04 School Year[+]

Dependent Variable	Reading Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Retained in Kindergarten	-1.00*	-0.83*	-0.84*	-0.80*	-0.38*	-0.37*
Retained Early (1st or 2nd grade)	-1.19*	-0.96*	-0.94*	-0.93*	-0.61*	-0.59*
Retained Late (3rd or 4th grade)	-1.01*	-0.72*	-0.71*	-0.70*	-0.42*	-0.40*
Child's Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	No	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	No	Yes	Yes	Yes	Yes
Age and Age Squared	No	No	No	Yes	Yes	Yes
Kindergarten Cognitive Tests	No	No	No	No	Yes	Yes
Kindergarten Behavioral Measures	No	No	No	No	No	Yes
No. of Observations	7646	7646	7646	7646	3806	3779
P-value for KI = EA = LA	0.031	0.030	0.036	0.028	0.103	0.124
P-value for KI = EA	0.027	0.127	0.194	0.111	0.057	0.071
P-value for EA = LA	0.046	0.012	0.012	0.012	0.151	0.151
P-value for KI = LA	0.943	0.271	0.210	0.302	0.800	0.867
R squared	0.180	0.310	0.323	0.326	0.512	0.5103

[+] Scores are standardized to have mean zero and variance equal to one

[++] KI, EA, and LA stand for the coefficient of the dummy variable for "retained in kindergarten", "retained early", and "retained late", respectively.

* Statistically significant at 5% level

Note: If the p value is small compared to the critical value, we reject the hypothesis of equality of coefficients. P values less than 0.05 are colored with yellow. Yes/No tells us if each group of variables is included as controls.

Table A3: Estimates of Parameters in the Outcome Equations

1998-99 School Year (Kindergarden)

	General Knowledge Scores		Reading Scores		Math Scores	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Constant	-18.756	0.000	-10.740	0.000	-17.493	0.000
Male	-0.004	0.423	-0.251	0.000	-0.077	0.000
White	0.282	0.000	0.000	0.926	0.024	0.000
Black	-0.318	0.000	-0.160	0.000	-0.296	0.000
Hispanic	0.006	0.609	-0.242	0.000	-0.282	0.000
Body Mass Index	0.008	0.000	-0.013	0.000	-0.008	0.000
Age	5.517	0.000	3.075	0.000	5.261	0.000
Age squared	-0.417	0.000	-0.218	0.000	-0.399	0.000
Number of Siblings	-0.075	0.000	-0.101	0.000	-0.057	0.000
Family Income	0.044	0.000	0.037	0.000	0.038	0.000
Nonenglish Spoken at Home	-0.379	0.000	-0.227	0.000	-0.142	0.000
TV Rule at Home	0.112	0.000	0.063	0.000	0.005	0.056
Mother in Household	0.180	0.000	0.135	0.000	0.049	0.019
Father in Household	0.171	0.000	0.130	0.000	0.161	0.000
Mother's Education	0.084	0.000	0.083	0.000	0.075	0.000
Fatehr's Education	0.056	0.000	0.060	0.000	0.063	0.000
Number of Kids in Class	-0.001	0.000	0.011	0.000	0.017	0.000
Number of Kids in Class squared	0.000	0.000	0.000	0.000	0.000	0.000
Teacher's Rating of Class Behavior	-0.008	0.000	-0.018	0.000	-0.007	0.000
Percentage of Minority Students (categorical)	-0.045	0.000	0.031	0.000	-0.001	0.727
Public School	-0.151	0.000	-0.146	0.000	-0.194	0.000
School's Average Daily Atendance	-0.002	0.032	0.006	0.000	0.000	0.546
TT1 Funds Received by School	-0.071	0.000	-0.145	0.000	-0.114	0.000
Crime a Problem	-0.002	0.770	-0.038	0.000	-0.015	0.001
Students Bring Weapons	-0.021	0.023	-0.095	0.000	-0.088	0.000
Children or Teachers Physically Attacked	0.031	0.000	-0.004	0.458	0.016	0.003
Security Measures in School	0.004	0.423	0.039	0.000	0.034	0.000
Parents Involved in School Activities	0.029	0.000	0.038	0.000	0.044	0.000
Cognitive Ability	0.453	0.000	0.967	0.000	1.000	--
General Ability	1.000	--	1.354	0.000	1.533	0.000
Mean of the 1st Mixture Component	0.025	0.000	0.304	0.000	-0.023	0.000
Variance of the 1st Mixture Component	0.401	0.000	0.800	0.000	0.170	0.000
Weight of the 1st Mixture Component	0.968	0.000	0.190	0.000	0.859	0.000
Mean of the 2nd Mixture Component	-0.738	0.000	-0.071	0.000	0.141	0.000
Variance of the 2nd Mixture Component	0.935	0.000	0.192	0.000	0.540	0.000
Weight of the 2nd Mixture Component	0.032	0.000	0.810	0.000	0.141	0.000

-- represents the value is not applicable because the variable is normalized to one.

