Grade Retention and Multidimensional Skill Formation in Young Children*

Fernando Saltiel[†] and Miguel Sarzosa[‡]

July 19, 2021

Abstract

This paper estimates the impacts of early grade retention on children's cognitive and non-cognitive skill development. We present and estimate a dynamic model of multidimensional skill formation, with endogenous retention outcomes. We use ECLS-K data covering students' test scores, non-cognitive skills, retention events along with parents' skills and investment choices. Low cognitive and non-cognitive skilled students are far more likely to be retained. Although in general, the skill production processes exhibit self-productivity, strong effects of parental investments, and that parents' skills affect children's' skill indirectly through investment, the technology of skill formation differs across retention status. Retention negatively impacts students' cognitive abilities, yet leads to positive effects for high-skilled students at baseline. Grade repetition boosts non-cognitive skills for low-skilled students at baseline. Targeted retention policy changes could improve children's skill development.

Keywords: Grade Retention, Dynamic Skill Development, Non-Cognitive Skills.

JEL Codes: I20, I24

*We thank Francesco Agostinelli, Felipe Barrera-Osorio, Emilia Del Bono, Matthew Lindquist, Tim Moore, Markus Poschke and Lesley Turner for their useful comments and suggestions. We are indebted to seminar participants at Vanderbilt University, SOLE 2021, Econometric Society Summer Meeting 2021, Econometric Society Winter Meeting 2020 and Purdue's Summer 2020 Brown Bag Lunch for their helpful feedback.

[†]McGill University. Email: **fernando.saltiel@mcgill.ca**

[‡]Purdue University. Email: **msarzosa@purdue.edu**

1 Introduction

The prevalence of grade retention implies this practice plays an important role in the skill development process.[1] As a result, an extensive literature has analyzed the effects of grade repetition on academic outcomes.[2] Yet the mere nature of retention implies that students are separated from their peers, potentially harming their non-cognitive skill development. On the other hand, if being held back leads students to better understand class material, thus gaining confidence in their learning abilities, repeating a grade may yield benefits in this dimension. However, the existing literature has paid limited attention to the impacts of grade repetition on children's non-cognitive skill outcomes.

In this paper, we estimate the effects of early grade retention on children's cognitive and non-cognitive skill development. We present and estimate a dynamic model of skill formation, which incorporates children's latent abilities, parental skills and investment choices. The model extends the existing literature on dynamic skill formation (Cunha et al., 2010; Agostinelli and Wiswall, 2016a; Attanasio et al., 2020) by accounting for children's endogenous retention outcomes. We consider latent cognitive and non-cognitive abilities, which are unobserved to the econometrician and proxied by observed measures. We further incorporate parental skills and allow for dedicated investments to affect their children's cognitive and non-cognitive skill development, fitting in with recent work by Attanasio et al. (2020) considering different investment dimensions. We allow for parental investments to reinforce or compensate children's initial skills. In our framework, grade retention depends on background characteristics, school-level retention policies and children's initial stock of skills. As a result, our model allows for current skills to indirectly affect future skills through grade retention and parental investment. We estimate retention-status-dependent CES production functions, which allow us to recover heterogeneous impacts across students' initial skill levels.

We implement the model using data from the ECLS:K-2011 study, which follows a nationally-representative sample of kindergarten students in the 2010-11 academic year through the Spring of 2013.[3] ECLS-K data includes detailed information on children's background characteristics, their academic performance and teacher-reported non-cognitive skill

[1]Across OECD counties, close to 10% of fifteen year-olds repeated a grade at least once (Ikeda and Garcia, 2014). We refer to grade retention and repetition interchangeably.

[2]Eide and Showalter (2001), Jacob and Lefgren (2004, 2009), Fruehwirth et al. (2016), Eren et al. (2017, 2018) and Schwerdt et al. (2017) present evidence in the United States. d'Haultfoeuille (2010), Alet et al. (2013) and Gary-Bobo et al. (2016) examine the effects of retention in France, Cockx et al. (2019) focus on the Belgian context, Koppensteiner (2014) present evidence in Brazil, Diaz et al. (2016) do so in Chile and Manacorda (2012) in Uruguay.

[3]ECLS included three additional survey rounds, following students 2015-16. Given our interest in early retention events, we focus on outcomes in the Spring 2013 survey. Throughout the paper, we refer to the 2013 survey round as the 'endline' round.

measures across both baseline and endline survey rounds. We further observe parents' socioemotional skills along with time and monetary investments in their children. We consider retention events in kindergarten and first grade—5.7% of students in the sample are held back, and examine the impact of repetition on cognitive and non-cognitive skill outcomes. Lastly, we observe detailed information on school-level characteristics, including various policies which can affect the prevalence of early retention.

Across both time periods, we find that observed test scores and non-cognitive skill variables measure latent skills with substantial error. Importantly, parental investments largely reinforce children's initial cognitive and non-cognitive skills and parents with higher socioemotional skills at baseline invest more in their children. Our model incorporates endogenous retention outcomes, which are strongly shaped by children's initial cognitive and non-cognitive skills, with limited contributions from parents' skills: 20% of students in the bottom of the joint skill distribution repeat either kindergarten or first grade compared to just 2% of their peers in the top. Moreover, school-level retention policies strongly affect the likelihood of early repetition.

We estimate retention-dependent skill formation technologies embedded in a Roy-like model of potential outcomes. In the process, we overcome the problem caused by 're-normalization'—the fact that imposing location and scale assumptions on skills distributions both at t and $t+1$ biases the parameters towards a Cobb-Douglas functional form (Agostinelli and Wiswall, 2016b)—using an alternative approach to the one proposed by Agostinelli and Wiswall (2016a). Our approach treats skills as unobservable for the entirety of the estimation process, a feature that is critical for the identification of a model with endogenous selection into treatment. Even though we find ample evidence on self-productivity and strong effects of parental investments in the production of cognitive and non-cognitive skills for both retained and non-retained students, we find that the technologies of skill formation in retained students significantly differ from those in non-retained children. In particular, we find that the production of non-cognitive skills has a significantly *higher* elasticity of substitution among retained children. In contrast, the production function of cognitive skills in retained children allows for substantially *less* substitutability between inputs than among non-retained students. In addition, parental investments play an important role in the skill formation process across both skill dimensions as well as for retained and non-retained students. Furthermore, parental socioemotional skills affect their children's skill outcomes largely through leading to increased investments.

We first estimate the impacts of retention on children's cognitive skill development. Early retention lowers students' cognitive skills at endline by 0.019 standard deviations (σ), with far larger estimated impacts for students who were in fact retained (TT = 0.147 σ). We examine heterogeneous impacts across the initial skill distribution, finding large negative impacts for low-skilled students, exceeding -0.4 σ, along with a positive ATE for their high-skilled peers. These results fit in with earlier evidence from Fruehwirth et al. (2016) on students' math and reading test scores. To provide comparable evidence with the existing literature on retention, we re-estimate our model using test scores as outcomes. We find far larger negative impacts of retention, thus remarking the extent to which test scores measure latent skills with error.

At the same time, grade retention yields small improvements in children's latent non-cognitive abilities, in the range of 0.04 standard deviations. While being held back does not significantly impact high-skilled students non-cognitive outcomes, retention significantly boosts low skilled children's non-cognitive skills. The estimated impacts of retention exhibit important differences across the two skill dimensions, remarking the importance of extending the literature to incoporate non-cognitive skill outcomes, in light of the importance of this skill dimension in driving successful outcomes in adulthood (Heckman et al., 2006).

Lastly, we take advantage of the estimated model parameters to inform the impacts of two policy exercises. First, since the estimated impacts of retention may not correspond to policy relevant margins (Heckman and Vytlacil, 2001; Mogstad et al., 2018), we examine how changing school-level retention policies affects children's latent skill development. We consider the impacts of not allowing schools to retain children without their parents' consent. Policy compliers are largely drawn from the bottom of the initial skill distribution. As such, this policy change would negatively impact children's cognitive skills while slightly improving their non-cognitive abilities.

Second, since retention implies students must spend an additional year in schooling, this policy implies sizable costs to school systems. We thus consider how replacing retention for a parental income transfer destined towards investing in their children could affect skill outcomes. The compensating income transfer would yield small improvements in children's cognitive skill development, along with small negative impacts on the non-cognitive margin. This policy simulation remarks one of the important advantages of our empirical strategy, as we can compare the impacts of income transfers previously considered in the literature (Agostinelli and Wiswall, 2016a; Attanasio et al., 2020), against a costly policy like retention, which directly affects children's skill development.

This paper makes various contributions to the literature on skill formation. First, ours is the first paper to embed a Roy model of grade retention in a model of dynamic skill accumulation. Our framework thus allows us to analyze the process of multidimensional skill formation across different scholing paths and to incorporate how parents' investment choices may react to their children being held back. We thus contribute to a growing structural literature on dynamic skill accumulation (Cunha and Heckman, 2007, 2008; Cunha et al., 2010; Agostinelli and Wiswall, 2016a; Attanasio et al., 2020), fitting in with recent papers estimating flexible CES production functions. We follow Sarzosa (2015) by incorporating a Roy model within this framework, yet extend his work by including parental skills in the production function as well as by analyzing how different policy reforms affect children's skill outcomes.

Furthermore, we contribute to a large literature on the impacts of grade retention on children's skill outcomes. Within the structural literature, ours is the first paper to estimate the effects of this practice on children's non-cognitive skills and to directly distinguish test scores from latent abilities. Various papers have previously estimated latent factor models to recover the effects of retention, including Gary-Bobo et al. (2016) in France and Cockx et al. (2019) in Belgium, yet these two papers consider latent ability to be unidimensional. Most closely related to this paper is Fruehwirth et al. (2016), who estimate the effect of primary school retention in the United States while accounting for latent cognitive and non-cognitive components of ability. We extend this strand of the literature by embedding the process of grade retention within a model of dynamic skill accumulation, documenting the impacts on non-cognitive skill development and directly incorporating the importance of compensating parental investment. As such, we contribute to literature on the impacts of retention on non-academic outcomes, including Ozek (2015), Diaz et al. (2016) and Eren et al. (2017, 2018). A number of these papers analyze how retention impacts non-cognitive skill proxies, yet the existing literature has not provided a comprehensive analysis of the impacts of this practice on children's non-cognitive skill formation.

The rest of the paper proceeds as follows. Section 2 describes the ECLS-K data, presents summary statistics and introduces reduced form estimates of the effects of grade retention on test score and non-cognitive skill measures. Section 3 introduces a model of grade retention and dynamic skill accumulation. Section 4 presents our estimation strategy. Section 5 presents the estimated model results. Section 6 presents the determinants of grade retention and the parameters of the production functions. Section 7 shows the estimated impacts of grade retention on children's cognitive and non-cognitive skill development. Section 8 presents evidence on the impacts of changes in school-level retention policies on children's cognitive and non-cognitive skills. Lastly, Section 9 discusses the results, concludes, and offers suggestions for future research in this context.

2 Data Sources and Summary Statistics

2.1 Data Sources

In this paper, we take advantage of data from the Early Childhood Longitudinal Study: Kindergarten Cohort (ECLS-K:2011), which follows a nationally-representative cohort of 18,200 kindergarten students in the 2010-11 academic year through fifth grade.[4] The survey was conducted by the National Center for Education Statistics (NCES), which first surveyed children in the fall of 2010, with follow-up surveys in the spring of 2011, spring of the 2011-12 and 2012-13 school years.[5] Our analysis focuses on the sample of children who were enrolled in kindergarten for the first time in 2010-11.

Importantly, ECLS-K:2011 includes information on students' grade progression. We follow the existing literature and define a student as having been retained if she is enrolled in the same grade as in the previous year. Given the low prevalence of retention in the sample, we define a student to have been retained if she repeated either kindergarten or first grade.[6] We consider outcomes in the Spring 2013 survey round, which measures outcomes at the end of second grade for on-time students.

The ECLS captures detailed information on student and household background characteristics, which encompass students' race, gender and their age in each survey round, along with information on their household composition, parents' education and family income. We also observe baseline measures of parents' reported psychological well-being, encompassing whether they felt sad, could not get going, and felt depressed, among other questions, which we use to model the importance of parents' characteristics in the skill formation process.

We further take advantage of various baseline measures of parents' behavior to examine the importance of parental investment for their children's skill development. In particular, we consider the number of books available in the household, weekly time parents spend reading to their children, the number and the types of activities done with their children, engagement in school activities along with their attitudes towards parenting and their relationship with their child.[7] These measures represent a combination of parents' time and

[4]The ECLS-K:2011 survey is directly comparable to its predecessor, the Early Childhood Longitudinal Study: Kindergarten Cohort, which initially followed a cohort of kindergartners in 1998-99. Throughout the rest of the paper we refer to the ECLS-K:2011 as the ECLS for simplicity.

[5]ECLS also fielded surveys in the fall of 2011 and 2012, yet only one-third of students were surveyed.

[6]Fruehwirth et al. (2016) also combine retention events across different grades due to the small prevalence of retention in the ECLS-K:1999. We drop students who repeat both kindergarten and first grade.

[7]The activities measure incorporates how much time parents spend telling stories to their children, singing songs with them, helping them in doing art, playing games, talking about nature, building things, playing sports, practicing numbers, reading books and picture books. The school engagement measure captures whether parents attended back to school night, participated in the PTA or parent advisory groups, attended school events and/or participated in committees/volunteering/fundraising at the school. The index capturing

monetary investments in their children. Similar measures have been considered in the skill development literature across different contexts (Cunha et al., 2010; Agostinelli and Wiswall, 2016a; Attanasio et al., 2020).

Furthermore, ECLS includes a school-level survey which captures relevant characteristics to the analysis of grade retention. We follow Fruehwirth et al. (2016) and consider information regarding school-level retention policies for kindergarten students. These measures indicate whether schools can retain students due to immaturity, at parents' request and/or without their consent, for academic deficiencies, if they have failed tests, and whether students can be retained more than once.[8]

To examine the impact of grade retention on students' cognitive skill development, we take advantage of information on students' performance in math, reading and science exams taken in both the baseline and endline survey rounds.[9] These assessments cover the same material independently of students' grade progression and we measure students' performance using NCES-reported item response theory (IRT) scores.

In each survey round, teachers rate students on a variety of dimensions related to their socioemotional development, and we rely on these measures to capture students' non-cognitive skills. In particular, we focus on scores across the Social Rating Scales (SRS), which encompasses teachers' ratings of their students in their interpersonal skills, their ability to exercise self-control and to interact with others along with their internalizing and externalizing problem behaviors.[10] We also consider teachers responses to the Child Behavior Questionnaire, which covers measures regarding students' attentional focusing — which measures their tendency to maintain attention on a task — and inhibitory control, which assesses a child's ability to plan and inhibit correct responses when following instructions in novel situations.

A possible concern in this context is that teachers may under- or overstate retained students' non-cognitive skills, as reference bias — such as comparing these students to their non-retained peers — may influence the reported measures we use in our analysis. For instance, Elder and Zhou (2021) show that using teacher reported measures results in understated Black-White gaps in non-cognitive skills. To address this concern, we follow the

attitudes towards parenting and the relationship to their child includes measures capturing whether being a parent is harder than expected, whether the child does things that bother the parent, whether the parent sacrifices for the child, whether they often feel angry with their kids, whether the parent shows love / express affection to the child, whether they have close times together and whether the child likes the parent.

[8]ECLS does not include additional measures of retention policies for first grade students, yet the vast majority of ECLS children attend the same school in kindergarten and first grade. As such, the kindergarten policies are proxies for school-level policies across the retention events considered in this paper.

[9]Students took the baseline science exam in the Spring of 2011, whereas the math and reading exams were taken in the Fall. All endline exams were taken in the Spring of 2013.

