Grade Retention and Multidimensional Skill Formation in Young Children[*]

Fernando Saltiel[†] and Miguel Sarzosa[‡]

July 31, 2020

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, we find that 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. Parents engage in compensating investment for their retained children. Targeted retention policy changes could improve children's skill development.

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

JEL Codes: I20, I24

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., 2019) 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, Solis (2017) and 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' non-cognitive skills along with time and monetary investments in their children. We consider retention events in kindergarten and first grade—6.1% 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 skills and higher skilled parents exhibit increased investment in their children's non-cognitive skills, with no discernible impacts in the investment in cognitive skills. 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: 16.5% 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. I addition, parental investment is a greater driver of cognitive skill accumulation among retained children than among non-retained. Furthermore, we find that parental skills affect their children's non-cognitive skill outcomes only through increased investments.

We first estimate the impacts of retention on children's cognitive skill development. Early retention lowers students' cognitive skills at endline by 0.02 standard deviations (σ), with larger estimated impacts for students who were in fact retained (TT = 0.082 σ). We examine heterogeneous impacts across the initial skill distribution, finding large negative impacts for

low-skilled students, exceeding -0.2 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 0.02 σ. 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). We additionally show that parents engage in compensating investments in light of their children being held back. This channel partly accounts for the estimated impacts on children's skill outcomes.

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 solely at their parents' request. 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 limited 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., 2019), 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., 2019), 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), Solis (2017) 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 the estimated model results. Section 5 shows the estimated impacts of grade retention on children's cognitive and non-cognitive skill development, along with evidence on parents' compensating investments. Section 7 presents evidence on the impacts of changes in school-level retention policies on children's cognitive and non-cognitive skills. Lastly, Section 8 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.

Critical to the analysis of grade retention, 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. We observe students' race, gender and their age in each survey round, as well as information on their household composition, parents' education, family income and poverty status. We also use a composite measure of socioeconomic status (SES), which is created by the NCES based on parental education, occupation and household income. Lastly, we consider baseline measures of parents' reported psychological well-being, encompassing whether they felt sad, could not get going, and felt depressed, among other questions.

We additionally take advantage of various measures of parents' behavior to examine the importance of parental investment for their children's cognitive and non-cognitive skill development. In the baseline survey round, we consider the number of books available in the household, weekly time parents spend reading to their children, activities done with their children, the frequency with which they hit their kids, engagement in school activities along with their attitudes towards parenting and their relationship with their child.[7] Further-

---

[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 and practicing numbers. The school engagement measure captures whether parents attended back to

more, the endline survey round includes similar measures of parental behavior, allowing us to examine whether parents engage in compensating investment in their retained children. These measures represent a combination of parents' time and 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., 2019, 2020).

Furthermore, ECLS includes a school-level survey which captures relevant characteristics to the analysis of grade retention. In particular, we 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, 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 as well as their internalizing and externalizing problem behaviors.[10] We additionally 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.

---

school night, participated in the PTA or parent advisory groups, attended school events and/or participated in committees/volunteering/fundraising at the school. Attitudes towards parenting include whether being a parent is harder than expected, whether the child does things that bother the parent, whether the parent sacrifices for the child and whether they often feel angry with their kids. The relationship index captures 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, 2018).

A common challenge for papers on grade retention is whether to focus on the impact of this event holding the age or the grade 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 schooling. For socio-emotional 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 the 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 1,099 who had previously enrolled in kindergarten. We drop an additional 883 students who did not take the baseline assessments or provided information on their gender, age or race. 2,174 parents did not respond to the baseline investment measures. An additional 980 children have missing values for the teacher-reported non-cognitive skill measures at baseline. We drop an additional 1,171 children due to missing information on grade progression. Furthermore, 1,006 students have missing information on their test scores at endline. Additionally, 1,116 students have missing parental investment measures at endline, with an additional 538 students having missing values for endline non-cognitive skill measures. As a result, the final sample size includes 6,817 students, of whom 419 repeated either kindergarten or first grade.

2.2   Regression Analysis

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, three-fourths of children live in two-parent households and about 20% of the children in the sample reside in households below the poverty line. Retained students are less likely to reside with both parents, and tend to come from households with less educated and lower income parents. Across parental investments, we only find significant differences on whether parents sang songs with their children every day. Our empirical strategy additionally considers variation in school-level retention policies for kindergartners. Similar to Fruehwirth et al. (2016)'s analysis for the 1998-99 ECLS cohort, we document significant 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. While retained students attend schools with more 'stringent' retention policies, these differences are only significant in the

immaturity and parental request policies.

Importantly, there are sizable differences in students' academic achievement at baseline, as retained students trail their non-retained peers by 0.82, 0.71 and 0.48 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 0.92, 1.13 and 0.61 standard deviations, respectively. At the same time, we find significant differences across the three non-cognitive skill measures, in the range of 0.26-0.32 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.1. 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 the first panel of 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 0.39 standard deviations. The estimated coefficients on the reading and science test scores are negative as well, reaching 0.62 and 0.12 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' attentional focus by 0.12 σ, denoting the extent to which selection-on-observables affects the estimated impacts of this practice.