Note: Standard errors are obtained via 200 bootstrap replications.

Table A4: Estimates of Parameters in the Outcome Equations

1998-99 School Year (Kindergarden)

	Approach to Learning		Self-Control		Interpersonal Skills	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Constant	-15.437	0.000	-6.529	0.000	-8.795	0.000
Male	-0.433	0.000	-0.309	0.000	-0.346	0.000
White	0.035	0.000	0.027	0.000	0.123	0.000
Black	-0.176	0.000	-0.157	0.000	-0.069	0.000
Hispanic	-0.100	0.000	0.018	0.035	0.050	0.000
Body Mass Index	-0.020	0.000	-0.017	0.000	-0.015	0.000
Age	5.189	0.000	2.288	0.000	3.031	0.000
Age squared	-0.425	0.000	-0.193	0.000	-0.255	0.000
Number of Siblings	-0.017	0.000	0.022	0.000	-0.008	0.000
Family Income	0.022	0.000	0.016	0.000	0.017	0.000
Nonenglish Spoken at Home	0.039	0.002	0.075	0.000	-0.028	0.010
TV Rule at Home	0.006	0.073	-0.027	0.000	-0.002	0.454
Mother in Household	-0.131	0.000	-0.070	0.000	0.024	0.206
Father in Household	-0.014	0.093	-0.003	0.706	-0.016	0.029
Mother's Education	0.027	0.000	-0.009	0.000	0.009	0.000
Fatehr's Education	0.027	0.000	0.029	0.000	0.025	0.000
Number of Kids in Class	-0.020	0.000	-0.017	0.000	-0.015	0.000
Number of Kids in Class squared	0.000	0.000	0.000	0.000	0.000	0.000
Teacher's Rating of Class Behavior	-0.016	0.000	-0.014	0.000	-0.016	0.000
Percentage of Minority Students (categorical)	0.028	0.000	0.000	0.970	0.040	0.000
Public School	0.094	0.000	0.135	0.000	0.132	0.000
School's Average Daily Atendance	-0.005	0.000	-0.001	0.131	-0.005	0.000
TT1 Funds Received by School	0.037	0.000	0.014	0.003	0.003	0.456
Crime a Problem	0.021	0.000	0.016	0.000	0.020	0.000
Students Bring Weapons	-0.038	0.000	-0.023	0.002	-0.048	0.000
Children or Teachers Physically Attacked	-0.038	0.000	-0.033	0.000	-0.096	0.000
Security Measures in School	0.024	0.000	0.027	0.000	0.007	0.150
Parents Involved in School Activities	0.024	0.000	0.039	0.000	0.021	0.000
General Ability	1.012	0.000	0.485	0.000	0.584	0.000
Behavioral Ability	0.774	0.000	1.000	--	1.001	0.000
Mean of the 1st Mixture Component	0.121	0.000	0.074	0.000	-0.448	0.000
Variance of the 1st Mixture Component	0.226	0.000	0.161	0.000	0.359	0.000
Weight of the 1st Mixture Component	0.643	0.000	0.671	0.000	0.170	0.000
Mean of the 2nd Mixture Component	-0.218	0.000	-0.150	0.000	0.092	0.000
Variance of the 2nd Mixture Component	0.451	0.000	0.490	0.000	0.185	0.000
Weight of the 2nd Mixture Component	0.357	0.000	0.329	0.000	0.830	0.000

-- represents the value is not applicable because the variable is normalized to one.

Note: Standard errors are obtained via 200 bootstrap replications.