[10]The SRS measure has been used extensively in the skill development literature (Neidell and Waldfogel, 2010; Bertrand and Pan, 2013; Elder and Zhou, 2021).

approach presented in Elder and Zhou (2021), who rely on teachers' ratings of their students math, science and reading skills and compare them against their observed performance in the corresponding subject-specific exams. This information thus allows us to examine whether teachers differentially rate retained students conditional on their observed test scores.[11] In Table A.1, we show that across a number of specifications, higher-scoring students earn higher ratings from their teachers, yet this relationship does not systematically differ across students' retention status. As such, reference bias is unlikely to play an important role in shaping teachers' reports of children's non-cognitive skills.[12]

A common challenge for papers on grade retention is whether to focus on the impact of this event holding the age or the grade of students fixed. In this paper, test score data availability implies that we must fix the age. As a result, the outcomes across retained and non-retained students will be observed at the same age, but the latter group will have completed an additional grade. For academic outcomes, Fruehwirth et al. (2016) argue that this assumption likely delivers a conservative estimate of the benefits of retention, as retained students have not been exposed to an additional year of class material. For socioemotional skill measures, on the other hand, examining the outcomes at the same age across retention status allows for a comparison which does not confound the causal impacts of the event from over-time variation in children's non-cognitive skill development.

We construct the sample as follows. We first drop 2,390 children who did not participate in the baseline survey round, along with an additional 811 who had previously enrolled in kindergarten. We drop an additional 613 students who are enrolled in special education in kindergarten and 866 students who did not take the baseline assessments or provided information on their gender, age or race. An additional 1,286 children have missing values for the teacher-reported non-cognitive skill measures at baseline. We drop an additional 1,468 children due to missing information on grade progression. Furthermore, 1,006 students have missing information on their test scores at endline. Additionally, 561 students having missing values for endline non-cognitive skill measures. As a result, the final sample size includes 9,010 students, of whom 520 repeated either kindergarten or first grade.[13,14]

[11] We use information from the Spring 2011 survey round, where retained students are those who had already repeated kindergarten prior to the baseline survey. In the Spring round, teachers rate students on their mathematics, science and reading skills on a scale ranging from 1 ('far below average') to 5 ('far above average'). These measures follow directly from teachers' responses to the Academic Rating Scales (ARS) survey. We compare teachers' ratings to children's test scores for each subject in the same survey round.

[12] We could have alternatively considered parent-reported child non-cognitive skill measures available in the ECLS. However, Del Bono et al. (2020) show that such measures are strongly influenced by parents' own non-cognitive skills, thus representing biased measures of children's socioemotional development.

[13] We do not drop individuals who have missing information, as we instead create indicators for missing variables and impute the sample mean.

[14] 32 students who are enrolled in kindergarten in the 2011-12 academic year appear in the data as enrolled

2.2 Descriptive Analysis

Table 1: Summary Statistics by Retention Status

	Full Sample (1)	Not Retained (2)	Retained (3)
Observed Characteristics			
Male	0.494	0.489	0.585***
Age (R1)	5.610	5.621	5.420***
Underrepresented Minority	0.406	0.405	0.429
Both Parents	0.721	0.727	0.616***
Parents' Education	14.019	14.052	13.477***
Log HH Income	10.835	10.843	10.706**
Parental Sadness	-0.009	-0.013	0.056
Parental Anxiety I	-0.003	-0.010	0.104
Parental Anxiety II	-0.005	-0.008	0.047
School Policies			
Retained for Immaturity	0.654	0.651	0.696*
Retained at Parents' Request	0.669	0.665	0.723**
Retained for Academic Deficiency	0.827	0.825	0.852
Retained if Failed Test	0.063	0.063	0.079
Retained More than Once	0.061	0.060	0.087*
Retained Without Parents' Consent	0.267	0.264	0.310*
Skill Measures (R1)			
Math Test Score	-0.000	0.052	-0.852***
Science Test Score	-0.000	0.029	-0.472***
Reading Test Score	0.000	0.047	-0.773***
Attentional Focus	0.000	0.042	-0.690***
Inhibitory Control	-0.000	0.032	-0.528***
Social Rating Scales	-0.000	0.023	-0.374***
Skill Measures (R6)			
Math Test Score	-0.000	0.061	-0.999***
Science Test Score	0.000	0.038	-0.628***
Reading Test Score	0.000	0.075	-1.231***
Attentional Focus	-0.000	0.020	-0.327***
Inhibitory Control	0.000	0.017	-0.283***
Composite Social Rating Scales	0.000	0.018	-0.299***
Observations	9,010	8,490	520

Source: ECLS-K:2011. Note. Table 1 displays summary statistics by students' early retention status. R1 refers to measures observed in the baseline survey, whereas R6 captures outcomes in the Spring 2013 survey. School policies encompass school-level retention policies for students in kindergarten. The stars in the third column capture the statistical significance of a t-test comparing the means of the variables of non-retained students to their retained peers. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Summary Statistics. In the first column of Table 1, we present summary statistics for the ECLS sample and across students' retention status. The sample is evenly split by gender, 40% of children are underrepresented minorities, and close to three-fourths of the sample lives in two-parent households. Retained students are less likely to reside with both parents, and tend to come from households with less educated parents. Our empirical strategy ad-

in second grade in 2012-13. We define these students are non-repeaters, though our results are robust to defining them as repeaters.

ditionally incorporates variation in school-level retention policies for kindergartners. There are important cross-school differences in these policies, as 67% of children attend schools in which children can be retained for immaturity, and 27% can be retained without their parents' consent. Moreover, retained students attend schools with more 'stringent' retention policies, these differences are significant for the immaturity, parental request, parental consent and multiple retention policies.

Importantly, there are sizable differences in students' academic achievement at baseline, as retained students trail their non-retained peers by 0.85, 0.77 and 0.47 σ in the math, reading and science test scores, respectively. We find similar differences across all teacher-reported measures of students' non-cognitive skills at baseline, exceeding 0.5 σ in the inhibitory control and attentional focus domains. Differences in academic achievement across retention status are larger through the Spring 2013 survey, as retained students trail their non-retained peers in the math, reading and science exams by 1, 1.23 and 0.63 standard deviations, respectively. At the same time, we find significant differences across the three non-cognitive skill measures, in the range of 0.28-0.33 σ.

Regression Analysis. As shown above, retained students differ from their non-retained peers along a variety of dimensions at baseline. To discern the relative importance of these components, we estimate an OLS regression of grade retention against all baseline characteristics and retention policies. We present the results in Table A.2. Conditional on background characteristics, students with lower math and reading test scores at baseline were more likely to have been retained through first grade, as were those with lower attentional focus. Males and students in poor households remain more likely to have been retained as well. School-level retention policies are jointly significant (p-value = 0.000).

In Table 2, we present preliminary evidence from an OLS regression on the impacts of grade retention on students' academic achievement through the Spring 2013 survey round. Conditional on background characteristics and baseline skill measures, grade retention is associated with a reduction in students' math test scores by 0.46 standard deviations. The estimated coefficients on the reading and science test scores are negative as well, reaching 0.72 and 0.13 σ, respectively. We note that the evidence presented in Table 1 indicated that retained students trailed their non-retained peers in the non-cognitive skill outcomes. However, the last three columns indicate that upon controlling for baseline characteristics, grade retention is no longer associated with lower non-cognitive skills through endline. In fact, grade retention leads to an increase in students' Social Rating Scales by 0.11 σ, denoting the extent to which selection-on-observables affects the estimated impacts of this practice.

Table 2: OLS Estimates of the Impacts of Retention

	Math Test Score (1)	Reading Test Score (2)	Science Test Score (3)	SRS (4)	Attentional Focus (5)	Inhibitory Control (6)
Grade Retention	-0.464***	-0.132***	-0.723***	0.111**	0.075	-0.004
	(0.035)	(0.031)	(0.042)	(0.042)	(0.042)	(0.043)
Math Test Score	0.557***	0.226***	0.302***	0.176***	0.082***	0.055***
	(0.013)	(0.013)	(0.014)	(0.016)	(0.016)	(0.016)
Science Test Score	0.180***	0.505***	0.196***	0.017	0.005	-0.002
	(0.009)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)
Reading Test Score	-0.023*	0.082***	0.218***	-0.014	-0.013	0.001
	(0.011)	(0.011)	(0.011)	(0.015)	(0.015)	(0.015)
Attentional Focus	0.098***	0.043***	0.064***	0.145***	0.070***	-0.002
	(0.012)	(0.011)	(0.013)	(0.016)	(0.015)	(0.016)
Inhibitory Control	-0.001	0.009	0.022	0.201***	0.237***	0.193***
	(0.013)	(0.013)	(0.013)	(0.017)	(0.016)	(0.017)
Composite Social Rating Scales	-0.001	-0.001	0.010	0.078***	0.131***	0.252***
	(0.010)	(0.010)	(0.010)	(0.013)	(0.013)	(0.013)
Background Characteristics			✓			
Observations			9,010			
R^2	0.584	0.589	0.536	0.294	0.285	0.257

Source: ECLS-K:2011. Note: Table 2 presents the estimated impacts of grade retention on students' endline test scores and teacher-reported non-cognitive skill measures. We present evidence from an OLS regression with controls for family and children's background characteristics as well as baseline skill measures. Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.15$, *** $p < 0.001$.

While the evidence presented in Table 2 offers suggestive evidence that retention may have heterogeneous impacts on children's multidimensional skills, these results do not constitute causal evidence. First, an extensive literature has shown that test scores and non-cognitive skill constructs measure latent ability with significant error (Carneiro et al., 2003; Heckman et al., 2006). As a result, and equating test scores with latent skills fails to capture the process of multidimensional skill accumulation. Moreover, this analysis does not incorporate the extent to which parental skills and their investments may shape the prevalence and the impacts of retention. Lastly, these results are not informative of potentially heterogeneous impacts of retention across the initial skill distribution. To address these concerns and recover the effects of retention on multidimensional skill development, we next introduce a dynamic Roy model of skill accumulation.

3 Model of Grade Retention and Skill Formation

In this section, we introduce a dynamic model of skill formation which accounts for endogenous grade repetition. Our empirical strategy incorporates students' cognitive and non-cognitive skills, which we posit to be unobserved to the econometrician and proxied by observed measures. The model incorporates various desirable features related to the skill development process. First, it accounts for the self- and cross-productivity of skills (Cunha and Heckman, 2007), while allowing for parents' skills and their investment choices to shape children's skill development. Second, parental investments are driven by household resources and the initial levels of the child's skills. At the same time, we allow for the initial stock of skills to affect the likelihood of grade retention, along with parents' skills, observed characteristics and school-level policies. As a result, initial skills can also affect future skills through grade retention and parental investment. Our model further considers retention-status-dependent production functions, thus allowing us to recover heterogeneous impacts across students' initial skill levels without the need for extrapolation.

3.1 Model Structure

In this framework, $\theta_{S,i,\tau}$ denotes student i's stock of skills $S = \{C, NC\}$ at time $\tau \in \{t, t+1\}$, $\theta_{P,i}$ reflects parents' time-invariant socioemotional skills and $\mathcal{I}_{i,t}$ captures parental investments in their children's skill development. R_i is a binary variable which equals one if student i has been retained between t and $t+1$ and zero otherwise. As such, the model of skill formation which incorporates parental characteristics and allows for retention-specific production functions is given by:

$$\theta_{S,i,t+1} = (\gamma^R_{C,S,t}\theta^{\rho^R_S}_{C,i,t} + \gamma^R_{NC,S,t}\theta^{\rho^R_S}_{NC,i,t} + \gamma^R_{P,S,t}\theta^{\rho^R_S}_{P,i} + \gamma^R_{I,S,t}\mathcal{I}^{\rho^R_S}_{i,t})^{1/\rho^R_s} + \eta^R_{i,S,t} \tag{1}$$

where $\gamma^R_{C,S,t} + \gamma^R_{NC,S,t} + \gamma^R_{P,S,t} + \gamma^R_{I,S,t} = 1$, and $\eta^{R_i}_{i,S,t}$ represents a mean-zero i.i.d. shock with variance $\sigma^2_{\eta_{i,S,t}}$, which is independent to contemporaneous skills, across skill dimensions ($S = \{C, NC\}$), by time and across retention status. We follow Cunha et al. (2010) and Attanasio et al. (2020) in introducing a CES production function of latent skills in period $t+1$, where $\sigma^R_s = \frac{1}{1-\rho^R_s}$ captures the elasticity of substitution between students' skill endowments at time t, latent parental skills and parental investments in the production function of skill S at time $t+1$.[15] Importantly, the parameters of the CES production function vary across students' retention status R_i, which allows us to examine how retention affects the skill development process. We posit the following model for retention:

$$R_i = \mathbb{1}\left[X_{i,t}\beta_t + \lambda_P\theta_{P,i} + \lambda_C\theta_{C,i,t} + \lambda_{NC}\theta_{NC,i,t} + \tau Z_i > \varepsilon_{i,t}\right] \tag{2}$$

[15]The CES production function allows us to examine the importance of static and dynamic complementarities in the skill development process (Cunha and Heckman, 2008; Cunha et al., 2010), in a more general framework than a Cobb-Douglas production function, as it places fewer restrictions on the possible interactions between the its inputs. We have also estimated richer specifications, including a nested CES specification, but we could not reject the estimated parameters differed from the CES specification. Furthermore, we favor the CES structure over a stochastic translog specification (Agostinelli and Wiswall, 2016a), since we consider inputs as unobservable and, as such, interactions between them are not identified.

where $\mathbb{1}$ is an indicator function which equals one if true, $X_{i,t}$ includes observed characteristics measured at baseline. $\varepsilon_{i,t}$ is a mean-zero error term with variance $\sigma^2_{\varepsilon_{i,t}}$, which is independent of observed characteristics, latent factors, and of $\eta^{R_i}_{i,S,t}$. Our model additionally considers the impact of school-level retention policies (Z_i) on the likelihood of grade repetition, which represent exclusions as they only affect students' skill development through grade retention. We also allow for parents' socioemotional skills (θ_P) to affect the likelihood of retention.

3.2 Parental Investment and Socioemotional Skills

Parental Investments. We allow for parental investments to affect children's skill development. Our approach does not consider an explicit model of parental choices, beliefs and preferences as in Del Boca et al. (2014), Doepke and Zilibotti (2017) and Doepke et al. (2019). On the other hand, we follow the existing literature on dynamic skill development (Cunha and Heckman, 2007, 2008; Cunha et al., 2010; Attanasio et al., 2020; Agostinelli and Wiswall, 2016a,b) and allow for investment to depend on household resources, parents' and children's latent skills—thus capturing the extent to which investments reinforce or compensate children's initial skill endowments. $\mathcal{I}_{i,t}$ denotes parents' investments in their child's skill development, which are given by the following model:

$$\mathcal{I}_{i,t} = \alpha_X X_{i,t} + \alpha_Y Y_{i,t} + \beta_P \theta_{P,i} + \beta_{C,t}\theta_{C,i,t} + \beta_{NC,t}\theta_{NC,i,t} + v_{I,i,t} \tag{3}$$

where $v_{I,i,t}$ is a mean-zero error term which is independent of observed characteristics, latent skills and investment factors.[16] Since parental investment choices may be endogenous to shocks in the skill development process ($\mathcal{I}_{i,t} \not\perp \eta^R_{i,S,t}$), we include a set of instruments $Y_{i,t}$ which are excluded from the production function. We follow Cunha et al. (2010); Agostinelli and Wiswall (2016a); Attanasio et al. (2020) and include household income measured in the baseline survey round as an instrument, as it represents a measure of household resources.

Socioemotional Skills. In light of our interest in understanding the impacts of retention on children's non-cognitive skill development, we additionally incorporate the importance of parents' socioemotional skills (θ_P) in the skill development process. We remark that the existing literature, including Cunha et al. (2010) and Del Bono et al. (2020), among others, has previously shown that parents' non-cognitive skills play an important role in directly shaping children's skill outcomes.

[16] $\mathcal{I}_{i,t}$ depends on the same observed characteristics as in equation (1).

3.3 Measurement System

As noted above, our empirical model incorporates the fact that skills and investments are inherently unobservable in nature. As such, using any of the observed skill measures discussed in Section 2 as proxies for the latent factors could lead to biased results of our main model parameters. To this end, we present a measurement system for children's skills and for parental characteristics which assumes that observed measures are a linear function of observable characteristics, the corresponding latent factors and random shocks.

3.3.1 Children's Skills

In this framework, students' skills play a critical factor in driving grade retention. However, cognitive and non-cognitive skills are latent rather than observable. We thus follow Cunha et al. (2010) and posit that observed test score and non-cognitive skill measures are a linear function of latent abilities and observable characteristics. We introduce a dedicated measurement system in which an observed measure $T_{S,k,\tau}$ ($k \in \mathcal{K}_S, \tau \in \{t, t+1\}$) corresponding to skill dimension $S = \{C, NC\}$ represents an error-ridden measure of latent skill $\theta_{S,\tau}$:

$$T_{S,k,i,\tau} = \beta_{S,k,\tau}\mathbf{x}_{i,\tau} + \alpha_{S,k,\tau}\theta_{S,i,\tau} + \upsilon_{S,k,i,\tau} \tag{4}$$

where $X_{i,\tau}$ denotes the set of observed characteristics affecting observed measure $T_{s,k}$ at time τ and $\alpha_{S,k,\tau}$ capture the factor loadings.[17] We follow the existing literature on factor models and assume that $X_{i,\tau} \perp \theta_{S,i,\tau} \ \forall \ S = \{C, NC\}, \tau \in \{t, t+1\}$. Moreover, $\upsilon_{S,k,i,\tau}$ represent mean-zero measurement errors with variance $\sigma^2_{\upsilon_{S,k,\tau}}$, which are independent across tests k, latent skills S and time periods τ, as well as across $X_{i,\tau}$ and $\theta_{S,i,\tau}$.

Identification. We first focus on the identification of the distribution of the latent skill factors at time t, $\widehat{F_{\theta_t^{NC},\theta_t^C}}$ and $\alpha_{S,k,t}$ in equation (4). Let us stack the scores in the measurement system at time t into $\mathbf{T}_t$ and the observed variables into $\mathbf{X}_{t,T}$. Then, note that the diagonal elements of the $[(\mathcal{K}_{NC} + \mathcal{K}_C) \times (\mathcal{K}_{NC} + \mathcal{K}_C)]$ matrix $COV\left(\mathbf{T}_t \,|\mathbf{X}_{t,T}\right)$ can be described by:

$$COV\left(T_{S,k,t}, T_{S,k,t} \,|\mathbf{X}_{t,T}\right) = \alpha^2_{S,k,t}\sigma^2_{\theta_{S,t}} + \sigma^2_{\upsilon_{S,k,t}} \tag{5}$$

where $\sigma^2_{\theta_{S,t}}$ indicates the variance of the latent factor S at time t. To describe its off-diagonal

[17] We have alternative estimated a triangular measurement system in which the set of non-cognitive measures loads on both factors. The estimated θ_C loadings on the observed non-cognitive measures were not significant, thus leading us to choose a dedicated measurement system. While the identification argument in this context involves additional steps, it largely follows the structure outlined in this section. See Sarzosa (2015) for additional details.

elements, first we split vector $\mathbf{T}_t$ into two blocks, where each block has measures dedicated to a different skill as follows:

$$\begin{bmatrix} \mathbf{T}_{NC,t} \\ \mathbf{T}_{C,t} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{t,T}\beta_{NC,t} + \Lambda_{NC,t}\theta_{NC,t} + \mathbf{v}_{NC,t} \\ \mathbf{X}_{t,T}\beta_{C,t} + \Lambda_{C,t}\theta_{C,t} + \mathbf{v}_{C,t} \end{bmatrix}$$

where $\Lambda_{S,t}$ is a vector containing the factor loadings in each block. We thus have two types of off-diagonal elements in the matrix $COV\left(\mathbf{T}_t \left| \mathbf{X}_{t,T}\right.\right)$, encompassing elements *within* a block:

$$COV\left(T_{S,k,t}, T_{S,k',t} \left| \mathbf{X}_{t,T}\right.\right) = \alpha_{S,k,t}\alpha_{S,k',t}\sigma^2_{\theta_{S,t}} \tag{6}$$

and across blocks:

$$COV\left(T_{NC,k,t}, T_{C,k',t} \left| \mathbf{X}_{t,T}\right.\right) = \alpha_{NC,k,t}\alpha_{C,k',t}\sigma_{\theta_t^{NC}\theta_t^{C}} \tag{7}$$

where $\sigma_{\theta_t^{NC}\theta_t^{C}}$ indicates the covariance of the latent factors at time t. We use the $(\mathcal{K}_{NC} + \mathcal{K}_C)(\mathcal{K}_{NC} + \mathcal{K}_C - 1)/2$ off-diagonal elements of $COV\left(\mathbf{T}_t \left| \mathbf{X}_{t,T}\right.\right)$ to recover the loadings, factor variances and covariance, yielding a total of $\mathcal{K}_{NC} + \mathcal{K}_C + 3$ parameters to be identified. Since latent factors have no location or scale of their own (Carneiro et al., 2003; Williams, 2019), we assume that $E(\theta_{S,i,t}) = 0$ and normalize the factor loading associated with one of the observed measures to unity ($\alpha_{S,1,t} = 1$). We consider three test scores for each skill measure S ($\mathcal{K}_{NC} = \mathcal{K}_C = 3$), implying that the number of off-diagonal elements in $COV\left(\mathbf{T}_t \left| \mathbf{X}_{t,T}\right.\right)$ suffice for identifying the unnormalized loadings, factor variances and covariance. Upon securing identification of the factor loadings and variances, we rely on the diagonal elements of $COV\left(\mathbf{T}_t \left| \mathbf{X}_{t,T}\right.\right)$ to identify $\sigma^2_{v_{S,k,t}}$.

We briefly outline the identification argument below. First note that the covariance term presented in equation (6), which includes the test score with the normalized loading, yields:

$$COV\left(T_{S,k,t}, T_{S,1,t} \left| \mathbf{X}_{t,T}\right.\right) = \alpha_{S,k,t}\sigma^2_{\theta_{S,t}}$$

Thus, we identify $\alpha_{S,k,t}$ and $\alpha_{S,k',t}$ from:

$$\frac{COV\left(T_{S,k,t}, T_{S,k',t} \left| \mathbf{X}_{t,T}\right.\right)}{COV\left(T_{S,k,t}, T_{S,1,t} \left| \mathbf{X}_{t,T}\right.\right)} = \frac{\alpha_{S,k,t}\alpha_{S,k',t}\sigma^2_{\theta_{S,t}}}{\alpha_{S,k,t}\sigma^2_{\theta_{S,t}}} = \alpha_{S,k',t}$$
$$\frac{COV\left(T_{S,k,t}, T_{S,k',t} \left| \mathbf{X}_{t,T}\right.\right)}{COV\left(T_{S,k',t}, T_{S,1,t} \left| \mathbf{X}_{t,T}\right.\right)} = \frac{\alpha_{S,k,t}\alpha_{S,k',t}\sigma^2_{\theta_{S,t}}}{\alpha_{S,k',t}\sigma^2_{\theta_{S,t}}} = \alpha_{S,k,t}$$

Having identified the loadings, we can identify $\sigma^2_{\theta_{NC,t}}$ and $\sigma^2_{\theta_{C,t}}$ from any covariance between tests scores of the NC and C bocks, respectively. Furthermore, we can identify $\sigma_{\theta_t^{NC}\theta_t^C}$ from the covariance between the tests that have a normalized loading: $COV\left(T_{NC,1,t}, T_{C,1,t} \left|\mathbf{X}_{t,T}\right.\right) = \sigma_{\theta_t^{NC}\theta_t^C}$. The identification of the factor loadings and factor variances allows us to use the diagonal elements (5) of the covariance matrix to identify the residuals' variances. Having identified all the loadings and variances, we rely on the argument put forth by Freyberger (2018) and applied by Heckman et al. (2016) to non-parametrically identify $F_{\theta_t^{NC},\theta_t^C}$ from the manifest variables $\mathbf{T}_t$.[18]

We follow a similar argument to identify the loadings, distribution of the latent factors and error terms at $t+1$. However, the identification strategy presented above normalizes the location and scale of latent skills at $t+1$ ($E[\theta_{S,i,t+1}] = 0$ and $\alpha_{S,1,t+1} = 1$, respectively). In this setting, Agostinelli and Wiswall (2016b) have shown that such 're-normalizations' would bias the parameters of the CES skill production function (equation 1) towards a Cobb-Douglas functional form. In Section 4, we present an estimation strategy which allows us to address this concern while incorporating the underlying Roy model of grade retention. Nonetheless, the essential identification strategy for the model parameters at $t+1$ follows the arguments for period t presented above.[19]

3.3.2 Parents' Measures

As discussed in Section 2, we have access to multiple measures on parental well-being. We similarly take advantage of these measures to posit a measurement system in which observed measures of parents' psychological well-being $T_{p,i,t}$ ($p \in \mathcal{P}$) measure latent parental socioemotional skills ($\theta_{P,i}$) with error as follows:

$$T_{p,i,t} = \beta_{p,t} X_{p,i,t} + \alpha_P \theta_{P,i} + \upsilon_{p,i,t} \tag{8}$$

where $\upsilon_{p,i,t}$ are mean-zero measurement errors which are independent across parental skill measures p, observed characteristics and latent parental socioemotional skills. In particular, we take advantage of the three measures of parents' psychological well-being available in the baseline survey round, where the first variable measures parents' mental well-being, whereas the other two variables encompass measures of parental anxiety. Since we have access to three observed measures of parents' skills, we can follow the identification argument outlined above to recover the distribution of the latent parental socioemotional skills factor.

A similar argument applies to the observed measures of parental investments. As noted

[18] As discussed below, we impose flexible parametric assumptions in the estimation process.

[19] We take advantage of the three test score and teacher-reported non-cognitive skill measures available in the endline survey round. Across both time periods, we normalize the factor loadings associated with the reading assessment and the SRS non-cognitive measure.

above, we observe measures four measures of parents' investments in their kids, which capture the number of books available in the household, the number and the types of activities done with their children, their engagement in school activities, and their parenting attitudes / their relationship with their child. Each observed investment measures $I_{k,i,t} k \in \mathcal{K}$ represents an error-ridden measure of latent parental investment in:

$$I_{k,i,t} = \beta_{k,t} X_{k,i,t} + \alpha_I \mathcal{I}_{i,t} + v_{k,i,t} \tag{9}$$

where $v_{k,i,t}$ are mean-zero measurement errors which are independent across investment measures k, observed characteristics and latent parental investments. We note that after recovering the distribution of latent parental investments, we estimate equation (3) to account for the potential endogeneity of parental investment to their children's latent skills at time t. In particular, following Cunha et al. (2010); Agostinelli and Wiswall (2016a), we first estimate equation (3) and subsequently include predicted $\widehat{\mathcal{I}_{i,t}}$ as an input in the production function.

4 Estimation

We follow a two-step estimation procedure as in Sarzosa (2015). In the first step, we recover the distribution of the latent factors and investment measures, along with the factor loadings and coefficients on observables using a Maximum Likelihood estimator. In this step, we additionally estimate the retention equation along with the parental investment equationss. Based on these results, in a second step, we estimate the skill production functions.

4.1 Latent Factors

We estimate the distribution of the latent skill factors at time t ($\widehat{F_{\theta_t^{NC},\theta_t^{C}}}$), $t+1$ ($\widehat{F_{\theta_{t+1}^{NC},\theta_{t+1}^{C}}}$), latent parental socioemotional skills ($\widehat{F_{\theta_t^{P}}}$) and latent investments ($\widehat{F_{\mathcal{I}_t}}$) each following a mixture of two normals, which imposes few restrictions on the underlying distribution of the factors. While Freyberger (2018) shows that the latent factors are identified non-parametrically, using a mixture of normals allows for numerical integration using the Gauss-Hermite quadrature (Judd, 1998).[20] We additionally assume the error terms in the measurement system ($\boldsymbol{v}_{S,\tau}$), parental socioemotional skill equations ($\boldsymbol{v}_P$), parental investment equations ($\boldsymbol{v}_K$), retention equation ($\boldsymbol{\varepsilon}_t$) and in the skills production functions ($\boldsymbol{\eta}_{S,t}^{R}$) are normally distributed.

[20]We remark that we rely on numerical integration based on the estimated distribution of the factors throughout the estimation procedure due to the unobservable nature of the latent factors.

4.2 Production Function

In the second step, we estimate the CES production function outlined in equation (1). Since we have already recovered the coefficients and factor loadings in the measurement system, we can construct the following vector for each latent skill factor S:[21]

$$\widehat{\boldsymbol{\xi}_{S,t+1}} = \mathbf{T}_{S,t+1} - \widehat{\boldsymbol{\beta}_{S,t+1}}\mathbf{X}_{t+1} = \widehat{\boldsymbol{\alpha}_{S,t+1}}\theta_{S,t+1} + \boldsymbol{\upsilon}_{S,t+1} \tag{10}$$

As explained in Section 3, our model estimates retention-status dependent skill production functions (equation (1)), and as such, it allows for endogenous selection into retention based on initial parental and children's skills, observable characteristics and school-level retention policies (equation (2)). Then, we write equation (10) as retention dependent:

$$\widehat{\xi^0_{S,k,i,t+1}} = \widehat{\alpha_{S,k,t+1}}\theta^0_{S,i,t+1} + \upsilon_{S,k,i,t+1} \qquad \text{if } R_{i,t} = 0 \tag{11}$$

$$\widehat{\xi^1_{S,k,i,t+1}} = \widehat{\alpha_{S,k,t+1}}\theta^1_{S,i,t+1} + \upsilon_{S,k,i,t+1} \qquad \text{if } R_{i,t} = 1 \tag{12}$$

Since we have assumed the error terms in the measurement system at $t, t+1$ and the production function are orthogonal and mutually independent from each other ($\boldsymbol{\upsilon}_{S,t+1} \perp \boldsymbol{\upsilon}_{S,t} \perp \boldsymbol{\eta}^R_{S,t}$), we replace $\theta^R_{S,t+1}$ by its production function $g^R_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t)$ into equations (11) and (12), yielding:[22]

$$\widehat{\xi^0_{S,k,i,t+1}} = \widehat{\alpha_{S,k,t+1}}g^0_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t) + \zeta^0_{S,k,i,t+1} \qquad \text{if } R_{i,t} = 0 \tag{13}$$

$$\widehat{\xi^1_{S,k,i,t+1}} = \widehat{\alpha_{S,k,t+1}}g^1_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t) + \zeta^1_{S,k,i,t+1} \qquad \text{if } R_{i,t} = 1 \tag{14}$$

where $\boldsymbol{\zeta}^R_{S,t+1} = \widehat{\boldsymbol{\alpha}_{S,t+1}}\boldsymbol{\eta}^R_{S,t} + \boldsymbol{\upsilon}_{S,t+1}$ is a compounded error term with mean zero and variance $\Omega_{\zeta^R_{S,t+1}}$. Note that its off-diagonal elements are given by $\widehat{\alpha_{S,k,t+1}}\widehat{\alpha_{S,k',t+1}}\sigma^2_{\eta^R_{S,t}} + \sigma^2_{\upsilon_{S,t+1}}$ and the diagonal elements follow $(\widehat{\alpha_{S,k,t+1}})^2\sigma^2_{\eta^R_{S,t}} + \sigma^2_{\upsilon_{S,k,t+1}}$. Since the elements across both the diagonal and off-diagonal are identified in the first estimation step, it is straight-forward to see that $\Omega_{\zeta^R_{S,t+1}}$ is identified as well.