Table A5: Estimates of Parameters in the Outcome Equations

2003-04 School Year (Kindergarden)

	Reading Score		Math Score	
	Coefficient	P-value	Coefficient	P-value
Constant	-44.950	0.000	-51.929	0.000
Male	-0.143	0.000	0.192	0.000
White	0.105	0.000	0.002	0.626
Black	-0.434	0.000	-0.703	0.000
Hispanic	-0.099	0.000	-0.193	0.000
Body Mass Index	-0.001	0.000	-0.001	0.000
Age	7.851	0.000	9.041	0.000
Age squared	-0.348	0.000	-0.403	0.000
Number of Siblings	-0.057	0.000	0.001	0.690
Family Income	0.027	0.000	0.027	0.000
Nonenglish Spoken at Home	-0.530	0.000	-0.301	0.000
TV Rule at Home	0.044	0.000	-0.119	0.000
Mother in Household	0.259	0.000	0.200	0.000
Father in Household	0.266	0.000	0.202	0.000
Mother's Education	0.093	0.000	0.095	0.000
Fatehr's Education	0.067	0.000	0.070	0.000
Number of Kids in Class	0.015	0.000	0.047	0.000
Number of Kids in Class squared	0.000	0.000	-0.001	0.000
Teacher's Rating of Class Behavior	-0.024	0.000	-0.019	0.000
Percentage of Minority Students (categorical)	-0.018	0.000	0.005	0.073
Public School	-0.128	0.000	0.051	0.000
School's Average Daily Atendance	0.018	0.000	0.008	0.000
TT1 Funds Received by School	-0.116	0.000	-0.098	0.000
Crime a Problem	-0.008	0.365	0.026	0.001
Students Bring Weapons	0.077	0.000	0.091	0.000
Children or Teachers Physically Attacked	-0.039	0.000	-0.022	0.022
Security Measures in School	0.069	0.000	0.057	0.000
Parents Involved in School Activities	0.033	0.000	0.043	0.000
Retained in Kindergarten	-0.041	0.254	-0.007	0.819
Retained Early	-0.346	0.000	-0.347	0.000
Retained Late	-0.403	0.000	-0.503	0.000
Cognitive Ability if T=0	-0.023	0.272	0.094	0.000
General Ability if T=0	1.425	0.000	1.537	0.000
Cognitive Ability if T=1	0.046	0.611	0.445	0.000
General Ability if T=1	2.254	0.000	2.065	0.000
Cognitive Ability if T=2	0.174	0.001	0.259	0.000
General Ability if T=2	1.850	0.000	1.797	0.000
Cognitive Ability if T=3	0.291	0.031	0.064	0.237
General Ability if T=3	1.116	0.000	1.072	0.000
Mean of the 1st Distribution of the Mixture	0.248	0.000	0.000	0.767
Variance of the 1st Distribution of the Mixture	1.028	0.000	0.351	0.000
Weight of the 1st Distribution of the Mixture	0.028	0.000	0.999	0.000
Mean of the 2nd Distribution of the Mixture	-0.007	0.000	0.299	0.766
Variance of the 2nd Distribution of the Mixture	0.355	0.000	6.406	0.000
Weight of the 2nd Distribution of the Mixture	0.972	0.000	0.001	0.000

Note: Standard errors are obtained via 200 bootstrap replications. Let T = 0,1,2, or 3 represent the retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively.

Table A6: Parameter Estimates of Choice Equations

	Kindergarten		Early (1st or 2nd Grade)		Late (3rd or 4th Grade)	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Constant	6.779	0.000	7.773	0.000	8.350	0.000
Male	0.317	0.000	0.317	0.000	0.317	0.000
White	0.047	0.000	0.047	0.000	0.047	0.000
Black	0.349	0.000	0.349	0.000	0.349	0.000
Hispanic	0.079	0.000	0.079	0.000	0.079	0.000
Body Mass Index	-0.019	0.000	-0.019	0.000	-0.019	0.000
Age	-1.794	0.000	-1.794	0.000	-1.794	0.000
Age squared	0.081	0.000	0.081	0.000	0.081	0.000
Number of Siblings	0.063	0.000	0.063	0.000	0.063	0.000
Family Income	-0.029	0.000	-0.029	0.000	-0.029	0.000
Nonenglish Spoken at Home	0.029	0.397	0.029	0.397	0.029	0.397
TV Rule at Home	0.070	0.000	0.070	0.000	0.070	0.000
Mother in Household	0.177	0.000	0.177	0.000	0.177	0.000
Father in Household	-0.057	0.000	-0.057	0.000	-0.057	0.000
Mother's Education	-0.049	0.000	-0.049	0.000	-0.049	0.000
Fatehr's Education	-0.040	0.000	-0.040	0.000	-0.040	0.000
Number of Kids in Class	-0.032	0.000	-0.032	0.000	-0.032	0.000
Number of Kids in Class squared	0.000	0.000	0.000	0.000	0.000	0.000
Teacher's Rating of Class Behavior	0.029	0.000	0.029	0.000	0.029	0.000
Percentage of Minority Students (categorical)	-0.029	0.000	-0.029	0.000	-0.029	0.000
Public School	-0.185	0.000	-0.185	0.000	-0.185	0.000
School's Average Daily Atendance	-0.096	0.000	-0.096	0.000	-0.096	0.000
TT1 Funds Received by School	-0.121	0.000	-0.121	0.000	-0.121	0.000
Crime a Problem	0.150	0.000	0.150	0.000	0.150	0.000
Students Bring Weapons	-0.113	0.000	-0.113	0.000	-0.113	0.000
Children or Teachers Physically Attacked	0.036	0.000	0.036	0.000	0.036	0.000
Security Measures in School	-0.094	0.000	-0.094	0.000	-0.094	0.000
Parents Involved in School Activities	0.029	0.000	0.029	0.000	0.029	0.000
Policy: Retained for any Reason	0.135	0.000	0.135	0.000	0.135	0.000
Policy: Retained for Immaturity	-0.091	0.000	-0.091	0.000	-0.091	0.000
Policy: Retained by Parents' Request	0.126	0.000	0.126	0.000	0.126	0.000
Policy: Retained without Parents' Permission	0.100	0.000	0.100	0.000	0.100	0.000
Cognitive Ability	-0.700	0.000	-0.768	0.000	-0.332	0.000
General Ability	-0.741	0.000	-1.351	0.000	-1.176	0.000
Behavioral Ability	-0.045	0.012	-0.122	0.000	0.017	0.596