4.3 Factor Normalizations

As discussed in Section 3.3, Agostinelli and Wiswall (2016b) show that assuming $E[\theta_{S,i,t+1}] = 0$ and $\alpha_{S,1,t+1} = 1$ biases the estimation of the production function. In particular, fixing the

[21] We omit individual i's subscripts for notational simplicity and note that variables in bold denote the vector encompassing all tests $k \in \mathcal{K}$ dedicated to latent skill S.

[22] For notational simplicity, we let $g^R_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t)$ represent the CES production function for latent skill factor S across retention status $R \in \{0, 1\}$.

location of the latent skills at $t + 1$ introduces a mean stationarity restriction that limits the possible functions that can be estimated to only those in the Cobb-Douglas family (i.e., $\rho_S^R \to 0$ in equation (1)). Re-normalizing the scale of the latent skills at $t + 1$ (i.e., $\alpha_{S,1,t+1} = 1$) makes the production function estimates vulnerable to biases stemming from possible differences in the scaling of the measures ($T_{S,1,t}$ and $T_{S,1,t+1}$) where normalizations were imposed. Our estimation strategy allows us to address both issues.

Agostinelli and Wiswall (2016a) propose an approach to identify skill production functions without re-normalizing skills at $t + 1$ by using transformations of moments to get rid of $E\left[T_{S,k,\tau}|\mathbf{X}_\tau\right]$ and $\alpha_{S,k,\tau}$ in equation (4). We cannot follow their approach, given the complexity of our setting, as we include an endogenous treatment that affects the production functions themselves. We rely on the non-observability of skills to implement a Roy model-like approach where retention is endogenous to those initial skills. Thus, we cannot use the moment transformations proposed by Agostinelli and Wiswall (2016a) with which they clean the measurements that allow them to treat the production function's inputs as observable.

Instead, we deal with the location issue by following Sarzosa (2015) and relying on the fact that the relationship between $E(\theta_{S,i,t+1})$ and the production function parameters is directly predictable by a quartic polynomial in the CES parameters ($\mathbf{P}_4(\gamma_S, \gamma_P, \gamma_I, \rho_S)$), as shown in Figure B.1. As a result, we use $\mathbf{P}_4(\gamma_{NC,S}^R, \gamma_{P,S}^R, \gamma_{I,S}^R, \rho_S^R)$ as a shifter of the mean of $\theta_{S,i,t+1}$ during estimation to avoid constraining the possible values of the CES production function towards a Cobb-Douglas functional form. That way, we counter the mechanical mean-shifting taking place when $\widehat{\rho} \neq 0$. We thus allow for $E[\xi_{S,k,t+1}^R] = -\alpha_{S,k,t+1}^R \mathbf{P}_4(\gamma_{NC}^R, \gamma_P^R, \gamma_I^R, \rho_S^R)$.

We note that assuming $E[\theta_{S,i,t+1}] = 0$ implies that the estimated parameters in the production function will not respond to aggregate mean changes in skills. This assumption does not affect our analysis, since our goal is to recover the impacts of retention on the skill development process, not overall shifts in skill means. We deal with the scale issue in a novel way. We take advantage of the fact that we are interested in comparing the skill trajectories of retained with those of non-retained children. Thus, we free the scale of the production functions of grade repeaters so that it is interpreted relative to the scale of the production functions of the non-retained. That is, we allow for a wedge ψ_S in the relationship between skills and scores for the retained. Equation (14) becomes:

$$\widehat{\xi_{S,k,i,t+1}^1} = \widehat{\alpha_{S,k,t+1}} \psi_S g_{S,t+1}^1(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t) + \zeta_{S,k,i,t+1}^1 \qquad \text{if } R_{i,t} = 1 \tag{15}$$

We interpret the wedge ψ_S as the returns to scale of the production function of skill S among the retained relative to the scale of the production function among the non-retained.

4.4 Model Implementation

Table B.1 shows the variables used in the implementation of the model. In the measurement system for test scores in both periods as well as in the retention decision, we include family background measures, such as household composition and parents' education, child-level characteristics, including gender, age and race and teacher's educational attainment. The observed measures used in the measurement system encompass the three test score and non-cognitive skill measures available at baseline and endline, as outlined in Section 2. The retention decision additionally includes school-level retention policies along with school characteristics, including the share of free and reduced price lunch students, whether the school is public or private and whether the school is in an urban area. The parental socioemotional and investment equations additionally include information on parents' age in the baseline survey, and the investment equation additionally incorporates household income as an exclusion restriction.

5 Model Results

5.1 Goodness of Fit

We first examine the accuracy of the model in predicting retention and test score outcomes. In the first column of Table B.2, we contrast students' observed and simulated retention outcomes, showing the model approximates the likelihood of retention across survey rounds (5.77%). We further take advantage of estimated model parameters to simulate students' observed skill measures. As shown in the remaining columns of Table B.2, the model closely matches the first and second moment of observed test scores and non-cognitive skill measures at baseline. In Figure B.2, we confirm these results by showing that the distribution of observed and simulated skill measures largely mirror each other.

5.2 Variance Decomposition

Table B.3 presents the estimated parameters of the measurement system at time t. Across the math, reading and science measures, older children as well as those coming from higher-SES families consistently earn higher test scores—with similar patterns emerging in the three observed non-cognitive skill measures. Furthermore, the latent skill factors at baseline load positively on the dedicated observed measures.

In the first panel of Figure 1, we present a variance decomposition of the measurement system to understand the relative importance of latent skills and background characteristics in determining test scores. First, while observable characteristics account for 20-30% of the variance in test scores, they explain 6-10% of the variance in teacher-reported non-cognitive skll measures, fitting in with evidence presented by Heckman et al. (2006). Latent cognitive ability explains an important share of the variance in test scores, ranging from 17% in science to close to 60% in math. Similarly, $\theta_{NC,t}$ accounts for 45-77% of the variance in the baseline non-cognitive skill measures. Lastly, 13-57% of the variance of observed measures in both skill domains can be attributed to the error term, remarking the extent to which test scores capture latent abilities with significant measurement error.

Figure 1: Variance Decomposition of Test Scores and Skills Distribution

(a) Variance Decomposition of Test Scores at t **(b)** Variance Decomposition of Test Scores at $t+1$

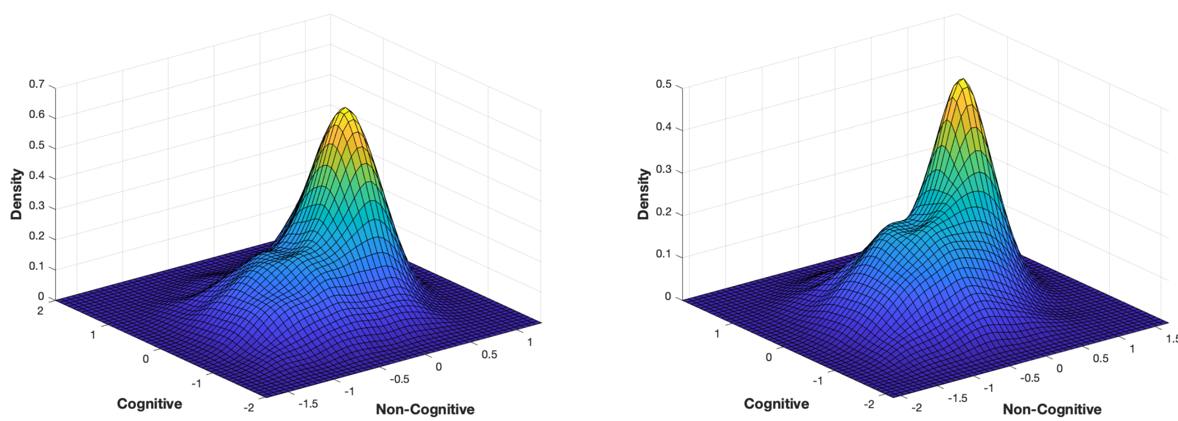
Note: The first panel of Figure 1 presents the share of the variance of test scores and non-cognitive skill measures explained by observable controls $\mathbf{X}_T$, latent cognitive and non-cognitive skills $\Theta_t = \left[\begin{array}{cc} \theta_{C,t} & \theta_{NC,t} \end{array} \right]$, and the share captured by the unobserved idiosyncratic error of the measurement system (Residuals) The second panel presents corresponding evidence from a variance decomposition of observed test scores and non-cognitive skill measures at $t+1$.

In the endline survey round, we similarly find higher test scores and non-cognitive skill outcomes for children from two parent and higher-SES households (Table B.4). In the second panel of Figure 1, we present a variance decomposition of test scores in the follow-up survey. For both cognitive and non-cognitive measures, observed characteristics and latent factors explain a sizable share of the variance. Nonetheless, the share of the variance all six observed measures explained by the error term ranges from 18% to 36%. Measurement error thus plays a critical role in our setting, remarking the limitations of empirical approaches which consider test scores — rather than latent abilities — as outcome variables when examining the impacts of retention across the skill development process.

In the first panel of Figure 2, we present the joint distribution of latent skill factors at time t. The model flexibly estimates latent factors following a mixture of normals, and both factors exhibit significant deviations from normality. We further allow for cognitive and non-cognitive skills to be correlated and find the correlation coefficient between $\theta_{C,t}$ and $\theta_{NC,t}$ equals 0.315. The second panel of Figure 2 similarly shows the joint distribution of latent skills at $t+1$ deviate significantly from normality. We find a significant correlation between both factors, reaching 0.305.

Figure 2: Joint Distribution of Latent Skill Factors

(a) Joint Distribution at t **(b)** Joint Distribution at $t+1$



Note: The first panel of Figure 2 presents the estimated latent skills joint distribution at t $[f_{\theta_{C,t},\theta_{NC,t}}(\cdot,\cdot)]$. It was obtained from a random draw of the estimated parameters from the measurement system at time t (presented in Table B.3). The distribution is centered at $(0,0)$. The correlation coefficient between cognitive and non-cognitive skills is 0.3149. The standard deviation of the non-cognitive skills marginal distribution is 0.554 and that of the cognitive skills distribution is 0.660. In the second panel, we present corresponding evidence for the joint distribution of latent factors at time $t+1$. Results follow from a random draw based on the estimates of the measurement system at time $t+1$ (Table B.4). The distribution is centered at $(0,0)$. The correlation coefficient between cognitive and non-cognitive skills is 0.3047. The standard deviation of the non-cognitive skills marginal distribution is 0.635 and that of the cognitive skills distribution is 0.635. Values in the top and bottom 1% in both dimensions were excluded in both panels.

5.3 Parental Skills and Investment

The empirical model specified in Section 3 further incorporates the importance of parents' socioemotional skills in the skill formation process. We present the estimated parameters of equation (8) in Table B.5. Parents in two-parent household exhibit higher measures of psychological well-being. Moreover, the factor loadings across all measures are positive and significant. In Table B.6, we present the estimated parameters of the investment measurement system, finding that all observed investment measures exhibit positive loadings on the latent investment factor.

Our empirical strategy further allows for parental investment choices to respond endogenously to children's latent skills at time t. We present the estimated parameters of equation

Table 3: Estimated Parameters of Parental Investment Equation

	Latent Investment	
	Coefficient	Std.Err.
HH Income	0.041***	(0.016)
Both Parents	-0.004	(0.029)
Parents' Education	0.089***	(0.006)
Male Parent	-0.225***	(0.040)
Parent's Age	0.002	(0.002)
URM Parent	-0.198***	(0.029)
Male Child	-0.050**	(0.024)
Child Age (R1)	-0.003	(0.003)
θ_P	0.045***	(0.009)
θ_{NC}	0.035*	(0.021)
θ_C	0.119***	(0.020)
Observations	9,010	

Note: Table 3 presents estimated coefficients of observed characteristics, parents' skills and children's cognitive and non-cognitive skills from the parental investment equation. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

(3) in Table 3. First, higher income parents exhibit higher investment in their children's skill development, fitting in with evidence presented by (Cunha et al., 2010; Attanasio et al., 2020). At the same time, more educated parents invest more in their children, conditional on other observed and unobserved characteristics. Importantly, we find that parental investments reinforce children's baseline cognitive skills, as a one standard deviation increase in $\theta_{C,t}$ is associated with a 0.12 σ increase in parents' latent investments. Parents also reinforce their children's non-cognitive skills, yet the estimated coefficient is smaller in magnitude.

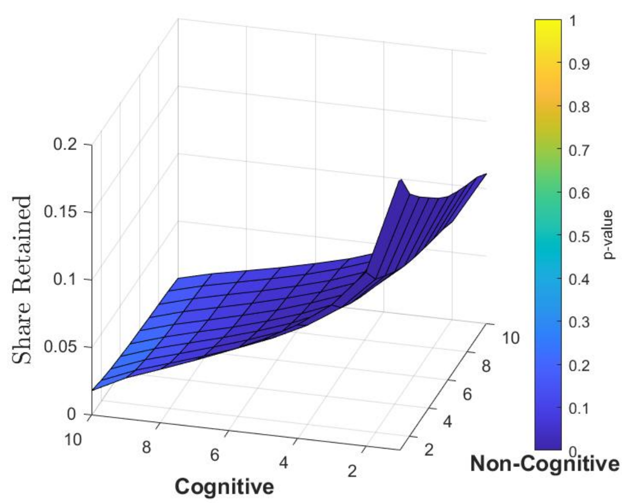
6 Grade Retention and Dynamic Skill Development

6.1 Determinants of Grade Retention

In Table B.7, we present marginal effects from the retention equation. Conditional on other observed characteristics, males and younger children are significantly more likely to have repeated an early grade. At the same time, children growing up in two parent households as well as those with more educated parents have a lower conditional likelihood of early retention. Conditional on other school characteristics, retention policies play an important role in driving the likelihood of early repetition, as children in schools which allow for retention based on parents' explicit requests and those which allow for retention without parents' consent are more likely to have repeated a grade.

Our results further remark the critical role played by students' latent abilities at time t in the retention process. First, a one standard deviation increase in students' cognitive ability reduces the likelihood of retention by four percentage points. We find sizable differences in the non-cognitive dimension as well, where a one σ increase in this dimension reduces the likelihood of retention by 1.1 pp. These findings are confirmed graphically in Figure 3, which shows that upwards of 11% of students in the bottom decile of the $\theta_{C,t}$ distribution are retained, with additional heterogeneity in the non-cognitive skill distribution: the conditional likelihood of retention increases from 11.2% for students in the top $\theta_{NC,t}$ decile to 20.1% for their peers in the bottom decile. Similarly, Figure B.3 presents the marginal distributions of cognitive and non-cognitive ability across retention status. Selection-into-retention patterns largely fit in with prior findings in the literature, as Gary-Bobo et al. (2016) and Cockx et al. (2019) find that students with lower cognitive ability are more likely to have been retained, and Fruehwirth et al. (2016) document similar results for non-cognitive skills in the earlier ECLS:K round. We lastly note that latent parental socioemotional skills have no discernible impact on the prevalence of early retention.

Figure 3: Probability of Retention at t by Initial Level of Skills



Note: Figure 3 presents the estimated probability of retention across joint deciles of the latent skills distribution in period t.

6.2 Skill Production Process

In Table 4, we present the parameters which govern the process of skill formation across children's retention status. The first row presents the estimated elasticity of substitution in the production of non-cognitive and cognitive skills, respectively, across retention status. For non-retained students, the estimated production function of $\theta_{NC,t+1}$ is not different from Cobb-Douglas, yet this is not the case for their retained counterparts, for whom the estimated σ^1_{NC} equals 1.68. The production of cognitive skills shows similar results. Among non-repeating students, the estimated elasticity of substitution is slightly larger than one, whereas for retained students, we find a lower σ^1_C, equaling 0.745.

We remark that despite normalizing the location of the factors ($E[\theta_{S,i,t+1}] = 0$), we recover elasticities of substitution which are different from one across three production functions, thus showing that our estimation approach overcomes the bias-towards-Cobb-Douglas issue documented in Agostinelli and Wiswall (2016b). All in all, the estimated production functions show that period t inputs are complementary in the production of skills at endline.

Table 4: Estimated Parameters of CES Production Function

	$\theta_{NC,t+1}$		$\theta_{C,t+1}$	
	$R_i = 0$	$R_i = 1$	$R_i = 0$	$R_i = 1$
	(1)	(2)	(3)	(4)
$\sigma^R_s = \frac{1}{1-\rho^R_s}$	0.991	1.684	1.059	0.745
	[0.966,1.016]	[1.467,1.900]	[1.030,1.088]	[0.649,0.840]
$\gamma_{NC,t}$	0.680	0.610	0.035	0.056
	[0.661,0.699]	[0.530,0.686]	[0.022,0.055]	[0.022,0.133]
$\gamma_{C,t}$	0.093	0.184	0.886	0.720
	[0.085,0.101]	[0.150,0.217]	[0.878,0.893]	[0.705,0.736]
$\gamma_{I,t}$	0.214	0.167	0.085	0.231
	[0.193,0.236]	[0.091,0.288]	[0.067,0.108]	[0.147,0.342]
γ_P	0.013	0.038	-0.006	-0.006
	[0.005,0.021]	[0.005,0.072]	[-0.014,0.001]	[-0.022,0.009]

Note: Table 4 presents the estimated parameters of equation (1) across retention status for the production of cognitive and non-cognitive skills at $t+1$. σ^R_s is the elasticity of substitution between skill endowments, parental skills and investments in retention status R. 95% confidence intervals are reported in brackets.

The following two rows present evidence on the self- and cross-productivity of latent cognitive and non-cognitive skills across retention status. First, non-cognitive skills are self-producing, as the estimated γ^R_{NC} parameter in the production function of non-retained and retained students equals 0.680 and 0.610, respectively. At the same time, period t cognitive skills have smaller *direct* influence on students' endline non-cognitive skills, yet the positive coefficients are consistent with the cross-productivity of skills. We document similar results in the production of cognitive skills. We find strong evidence of self-productivity of θ_C, especially among non-retained students, for whom $\gamma^1_{C,t}$ equals 0.886. We also find evidence of the cross-productivity in the production function of cognitive skills for both retained and non-retained students.

The fourth row shows that parents' investments play an important role in the skill development process. Across both retention outcomes, the estimated γ parameters exceed 0.16 in the production of non-cognitive skills. Meanwhile, parents' investments play a smaller role in the development of retained children's cognitive skills, while exhibiting a far larger impact for their non-retained peers. The positive coefficients on parents' investments further remark the extent to which children's period t skills drive outcomes at $t+1$ given the evidence of reinforcing investments documented in Table 3. Lastly, parents' socioemotional skills (θ_P) have a limited direct impact on their children's skill outcomes in period $t+1$. Nonetheless, as we had previously found (Table 3) that parents with higher θ_P engage in increased investments for their children, parents' socioemotional skills indirectly improve their children's skill outcomes through their investment decisions.

We present graphical evidence of the *total* productivity of latent skills in Figure B.4. While self-productivity dominates the skill production process, baseline cognitive skills increase the marginal returns to non-cognitive skills in the production of $\theta_{NC,t+1}$ ($\frac{\partial^2 \theta_{NC,t+1}}{\partial \theta_{NC,t} \partial \theta_{C,t}}$)—this result similarly emerges for the production of $\theta_{C,t+1}$ and holds across retention outcomes. These results fit in with existing evidence on the production of non-cognitive skills across different contexts (Cunha et al., 2010; Sarzosa, 2015), yet we extend the existing literature by considering differences in the production function across retention outcomes. The evidence presented so far show that the skill development process exhibits self- and cross-productivity. Moreover, parents' investments affect children's skill outcomes at endline, and the importance of initial inputs varies across retention status.[23]

7 Estimated Impacts of Grade Retention

The empirical strategy outlined in Section 3 allows us to recover counterfactual skill development outcomes across retention histories. We take advantage of the estimated model parameters to estimate various treatment effects of retention on cognitive and non-cognitive skills at endline. In this context, the estimated average treatment effect (ATE) of retention is given by:

$$ATE_S \equiv \iint E[g^1_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t) - g^0_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t)] dF_{\theta_C, \theta_{NC}}(\underline{\theta}_{C,t}, \underline{\theta}_{NC,t}) \quad (16)$$

While equation (16) allows us to recover the average estimated impacts of retention on skill dimension S, the characteristics of retained students differ significantly from their non-retained peers. We may thus be interested in further examining the impacts of retention on students who were in fact retained (Fruehwirth et al., 2016; Gary-Bobo et al., 2016; Cockx et al., 2019). This object is given by the treatment on the treated (TT) parameter, defined

[23] In Appendix C, we present evidence on the unconditional productivity of period t inputs on unconditional skill outcomes at $t+1$. We find strong evidence of self-, cross-productivity as well as an important role for parental investments in shaping children's skill outcomes at endline.

as follows:

$$TT_S \equiv \iint E[g^1_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t) - g^0_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t)|R = 1] dF_{\theta_C,\theta_{NC}}(\underline{\theta}_{C,t}, \underline{\theta}_{NC,t}) \tag{17}$$

The treatment effect parameters defined above are calculated by integrating across the latent ability distribution, yet the impacts of retention may vary across the $\boldsymbol{\theta_t}$ vector, depending on the productivity of skills across retention outcomes. We thus consider the heterogeneous effects of retention by initial skill level as follows:

$$ATE_{S,(\underline{\theta}_{C,t},\underline{\theta}_{NC,t})} \equiv E[g^1_{S,t+1}(\cdot) - g^0_{S,t+1}(\cdot)|\theta_{C,t} = \underline{\theta}_{C,t}, \theta_{NC,t} = \underline{\theta}_{NC,t}]$$

7.1 Effects on Cognitive Skill Development

We present the estimated average treatment effect of grade retention on children's cognitive skill development in the first row of Table 5. Early retention negatively impacts children's cognitive skill development, as the estimated ATE indicates that repeating a grade lowers $\theta_{C,t+1}$ by 0.019 standard deviations. The estimated impact of retention on cognitive skills at $t+1$ is significantly smaller than the estimated effects in the OLS results presented in Table 2. The difference in the estimates emerges not only from the fact that our model incorporates school-level policies, latent parental skills and endogenous investment choices, but also from the prevalence of selection-into-retention on unobservables (Figure 3). In fact, since retained students trail their non-retained peers across both skill dimensions, the estimated effect for students who were retained may differ substantially from the average population effects. As such, we additionally estimate the treatment on the treated parameter, finding that retention lowers retained students endline cognitive skills by 0.147 standard deviations.

The existing literature has so far found mixed evidence on the impacts of early retention on academic achievement. For instance, Fruehwirth et al. (2016) find small, yet significant losses in math and reading test scores arising from retention before second grade, in the range of 5-10 percent. On the other hand, Schwerdt et al. (2017) find that third grade retention increases short term GPA for students in Florida, Jacob and Lefgren (2004) show that retention leads to short-term gains in math test scores for Chicago students, in the range of 33 percent and early retention in France leads to short-term gains in achievement, which become negative in the medium-term (Alet et al., 2013). To provide comparable estimates to the existing literature, we re-estimate the model presented in Section 3 using endline test scores as outcomes.[24] In this specification, grade retention would lower students' math,

[24]We take advantage of model parameters to estimate a Roy model with unobserved heterogeneity as in

reading and science test scores by 0.197, 0.347 and 0.026 standard deviations, respectively (Figure B.5). The estimated (negative) impacts of retention would far exceed those estimated under our preferred model, which differentiates test scores from latent skills. Following the evidence presented in Figure 1, these results further highlight the extent to which test scores measure latent abilities with error.

Table 5: Estimated Impacts of Grade Retention

	$\theta_{C,t+1}$	$\theta_{NC,t+1}$
ATE	-0.0191**	0.0414***
	(0.0094)	(0.0103)
ATT	-0.1467***	0.0279***
	(0.0094)	(0.0103)
PRTE	-0.1231	0.0322
	(0.1655)	(0.1817)

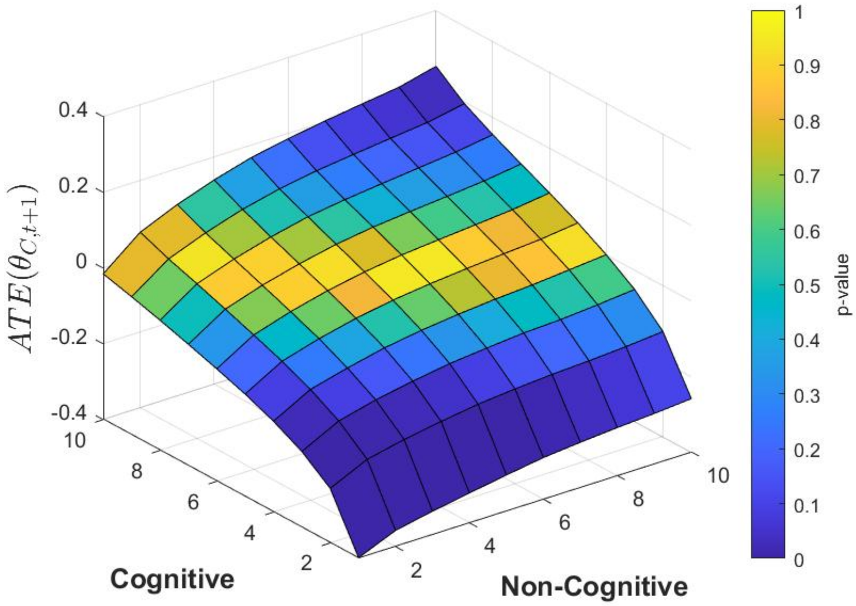
Note: Table 5 presents the estimated effects of grade retention on children's cognitive (column 1) and non-cognitive (column 2) skill outcomes in period $t+1$. The first row presents the estimated average treatment effect (equation (16)). The second row presents the treatment on the treated parameter (equation (17)). Standard errors reported in parenthesis: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The average effects of retention are estimated by integrating across the latent ability distribution, yet there may be heterogeneous impacts of the event across students' baseline cognitive and non-cognitive skills. We thus present the estimated heterogeneous effects of retention on $\theta_{C,t+1}$ across the initial skill distribution in Figure 4. The impacts of grade retention on cognitive skills exhibit significant heterogeneity across the period t skill distribution. For instance, for students in the bottom decile of the joint distribution, repeating an early grade would lower these students' endline cognitive skills by 0.404 standard deviations. Nonetheless, the estimated effects of early retention are increasing across both dimensions of children's initial skills, such that the estimated effects are not different from zero for students in the median of the baseline skill distribution. In fact, we find positive impacts for students in the top decile of the joint skill distribution, for whom repeating a grade would increase their latent cognitive skills by 0.266 σ.

Since grade retention is geared towards improving academic outcomes for struggling students, finding increasing returns to repetition across the skill distribution may be surprising at first. We first note that Fruehwirth et al. (2016) similarly find larger impacts of early grade retention on students' math and reading test scores in the United States, and grade retention for eighth graders in Belgium has the most adverse consequences for low-ability

Heckman et al. (2006, 2018).

Figure 4: Average Treatment Effect of Retention on $t+1$ Cognitive Skills



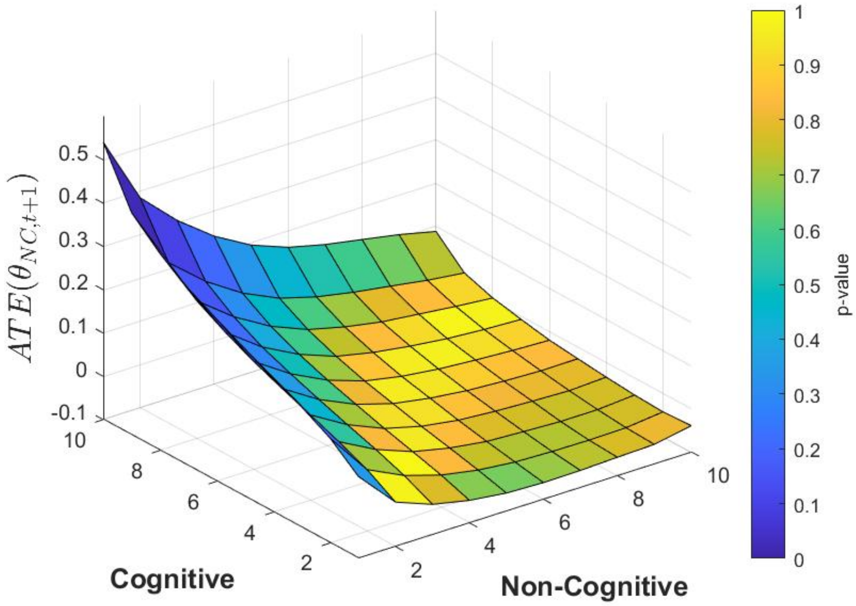
Note: Figure 4 presents the estimated impacts of grade retention on students' cognitive skill development ($\theta_{C,t+1}$) across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

students (Cockx et al., 2019). This finding may emerge if high-skilled students are best positioned to take advantage of retention, by becoming one of the highest-ability students in their new peer group, thus receiving additional benefits from repeated exposure to the same material. Alternatively, retention may drive parents to increase their investments, thus mitigating the adverse impacts of the event.

7.2 Effects on Non-Cognitive Skill Development

As discussed above, grade retention implies that students will be separated from their classmates, potentially resulting in stigmatization and thus affecting their non-cognitive skill development. In the second column of Table 5, we present evidence on the estimated impacts of retention on students' non-cognitive skill outcomes. Early grade repetition results in small increases in children's non-cognitive skills, as the estimated average treatment effect improves children's $\theta_{NC,t+1}$ by 0.041 standard deviations. The estimated ATE follows the same sign as the OLS estimates presented in Table 2, yet is smaller in magnitude vis-à-vis reduced form estimates. We additionally present evidence on the treatment on the treated parameter, which indicates that retained students enjoyed small benefits from being held back, as the estimated TT parameter equals 0.028 standard deviations.

Figure 5: Average Treatment Effect of Retention on $t + 1$ Non-cognitive Skills



Note: Figure 5 presents the estimated impacts of grade retention on students' non-cognitive skill development ($\theta_{NC,t+1}$) across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

Figure 5 extends the analysis to examine heterogeneous impacts of retention on students' non-cognitive skills across the initial latent skill distribution. For students above the cognitive and non-cognitive skill median, the impacts of retention on $\theta_{NC,t+1}$ are not different from zero. The estimated ATE becomes larger for students below the $\theta_{NC,t}$ median, while exhibiting limited heterogeneity across the cognitive ability distribution. As a result, for students in the bottom decile of the non-cognitive skill distribution, grade retention boosts their period $t + 1$ skills by upwards of 0.1 σ, reaching 0.50 standard deviations for low non-cognitive skilled students who are in the top decile of the cognitive skill distribution. This result may emerge in a context in which retention leads low self-confidence students to better understand class material, thus gaining confidence in their learning abilities and boosting their endline non-cognitive skills.