Note: Standard errors are obtained via 200 bootstrap replications.

TableA7: Estimates of Parameters in the Distribution of Ability

	Cognitive		General		Behavioral	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Mean of the 1st Mixture Component	0.139	0.001	-0.309	0.000	-0.454	0.000
Variance of the 1st Mixture Component	0.095	0.000	0.085	0.000	0.457	0.000
Weight of the 1st Mixture Component	0.396	0.000	0.351	0.000	0.514	0.000
Mean of the 2nd Mixture Component	-0.072	0.001	0.230	0.000	0.860	0.000
Variance of the 2nd Mixture Component	0.071	0.000	0.089	0.000	0.092	0.000
Weight of the 2nd Mixture Component	0.403	0.000	0.406	0.000	0.296	0.000
Mean of the 3rd Mixture Component	-0.130	0.242	0.061	0.551	-0.111	0.649
Variance of the 3rd Mixture Component	0.098	0.000	0.226	0.000	0.410	0.000
Weight of the 3rd Mixture Component	0.201	0.000	0.243	0.000	0.190	0.000

Note: Standard errors are obtained via 200 bootstrap replications.

Table A8: Average Reading Test Scores by Potential and Actual Retention Status: 2003-04 School Year

Potential Retention Status	Actual Retention Status	A kid who is actually (i.e. conditional on retention status being:) Not Retained	Retained in Kindergarten	Retained Early	Retained Late	Unconditional
would obtain if the kid was	Not Retained	0.127*	-0.543*	-0.670*	-0.702*	0.067*
	Retained in Kindergarten	0.112*	-0.822*	-1.042*	-1.067*	0.026
	Retained Early	-0.204*	-1.034*	-1.205*	-1.221*	-0.279*
	Retained Late	-0.282*	-0.906*	-0.991*	-1.000*	-0.336*

(Highlighted) Factual Results: diagonal cells 0.127*, -0.822*, -1.205*, -1.000*

Note: Let T = 0,1,2, or 3 represent the actual retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively. Let S(i) be the potential test score at 2003-04 school year if the kid were retained at time i=0,1,2,3. The row i, column j element of this table calculates E[S(i) | D=j]. For example, a kid who was actually not retained would get -0.204 on average if the kid were retained at 1 or 2 grade instead. When calculating them, we keep kid's age fixed at 11.