As discussed earlier, while an extensive literature has considered the impacts of retention on academic outcomes, there is limited evidence on the effects of this practice on students' non-cognitive skill development. Eren et al. (2018) find that retention for eighth graders in Louisiana increases the number of absences and the prevalence of disciplinary incidents.[25]

[25]A larger strand of the literature has found mixed evidence on non-academic outcomes. Ozek (2015) finds that grade retention for third graders in Florida increases the likelihood of suspensions in the short run, yet the effects do not persist over time. Eren et al. (2017) find that grade retention in eighth grade reduces the probability of being convicted for a juvenile crime, whereas Eren et al. (2018) find that opposite

These papers present preliminary evidence on how repetition affects non-cognitive skills, yet individual measures of non-cognitive skills capture latent skills with substantial error, as documented in Figure 1b. To assess the importance of measurement error, we re-estimate our model using teacher-reported endline non-cognitive skill measures as outcomes. This model indicates that retention would boost students' attentional focus, inhibitory control and SRS by 0.131, 0.048 and 0.025 σ (Figure B.6), respectively, exceeding the estimated impacts under the skill formation model presented in Section 3. All in all, our results show that grade retention differentially impacts cognitive and non-cognitive abilities, while resulting in heterogeneous impacts across students' initial skill endowments.

7.3 Policy-Relevant Treatment Effects

Conceptual Framework. The various treatment effects presented so far may not necessarily correspond to policy relevant parameters, as changes in retention policies may affect students with different observed and unobserved characteristics relative to the full sample. In this context, we adapt the literature on policy-relevant treatment effects to a dynamic skill formation model. We thus consider policy changes which do not directly affect outcomes (Heckman and Vytlacil, 2001; Carneiro et al., 2010; Mogstad et al., 2018).

Let $R_i(p)$ represent student i's retention outcome under retention policy regime p. Under an alternative policy p' ($p' \in \mathcal{P}$), retention outcomes are captured by $R_i(p')$. We consider policies which reduce the likelihood of retention, such that policy compliers are defined by $\{R_i(p) = 1, R_i(p') = 0\}$.[26] Our empirical framework incorporates school-level retention policies as drivers of the retention process. As such, we examine the impacts of no longer allowing schools to retain children without their parents' consent. The policy relevant treatment effect (PRTE) from switching to policy regime p' on children's latent skill outcomes at period $t+1$ is given by:

$$PRTE_S(p,p') \equiv \iint E[\theta_{S,t+1}(p') - \theta_{S,t+1}(p) | R(p) = 1, R(p') = 0] dF_{\boldsymbol{\theta}}(\underline{\boldsymbol{\theta}} | R(p) = 1, R(p') = 0) \quad (18)$$

where $\theta_{S,i,t+1}(p)$ captures student i's latent skills $S \in \{C, NC\}$ at period $t+1$ under policy regime p. Equation (18) allows us to further recover the observed and unobserved characteristics of policy compliers (Heckman et al., 2018). Since the PRTE directly depends on the distribution of compliers' latent skills at period t, we can thus define complier weights,

result on the likelihood of violent crime conviction among adults. Meanwhile, Diaz et al. (2016) find that retention in Chilean primary schools leads to sizable increases in the prevalence of juvenile crime.

[26] This analysis carries an implicit monotonicity assumption, which rules out the presence of defiers.

which capture the share of compliers joint distribution of period-t latent skills, as follows:

$$W_{p'}(\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}}) = \frac{E[R(p) = 1, R(p') = 0 | \boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}}] \times P(\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}})}{E[R(p) = 1, R(p') = 0]} \tag{19}$$

where $W_{p'}(\boldsymbol{\theta_t})$ measures the share of compliers across different deciles of the joint skill distribution at period t.[27]

Empirical Evidence. A policy change which limits parents' capacity to request their children to be retained would reduce the prevalence of early retention by 0.42 percentage points — corresponding to a 7.8% drop in baseline retention rates in the ECLS sample. Relative to the full sample, policy compliers are less likely to be female, older and from two parent households (Table B.8). Moreover, we find sizable differences in the latent skill dimension, as compliers have lower initial cognitive and non-cognitive skills, trailing the full sample average by 0.40 and 0.20 standard deviations, respectively. In Figure B.7, we present evidence on the complier weights introduced in equation (19). Policy compliers are largely drawn from the bottom of the period t latent skills distribution. For instance, 13% of compliers come from the bottom decile of the $\theta_{C,t}$ distribution, compared to just 7.9% of their peers in the top decile. On the other hand, we find far smaller differences in the non-cognitive dimension.

The third row of Table 5 presents the estimated average PRTE, which directly follows from complier weights presented in Figure B.7 and the heterogeneous PRTEs across the skill distribution. The policy change would thus worsen students' period $t + 1$ cognitive skills by 0.123 standard deviations, while slightly improving their non-cognitive abilities by 0.035 standard deviations, though neither estimate is statistically significant at the 10% level.

8 Compensating Income Transfer

While the existing literature has yet to reach a consensus on the net impacts of retention, holding children back entails an additional year in public schooling.[28] Average per-pupil expenditures, which amounted to $11,841 in 2015-16 (Cornman et al., 2018), imply that the direct costs of retention are significant. Our empirical framework incorporates the impor-

[27] $P(\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}})$ equals the share of students in joint skill decile $\underline{\boldsymbol{\theta_t}}$ and $E[R(p) = 1, R(p') = 0 | \boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}}]$ captures the share of students in that decile who are policy compliers.

[28] Various papers have considered the impact of retention on subsequent schooling attainment, finding mixed results (Manacorda, 2012; Gary-Bobo et al., 2016; Eren et al., 2017). Moreover, these papers follow RD designs, whose results do not necessarily generalize outside the cutoff. In this context, we assume that early retention would lead students to spend an additional year in public schooling, though our analysis could be extended to incorporate alternative retention-cost estimates.

tance of parental income on their cognitive and non-cognitive investment choices (equation (3)). As such, we consider an alternative policy regime in which parents would receive a compensating transfer equivalent to the costs of retention. In particular, we assume the income transfer would be entirely dedicated to children's skill investments, whose impacts would be evenly split across the two skill dimensions. We thus compare outcomes for on-time students whose parents received the compensating income transfer compared to having retained them. The estimated effect of this policy counterfactual is given by:

$$ATE_S^Y \equiv \iint E[g^0_{S,t+1}(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t(\Delta Y)) - g^1_{S,t+1}(\cdot)]dF_{\theta_C,\theta_{NC}}(\underline{\theta}_{C,t}, \underline{\theta}_{NC,t}) \qquad (20)$$

where ΔY represents the magnitude of the compensating income transfer, which follows directly from annual per-pupil expenditure costs. The estimated impact of the compensating income policy (ATE_S^Y) depends on the estimated impacts of retention (Figures 4-5) and the productivity of investments in the skill formation process (Table 4). In our setting, the compensating income transfer would increase parents' latent investments in period t by 0.22 standard deviations.

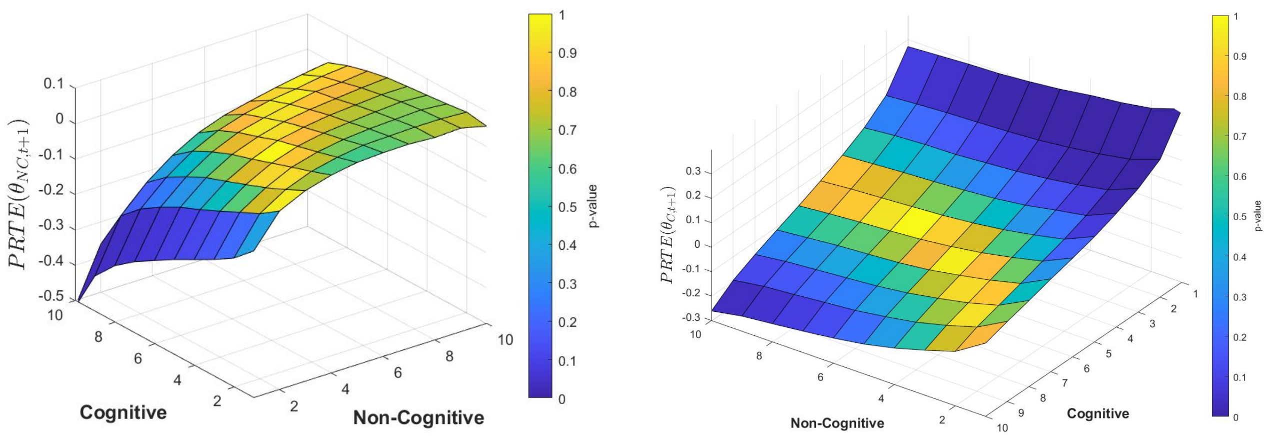
We present the estimated effects of the compensating income transfer in Figure 6. This policy change would modestly worsen children's non-cognitive skill outcomes at endline (Panel A), with average impacts equaling 0.029 standard deviations. The estimated effects of the income transfer in lieu of retention are positive only for high-skilled children at time t, reaching 0.05 σ for those in the top decile, yet not statistically different from zero. For their low-skilled peers, the policy reform would result in negative impacts in this dimension, given the positive impacts of retention documented in Figure 5.

On the other hand, the compensating income transfer would result in small positive impacts on children's cognitive skills, yielding an average improvement of 0.023 σ In this dimension, we find that the policy change would yield larger $\theta_{C,t+1}$ increases for low-skilled children, as the estimated ATE_C^Y reaches 0.5 standard deviations for children in the bottom joint decile of the skill distribution. As a result, since low-skilled children are more likely to have repeated a grade, the policy change for children who are retained (TT_{NC}^Y) would increase their cognitive skills by 0.152 standard deviations through endline.

The existing literature on dynamic skill formation models has previously examined the impacts of early-life income transfers on children's skill development (Cunha et al., 2010; Agostinelli and Wiswall, 2016a; Attanasio et al., 2020). Our analysis extends this strand of the literature by allowing us to perform a cost-benefit analysis of income transfers relative to a costly policy, such as grade retention. Our results indicate that replacing repetition events

Figure 6: Estimated Impacts of Compensating Income Transfer on $\boldsymbol{\theta_{t+1}}$

(a) Effects on $\theta_{NC,t+1}$ **(b)** Effects on $\theta_{C,t+1}$



Note: Figure 6 presents the estimated impacts of compensating parents with the estimated cost of retention vis-à-vis holding their children back. We present the estimated impacts on non-cognitive (Panel A) and cognitive skills (Panel B) in period $t+1$ across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

with direct income transfers could yield small improvements in children's skill development, depending on the outcome in consideration and their initial skill level.

9 Conclusion

This paper introduces a model of dynamic skill formation which incorporates parental skills, their investments and endogenous retention outcomes. We allow for grade retention to depend on children's latent skills, their background characteristics and school-level policies. As such, we incorporate retention-dependent production functions, which allow us to estimate how this practice affects children's cognitive and non-cognitive skills formation. We thus extend to the existing literature on the impacts of grade repetition on academic outcomes by distinguishing test scores from latent abilities and analyzing the impacts of retention on non-cognitive skills, while incorporating the importance of parents' skills and their investments in the skill formation process.

Our analysis shows that children with low cognitive and non-cognitive skills are far more likely to be retained. School-level policies play an important role in the retention process, as well. While retention has limited average impacts on children's latent skill outcomes, we find significant heterogeneity across the initial skill distribution. As a result, retention harms low-skilled students' cognitive outcomes, while yielding improvements in the non-cognitive dimension. We take further advantage of the estimated model parameters to examine how policy changes can affect children's skill development outcomes. Changing school policies to reduce the prevalence of retention could improve children's skill outcomes. Similarly, investment-targeted income transfers in lieu of retention positively impact low-skilled students' cognitive development. However, since the impacts of these policy changes are heterogeneous by children's initial skills, policy reforms should incorporate these considerations, as one size will not fit all.

References

Agostinelli, F. and M. Wiswall (2016a). Estimating the technology of children's skill formation. Technical report, National Bureau of Economic Research.

Agostinelli, F. and M. Wiswall (2016b). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. Technical report, National Bureau of Economic Research.

Alet, E., L. Bonnal, and P. Favard (2013). Repetition: Medicine for a short-run remission. *Annals of Economics and Statistics*, 227–250.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., C. Meghir, and E. Nix (2020). Human capital development and parental investment in india. *The Review of Economic Studies 87*(6), 2511–2541.

Bertrand, M. and J. Pan (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics 5*(1), 32–64.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). 2001 Lawrence R. Klein Lecture 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.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2010, jan). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica 78*(1), 377–394.

Cockx, B., M. Picchio, and S. Baert (2019). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics 34*(3), 403–424.

Cornman, S. Q., L. Zhou, M. R. Howell, and J. Young (2018). Revenues and expenditures for public elementary and secondary education: School year 2015-16 (fiscal year 2016). first look. nces 2019-301. *National Center for Education Statistics*.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *Review of Economic Studies 81*(1), 137–185.

Del Bono, E., J. Kinsler, and R. Pavan (2020). Skill formation and the trouble with child non-cognitive skill measures. Technical report, IZA Discussion Papers.

d'Haultfoeuille, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics 154*(1), 1–15.

Diaz, J., N. Grau, T. Reyes, and J. Rivera (2016). The Impact of Grade Retention on Juvenile Crime. Technical report, University of Chile, Department of Economics.

Doepke, M., G. Sorrenti, and F. Zilibotti (2019). The economics of parenting. *Annual Review of Economics 11*, 55–84.

Doepke, M. and F. Zilibotti (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica 85*(5), 1331–1371.

Eide, E. R. and M. H. Showalter (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review 20*(6), 563–576.

Elder, T. and Y. Zhou (2021). The black-white gap in noncognitive skills among elementary school children. *American Economic Journal: Applied Economics 13*(1), 105–32.

Eren, O., B. Depew, and S. Barnes (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics 153*, 9–31.

Eren, O., M. F. Lovenheim, and N. H. Mocan (2018). The Effect of Grade Retention on Adult Crime: Evidence from a Test-Based Promotion Policy. Technical Report 0898-2937, National Bureau of Economic Research.

Freyberger, J. (2018). Non-parametric panel data models with interactive fixed effects. *The Review of Economic Studies 85*(3), 1824–1851.

Fruehwirth, J. C., S. Navarro, and Y. Takahashi (2016). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics 34*(4), 979–1021.

Gary-Bobo, R. J., M. Goussé, and J. Robin (2016). Grade retention and unobserved heterogeneity. *Quantitative Economics 7*(3), 781–820.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016, April). Dynamic treatment effects. *Journal of Econometrics 191*(2), 276–292.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*(S1), S197–S246.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *American Economic Review 91*(2), 107–111.

Ikeda, M. and E. Garcia (2014, February). Grade repetition. *OECD Journal: Economic Studies 2013*(1), 269–315.

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

Jacob, B. A. and L. Lefgren (2009, July). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics 1*(3), 33–58.

Judd, K. L. (1998). *Numerical methods in economics*. MIT press.

Koppensteiner, M. F. (2014). Automatic grade promotion and student performance: Evidence from Brazil. *Journal of Development Economics 107*, 277–290.

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics 94*(2), 596–606.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica 86*(5), 1589–1619.

Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics 92*(3), 562–576.

Ozek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. *Education Finance and Policy 10*(3), 350–377.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association 53*(284), 873–880.

Sarzosa, M. (2015). The dynamic consequences of bullying on skill accumulation.

Schwerdt, G., M. R. West, and M. A. Winters (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics 152*, 154–169.

Williams, B. (2019). Identification of the linear factor model. *Econometric Reviews*, 1–18.

Appendix

A Reduced Form Results

Table A.1: Teachers' Ratings of Student Skills, Test Scores and Retention

	Math			Reading			Science		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Test Score	0.583***	0.579***	0.683***	0.647***	0.621***	0.723***	0.342***	0.303***	0.411***
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)
Repeated KG	-0.062	-0.071	-0.083*	-0.051	-0.069*	-0.077*	0.012	-0.033	-0.035
	(0.046)	(0.047)	(0.046)	(0.038)	(0.040)	(0.041)	(0.049)	(0.052)	(0.051)
Background Characteristics		✓			✓			✓	
School Fixed Effects			✓			✓			✓
R^2	0.335	0.338	0.467	0.420	0.427	0.522	0.116	0.132	0.271
Observations	10,961	10,961	10,961	10,961	10,961	10,961	10,961	10,961	10,961

Source: ECLS-K:2011. Note: Table A.1 presents evidence of the relationship between teacher-reported Academic Rating Scales (ARS) of students in the Spring 2011 survey round in mathematics, reading and science vis-à-vis students' performance in these subjects in the corresponding survey round. Teachers' ratings constitute the outcome variables of interest. 'Repeated KG' captures students who had taken kindergarten at least once prior to the 2010-11 academic year. Background characteristics include gender, age, race, parental education and household structure as controls.. Standard errors clustered at the school level in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Determinants of Grade Retention

	(1)
Math Test Score	-0.030***
	(0.005)
Science Test Score	-0.002
	(0.003)
Reading Test Score	-0.012***
	(0.004)
Attentional Focus	-0.022***
	(0.004)
Inhibitory Control	0.002
	(0.005)
Composite Social Rating Scales	0.002
	(0.003)
Male	0.016***
	(0.005)
Age (R1)	-0.060***
	(0.008)
Underrepresented Minority	-0.022***
	(0.006)
Both Parents	-0.011*
	(0.006)
Parents' Education	0.002
	(0.001)
School-Level Policies	Yes
Observations	9010
R^2	0.068

Source: ECLS-K:2011. Note: Table A.2 presents estimated results from an OLS regression where grade retention is the outcome variable and the set of explanatory variables include baseline test scores, non-cognitive skill measures, individual, family background characteristics along with school-level retention policies. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Model Estimates

Table B.1: Empirical Specification

	Test Scores (1)	Non-Cog (2)	Parents (3)	Investment (4)	Retention (5)	CES: $\theta_{C,t+1}$ (6)	CES: $\theta_{NC,t+1}$ (7)
Observables							
Gender	✓	✓	✓	✓	✓		
Age	✓	✓	✓	✓	✓		
Race	✓	✓	✓	✓	✓		
Both Parents	✓	✓	✓	✓	✓		
Parents' Education	✓	✓	✓	✓	✓		
Parents' Age			✓	✓	✓		
Teachers' Education	✓	✓			✓		
Instruments							
HH Income				✓			
Retention Policies					✓		
Latent Factors							
$\theta_{C,t}$	✓	✓	✓	✓	✓	✓	✓
$\theta_{NC,t}$	✓	✓	✓	✓	✓	✓	✓
θ_P			✓	✓	✓	✓	✓
$\mathcal{I}_t$				✓		✓	✓

Source: ECLS-K:2011. Note: Table B.1 shows the variables used in the empirical model presented in Section 3.

Table B.2: Goodness of Fit of the Model

	Retained	Attention Focus	Inhibitory Control	SSRS	Math	Science	Reading
Means							
Actual	0.0577	0.0014	0.0015	0.0003	0.0008	0.0016	0.0013
Predicted	0.0537	0.0032	0.0037	0.0027	-0.0006	0.0010	0.0003
Std. Devs.							
Actual	0.2332	1.0002	0.9996	1.0012	0.9996	1.0002	0.9994
Predicted	0.2254	1.0988	0.9776	1.0889	1.1073	1.0872	1.1125

Note: Table B.2 presents predicted and observed means and standard deviations of baseline test scores, non-cognitive skill measures along with the prevalence of retention. Predicted values come from simulations based on the estimated parameters of the model. Predicted means and standard deviations are not statistically different from the actual means and standard deviations at any conventional level of significance.

Table B.3: Estimated Parameters of Measurement System at t

	Attentional Focus	Inhibtory Control	SRS	Math	Science	Reading
Male	-0.309***	-0.322***	-0.277***	0.058***	0.049***	-0.085***
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)
Age (R1)	0.027***	0.020***	0.011***	0.062***	0.040***	0.048***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
URM Child	-0.007	0.003	0.004	-0.305***	-0.527***	-0.143***
	(0.019)	(0.018)	(0.020)	(0.020)	(0.020)	(0.020)
Both Parents	0.236***	0.225***	0.228***	0.215***	0.077***	0.215***
	(0.021)	(0.020)	(0.022)	(0.021)	(0.021)	(0.021)
Parents' Education	0.035***	0.017***	0.015***	0.123***	0.122***	0.129***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Teacher's Education'	0.006	-0.002	0.018**	0.007	0.006	0.015*
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)
Non-Cognitive Skills	1.152***	1.304***	1			
	(0.016)	(0.017)	■			
Cognitive Skills				1.072***	0.575***	1
				(0.016)	(0.013)	■
Observations			9010			

Note: Table B.3 presents the estimated parameters from the measurement system at period t. "Lives Both Parents" is a binary variable capturing if students reside with both biological parents. "Parental Edu" refers to the educational attainment of the primary parent, measured as the total years of schooling. "SES" is an index of socioeconomic status based on parental education, occupation and income. Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.4: Estimated Parameters of Measurement System at $t + 1$

	Attentional Focus	Inhibtory Control	SRS	Math	Science	Reading
Male	-0.376***	-0.462***	-0.300***	0.215***	0.144***	-0.156***
	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)	(0.019)
Age (R1)	0.018***	0.014***	0.009***	0.027***	0.029***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
URM Child	0.040**	0.025	0.011	-0.387***	-0.405***	-0.208***
	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)	(0.021)
Both Parents	0.310***	0.320***	0.369***	0.268***	0.161***	0.216***
	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)
Parents' Education	0.035***	0.023***	0.023***	0.093***	0.109***	0.109***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Teacher's Education'	0.013	0.012	0.029***	-0.002	0.008	0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Non-Cognitive Skills	1.044***	1.105***	1			
	(0.013)	(0.013)	■			
Cognitive Skills				0.980***	0.879***	1
				(0.015)	(0.014)	■
Observations			9010			

Note: Table B.4 presents the estimated parameters from the measurement system at period $t + 1$. "Lives Both Parents" is a binary variable capturing if students reside with both biological parents. "Parental Edu" refers to the educational attainment of the primary parent, measured as the total years of schooling. "SES" is an index of socioeconomic status based on parental education, occupation and income. Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.5: Estimated Parameters of Parental Skill Measurement System

	Anxiety I	Anxiety II	Sadness
Both Parents	0.163***	0.092***	0.035***
	(0.031)	(0.027)	(0.012)
Parents' Education	-0.015**	0.022***	0.003
	(0.006)	(0.005)	(0.002)
Male Parent	0.220***	0.127***	0.062***
	(0.043)	(0.038)	(0.016)
Parent's Age	0.012***	0.003	-0.002**
	(0.002)	(0.002)	(0.001)
URM Parent	0.286***	-0.069***	0.032***
	(0.030)	(0.027)	(0.012)
Parental NC Skills	0.517***	0.539***	1
	(0.009)	(0.008)	■
Observations		8088	

Note: Table B.5 presents the estimated parameters from the measurement system of parental socioemotional skills. The first column presents evidence from the parental sadness measure. The following two columns show results for two measures of parental anxiety. All measures are reversed. "Both Parents" is a binary variable capturing if students reside with both biological parents. "Parents' Education" refers to the educational attainment of the primary parent, measured as the total years of schooling. "SES Index" is an index of socioeconomic status based on parental education, occupation and income. Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.6: Estimated Parameters of Measurement System of Parental Investments

	Number of Books	School Engagement	NC Investment	Activities
Latent Parental Investment	0.231***	0.384***	0.288***	1
	(0.017)	(0.034)	■	
Observations		9010		

Note: Table B.6 presents the estimated parameters from the measurement system of parental investments. The first column presents evidence from . Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.7: Determinants of Grade Retention (Marginal Effects at the Mean)

	(1)
Male	0.014***
	(0.004)
Age (R1)	-0.006***
	(0.000)
Underrepresented Minority	-0.001
	(0.004)
Both Parents	-0.020***
	(0.004)
Parents' Education	-0.004***
	(0.001)
Teachers' Education'	0.001
	(0.002)
Retained for Immaturity	0.003
	(0.005)
Retained at Parents' Request	0.010*
	(0.005)
Retained for Academic Deficiency	-0.004
	(0.007)
Retained if Failed Test	0.003
	(0.007)
Retained More than Once	0.013*
	(0.007)
Retained Without Parents' Consent	0.009*
	(0.004)
% Free and Reduced Price Lunch	-0.000*
	(0.000)
Public School	-0.018**
	(0.006)
Urban School	-0.012**
	(0.004)
Parents' NC Skills	0.000
	(0.001)
Non-Cognitive Skills	-0.011***
	(0.003)
Cognitive Skills	-0.040***
	(0.003)
Observations	9,010

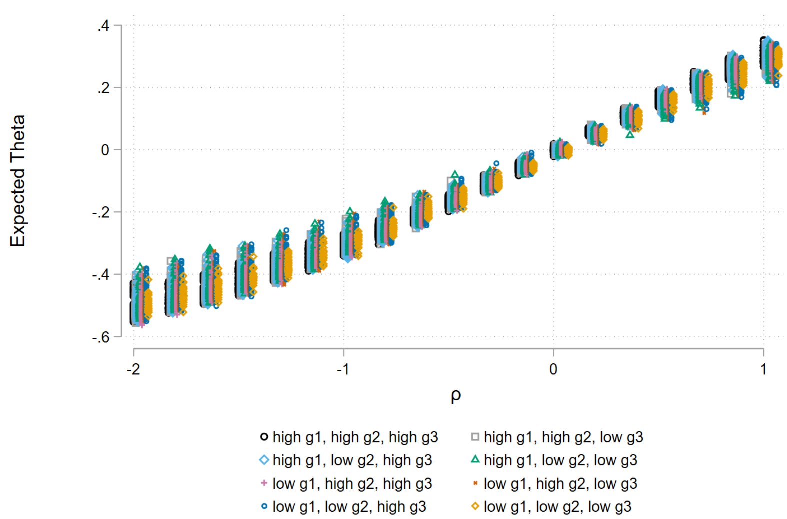
Note: Table B.7 presents estimated marginal effects from the retention equation (2). Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.8: Characteristics of Policy Compliers for Parental Request Retention Policy

	Full Sample	PRTE Sample	Difference
Observed Characteristics			
Male	0.4943	0.5390	0.0446
	(0.5000)	(0.4980)	
Age in Months (R1)	67.3184	65.7506	-1.5679***
	(4.0817)	(3.6504)	
URM Child	0.4063	0.4295	0.0232
	(0.4912)	(0.4947)	
Both Parents	0.7207	0.5939	-0.1269
	(0.4480)	(0.4897)	
Parental Education	14.0189	13.5765	-0.4424
	(2.4779)	(2.3928)	
Teacher's Education	16.9132	16.9321	0.0189
	(1.0979)	(1.0814)	
Parents' NC Skills	0.0018	-0.0127	-0.0144
	(1.3559)	(1.4019)	
Unobserved Characteristics			
Non-Cognitive Skills	-0.0001	-0.2040	-0.2039
	(0.6767)	(0.6925)	
Cognitive Skills	0.0005	-0.4030	-0.4035***
	(0.7178)	(0.6695)	

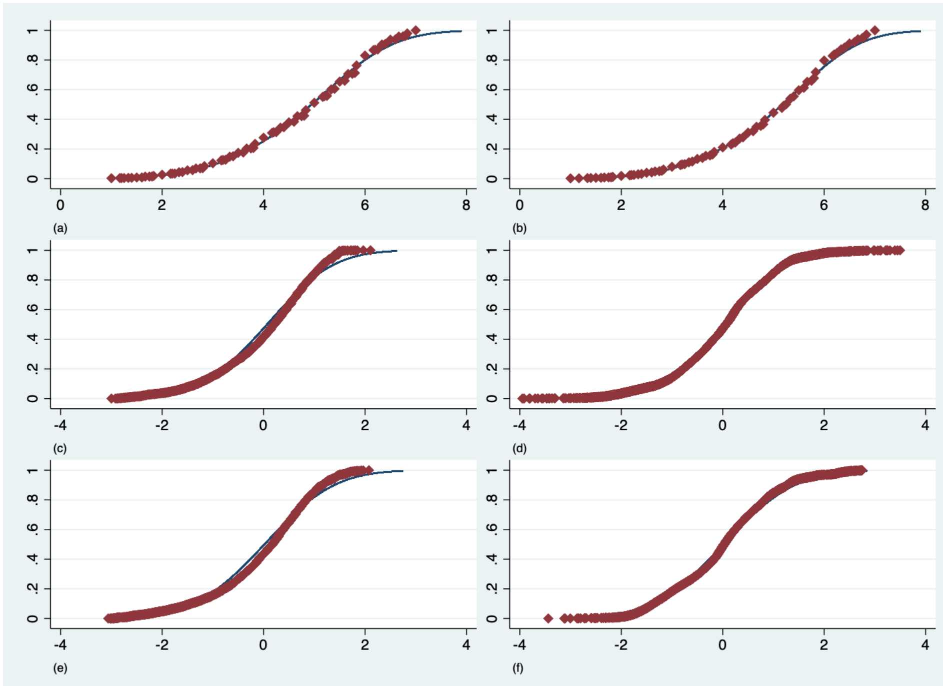
Note: Table B.8 presents observed and unobserved characteristics of policy compliers for a change in the school-level policy allowing parents to unilaterally request for their children to be retained.

Figure B.1: Relationship between the mean of $\widehat{\boldsymbol{\theta_{t+1}}}$ and ρ



Note: Figure B.1 presents the results of 1,400 different combinations of γ_1, γ_2, γ_3 and ρ parameters in the CES production function (equation (2): $\theta_{t+1} = [\gamma_1 x^\rho + \gamma_2 y^\rho + \gamma_3 z^\rho + (1 - \gamma_1 - \gamma_2 - \gamma_3) n^\rho]^(1/\rho))$ where x, y, z and n come from 5,000 random draws from independent normal distributions.

Figure B.2: Actual vs. predicted test scores cumulative distributions at $t = 1$

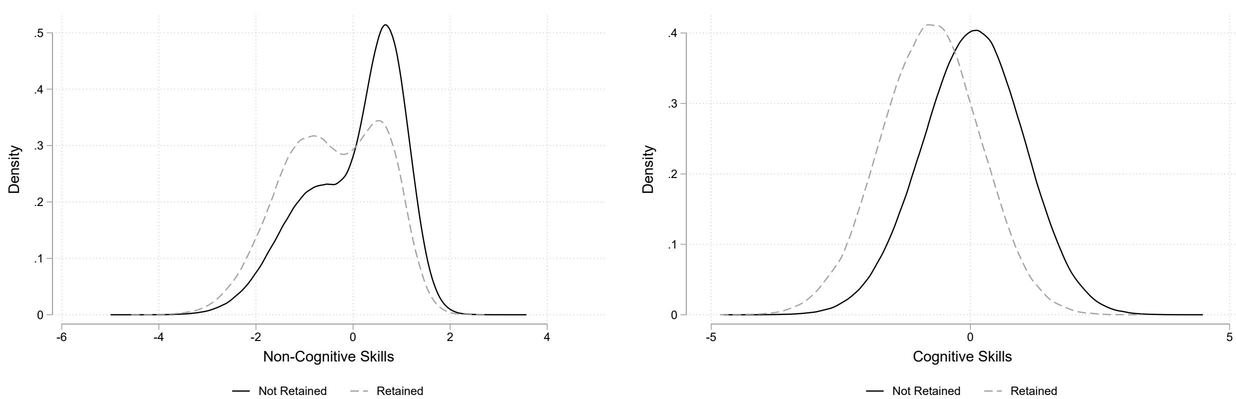


Note: Figure B.2 presents the predicted and observed distribution of baseline test scores and non-cognitive skill measures. Predicted values come from simulations based on the estimated parameters of the model. Actual (diamond) and predicted (line) cumulative distributions plotted of the following test scores: (a) attention focus (b) inhibitory control (c) SSRS (d) math (e) science (f) reading.