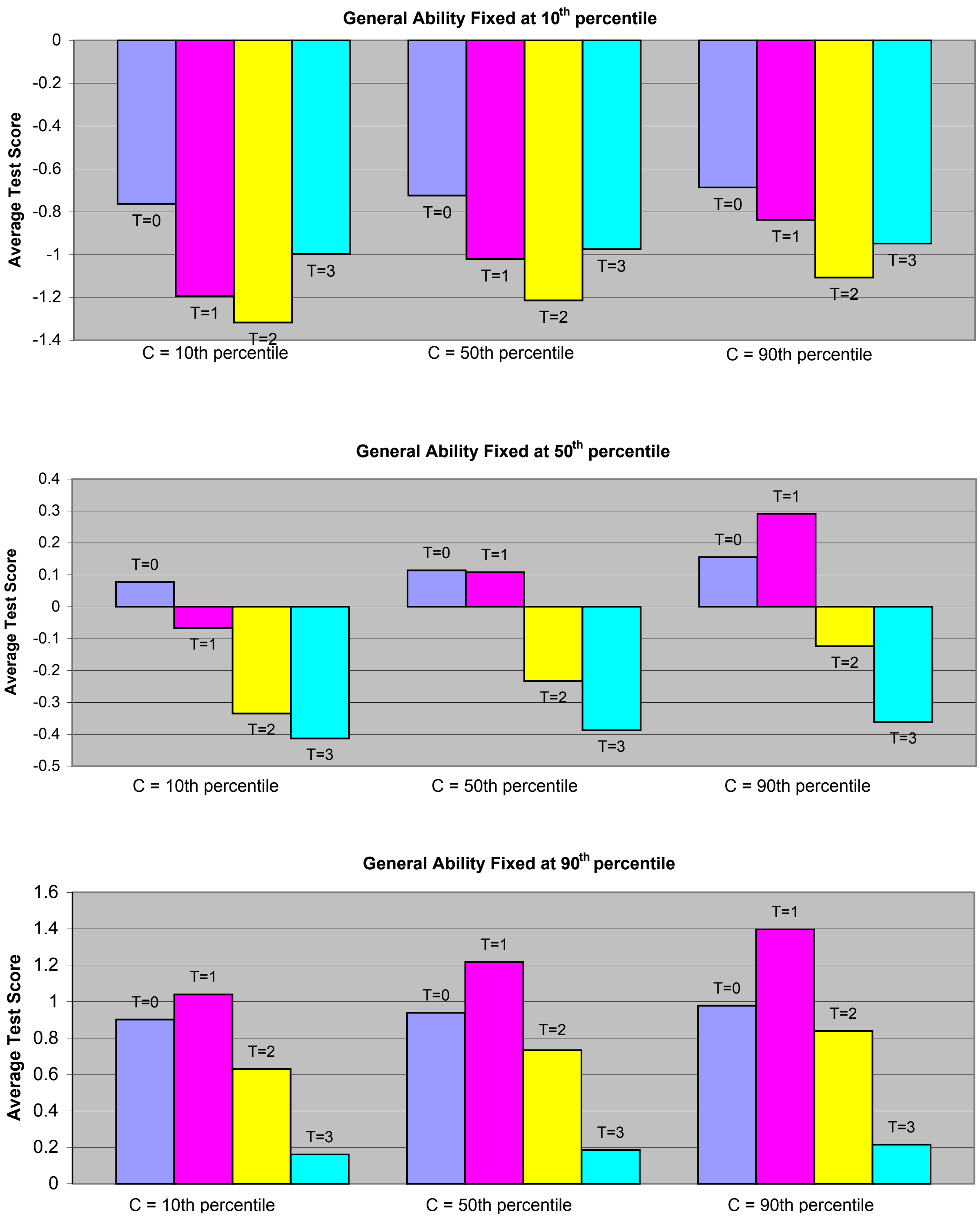
Table A9: Average Math Test Scores by Potential and Actual Retention Status: 2003-04 School Year

Potential Retention Status	Actual Retention Status	A kid who is actually (i.e. conditional on retention status being:) Not Retained	Retained in Kindergarten	Retained Early	Retained Late	Unconditional
would obtain if the kid was	Not Retained	0.127*	-0.551*	-0.687*	-0.696*	0.066*
	Retained in Kindergarten	0.139*	-0.754*	-0.942*	-0.924*	0.059
	Retained Early	-0.210*	-0.993*	-1.155*	-1.151*	-0.281*
	Retained Late	-0.390*	-0.921*	-1.006*	-1.017*	-0.437*

(Highlighted cells:) Factual Results

Note: Let T = 0,1,2, or 3 represent the actual retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively. Let S(i) be the potential test score at 2003-04 school year if the kid were retained at time i=0,1,2,3. The row i, column j element of this table calculates E[S(i) | D=j]. For example, a kid who was actually not retained would get -0.210 on average if the kid were retained at 1 or 2 grade instead. When calculating them, we keep kid's age fixed at 11.

Figure A10: Average Math Score at 2003-04 School Year for Different Percentiles of the Cognitive Ability Distribution, Conditonal on Retention Status and General Ability.



T=0,1,2 and 3 are "not retained", "retained in kindergarten", "retained early (at 1 or 2 grade)", and "retained late (at 3 or 4 grade)", respectively. When calculating these values, we keep age of children fixed at 11. Scores are standardized.