Figure B.3: Sorting into Grade Retention by Skill Levels

(a) Non-cognitive Skills

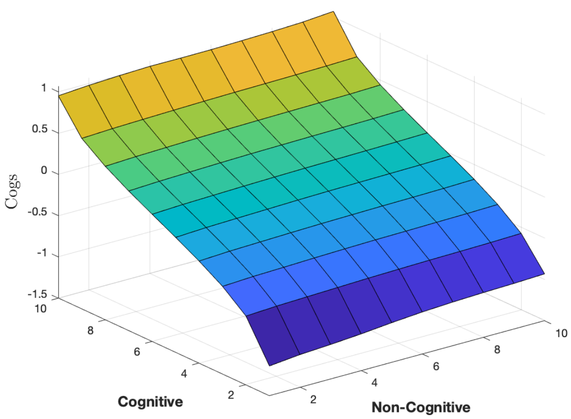
(b) Cognitive Skills



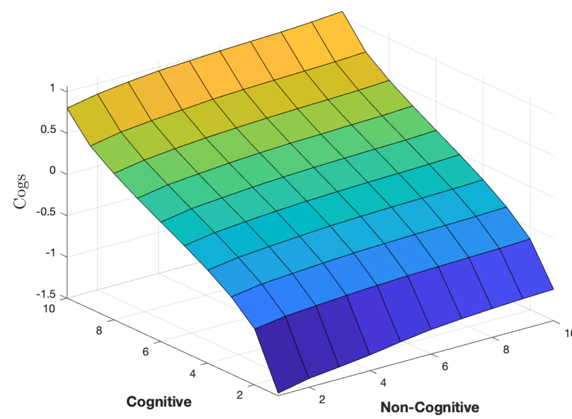
Note: Figure B.3 presents the marginal densities of latent non-cognitive and cognitive skills in period t by retention status.

Figure B.4: Latent Skills at $t+1$ as a function of $\boldsymbol{\theta_t}$

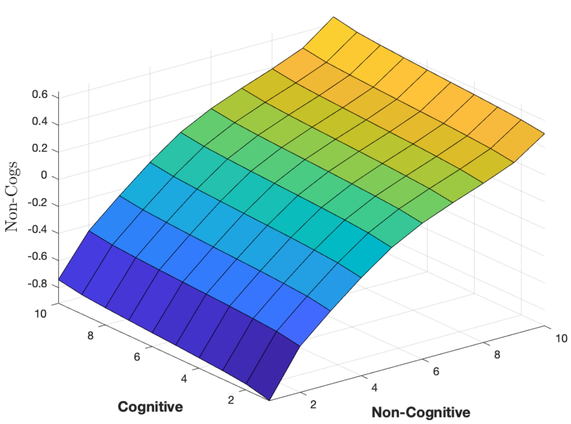
(a) $\theta_{C,t+1}$ as a function of $\boldsymbol{\theta_t}$: Non-Retained



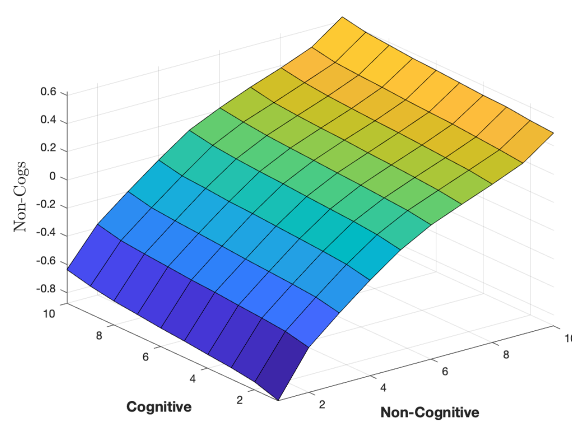
(b) $\theta_{C,t+1}$ as a function of $\boldsymbol{\theta_t}$: Retained



(c) $\theta_{NC,t+1}$ as a function of $\boldsymbol{\theta_t}$: Non-Retained



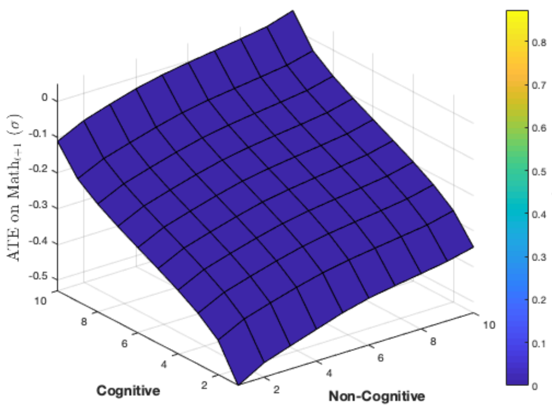
(d) $\theta_{NC,t+1}$ as a function of $\boldsymbol{\theta_t}$: Retained



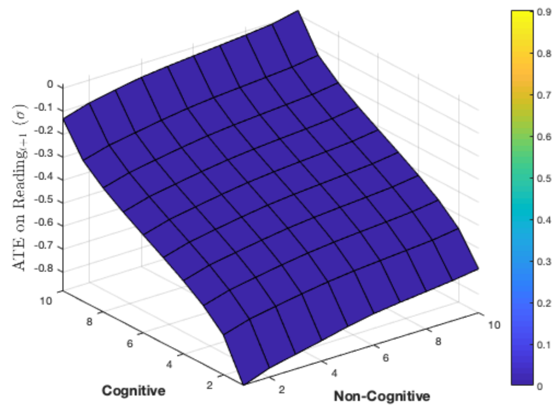
Note: The first two panels of Figure B.4 present the relationship between latent skills at t and cognitive skills at $t+1$ for non-retained and retained students, respectively. The last two panels present corresponding evidence for the relationship between $\theta_{C,t}$, $\theta_{NC,t}$ and $\theta_{NC,t+1}$ for non-retained and retained students, respectively.

Figure B.5: Estimated Effects of Retention on Test Scores at $t + 1$

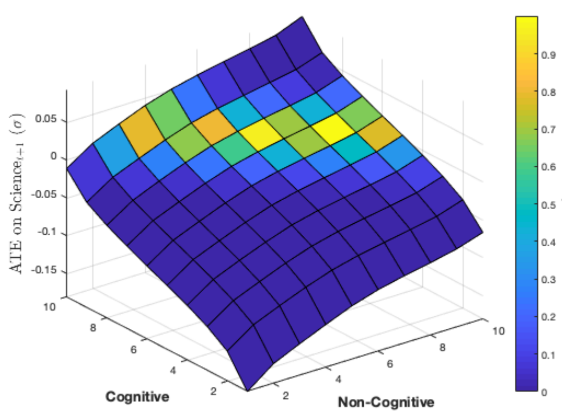
(a) Math Test Score



(b) Reading Test Score



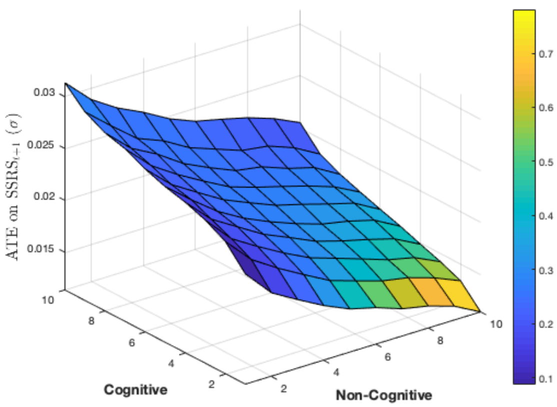
(c) Science Test Score



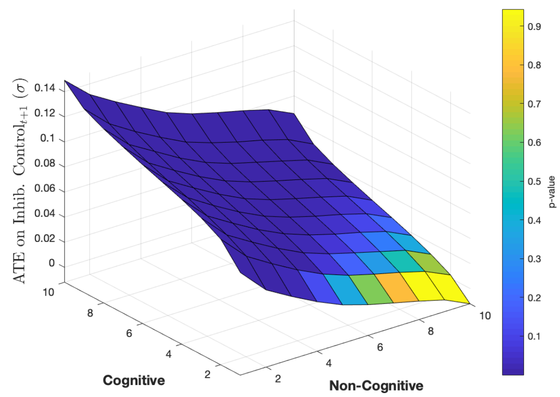
Note: Figure B.5 presents the estimated impacts of grade retention on endline test scores following a Roy model of retention using observed test scores as outcomes.

Figure B.6: Estimated Effects of Retention on Non-Cognitive Skills at $t + 1$

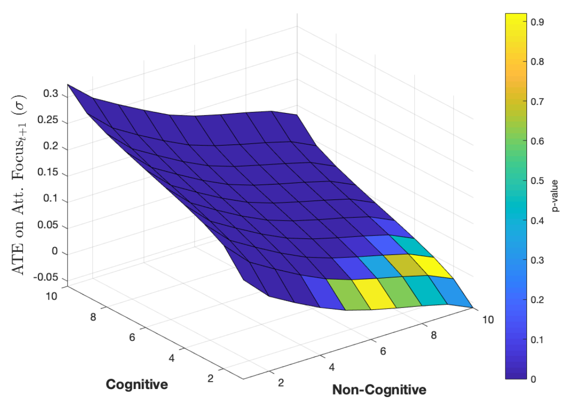
(a) Social Rating Scale



(b) Inhibitory Control

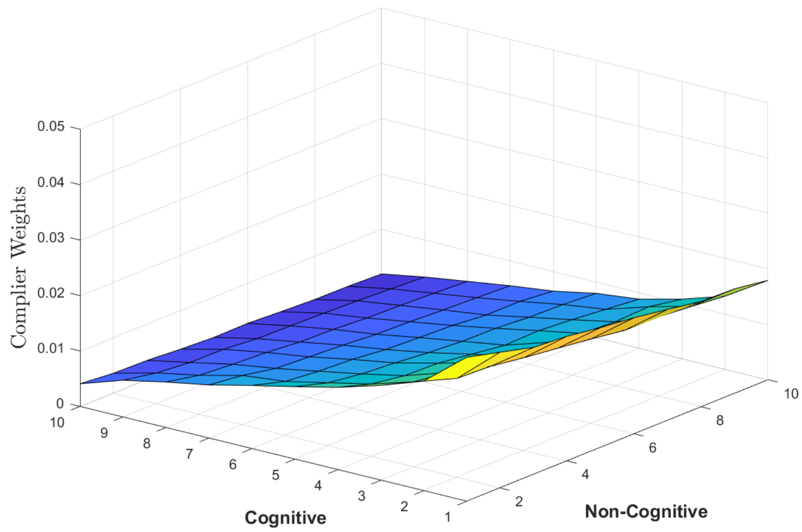


(c) Attentional Focusing



Note: Figure B.6 presents the estimated impacts of grade retention on endline non-cognitive skill measures following a Roy model of retention using observed non-cognitive skills as outcomes.

Figure B.7: PRTE: Complier Weights



Note: Figure B.7 presents estimated complier weights across the latent skill distribution in period t (equation (19)). The simulated policy change limits schools' ability to retain students solely at their parents' request.

C Unconditional Production of Latent Skills at $t + 1$

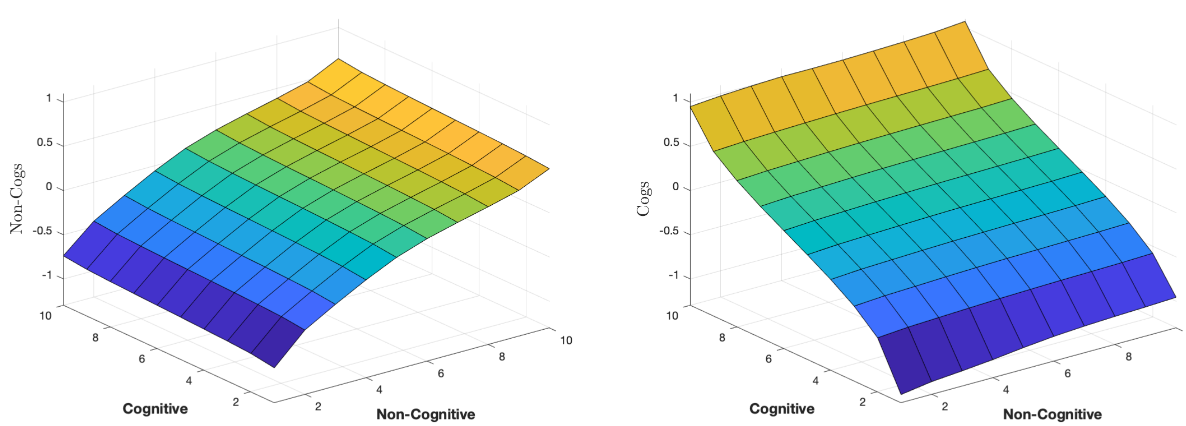
Our model differs from the existing skill development literature by accounting for endogenous retention outcomes, yet we can still examine the unconditional production function of skills relying on the Quandt (1958) switching regression framework. Latent skill outcomes at period $t + 1$ are thus given by:

$$\theta_{S,t+1}(\cdot) = R(\cdot)g^1_{S,t+1}(\theta_t, \theta_P, \mathcal{I}_t) + (1 - R(\cdot))g^0_{S,t+1}(\theta_t, \theta_P, \mathcal{I}_t)$$

We take advantage of this framework to examine the productivity of period t inputs on unconditional skill outcomes at $t+1$ (Figure C.1). We find strong evidence of self-productivity of skills ($\frac{\partial \theta_{S,t+1}}{\partial \theta_{S,t}}$), as a one σ increase in period t cognitive skills leads to 0.684 σ increase in $\theta_{C,t+1}$, with a corresponding impact of 0.605 σ in the non-cognitive dimension. Skills further exhibit cross-productivity ($\frac{\partial \theta_{S,t+1}}{\partial \theta_{-S,t}}$): a one σ increase in cognitive and non-cognitive skills in period t yields a 0.239 and 0.248 σ increase in the opposite skill dimension in $t + 1$. While Figure C.1 does not include parental investments, we find that a one σ increase in parents $\mathcal{I}_t$ is associated with increased cognitive and non-cognitive skills at $t + 1$ by 0.21 and 0.32 standard deviations, respectively.

Figure C.1: Latent Skills at $t + 1$ as a function of $\boldsymbol{\theta_t}$

(a) $\theta_{C,t+1}$ as a function of $\boldsymbol{\theta_t}$ **(b)** $\theta_{NC,t+1}$ as a function of $\boldsymbol{\theta_t}$



Note: The first panel of Figure C.1 presents the relationship between latent skills at t and cognitive skills at $t + 1$. The second panel presents corresponding evidence for the relationship between $\theta_{C,t}$, $\theta_{NC,t}$ and $\theta_{NC,t+1}$.