Nonetheless, the essential identification challenge in this context remains sorting-into-retention based on unobserved characteristics. A possible solution to this issue is to estimate an instrumental variables strategy, leveraging school-level policies as exclusions. We present the results in the second panel of Table 2. Unlike the OLS results, we find that retention is

Table 1: Summary Statistics by Retention Status

	Full Sample (1)	Not Retained (2)	Retained (3)
Observed Characteristics			
Male	0.501	0.495	0.599***
Age (R1)	5.618	5.630	5.435***
White	0.573	0.574	0.563
Black	0.100	0.098	0.138**
Hispanic	0.210	0.212	0.177
Both Parents	0.756	0.763	0.656***
Parents' Education	14.295	14.322	13.878***
Log HH Income	10.901	10.912	10.735***
SES Index	0.091	0.100	-0.052***
Poverty Status	0.202	0.195	0.308***
Parental Investment (R1)			
Tell Stories	0.409	0.408	0.413
Sing Songs	0.428	0.425	0.475*
Help Doing Art	0.221	0.220	0.236
Play Games	0.231	0.230	0.239
Talk About Nature	0.112	0.112	0.115
Build Things	0.163	0.162	0.186
Play Sports	0.237	0.235	0.267
Practice Numbers	0.605	0.604	0.618
Read Books	0.531	0.533	0.504
Number of Books	93.741	94.490	82.308
School Policies			
Retained for Immaturity	0.668	0.665	0.714*
Retained at Parents' Request	0.671	0.667	0.737**
Retained for Academic Deficiency	0.835	0.834	0.850
Retained if Failed Test	0.065	0.064	0.084
Retained More than Once	0.065	0.064	0.086
Retained Without Parents' Consent	0.272	0.269	0.308
Skill Measures (R1)			
Math Test Score	0.000	0.054	-0.827***
Reading Test Score	0.000	0.047	-0.713***
Science Test Score	-0.000	0.031	-0.479***
Composite Social Rating Scales	-0.000	0.022	-0.330***
Attentional Focus	-0.000	0.044	-0.669***
Inhibitory Control	0.000	0.034	-0.516***
Skill Measures (R6)			
Math Test Score	0.000	0.060	-0.920***
Reading Test Score	0.000	0.074	-1.129***
Science Test Score	0.000	0.040	-0.614***
Composite Social Rating Scales	0.000	0.017	-0.259***
Attentional Focus	-0.000	0.021	-0.317***
Inhibitory Control	0.000	0.019	-0.284***
Observations	6,817	6,398	419

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 2012-13 survey. 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

associated with a positive impact on math test scores, along with large negative impacts on students' Social Rating Scales and inhibitory control (albeit statistically insignificant).

However, Heckman and Vytlacil (2005); Heckman et al. (2006) have shown that instrumental variable approaches do not necessarily recover a parameter of interest, and the issue is exacerbated in models with multiple instruments (Mogstad et al., 2019). Moreover, 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. Furthermore, the results presented so far are not informative of 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

3.1   Model Structure

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, dedicated parent 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.

In this context, $\theta_{S,i,\tau}$ denotes student i's stock of skills $S = \{C, NC\}$ at time $\tau \in \{t, t+1\}$, $\theta_{P,i}$ reflects parents' time-invariant latent skills and $\mathcal{I}_{S,i,t}$ denotes parental investment in their children's skill development. R_i indicates whether she has been retained between t and $t+1$. As such, the model of skill formation which allows for endogenous retention and parental

Table 2: Estimated Impacts of Retention: OLS and IV Estimates

Panel A. OLS Estimates of Grade Retention

	Math Test Score	Reading Test Score	Science Test Score	SRS	Attentional Focus	Inhibitory Control
	(1)	(2)	(3)	(4)	(5)	(6)
Grade Retention	-0.393***	-0.626***	-0.120***	0.024	0.120**	0.071
	(0.033)	(0.035)	(0.033)	(0.046)	(0.044)	(0.044)
Math Test Score	0.558***	0.297***	0.237***	0.050**	0.169***	0.072***
	(0.013)	(0.014)	(0.014)	(0.019)	(0.018)	(0.018)
Reading Test Score	-0.024	0.218***	0.071***	-0.034	-0.016	-0.014
	(0.013)	(0.013)	(0.013)	(0.018)	(0.017)	(0.017)
Science Test Score	0.185***	0.202***	0.497***	0.015	0.037**	0.013
	(0.010)	(0.011)	(0.010)	(0.014)	(0.014)	(0.014)
SRS	0.006	-0.006	0.009	0.160***	0.054***	0.093***
	(0.010)	(0.010)	(0.010)	(0.013)	(0.013)	(0.013)
Attentional Focus	0.099***	0.065***	0.043***	0.029	0.156***	0.079***
	(0.013)	(0.014)	(0.013)	(0.018)	(0.017)	(0.017)
Inhibitory Control	-0.008	0.033*	-0.002	0.235***	0.218***	0.273***
	(0.013)	(0.014)	(0.013)	(0.018)	(0.018)	(0.018)
Background Characteristics			✓			
Parental Investment			✓			
Observations			6,817			
R^2	0.609	0.552	0.606	0.235	0.311	0.300

Panel B. IV Estimates of Grade Retention

	Math Test Score	Reading Test Score	Science Test Score	SRS	Attentional Focus	Inhibitory Control
	(1)	(2)	(3)	(4)	(5)	(6)
Grade Retention	0.579	-0.080	0.007	-1.331	0.103	-0.648
	(0.815)	(0.836)	(0.771)	(1.140)	(1.019)	(1.047)
Math Test Score	0.592***	0.316***	0.241***	0.002	0.168***	0.047
	(0.032)	(0.033)	(0.030)	(0.045)	(0.040)	(0.041)
Reading Test Score	-0.016	0.223***	0.073***	-0.046*	-0.017	-0.020
	(0.015)	(0.016)	(0.014)	(0.021)	(0.019)	(0.019)
Science Test Score	0.189***	0.205***	0.498***	0.008	0.037*	0.010
	(0.011)	(0.012)	(0.011)	(0.016)	(0.014)	(0.015)
SRS	0.002	-0.008	0.008	0.166***	0.054***	0.096***
	(0.011)	(0.011)	(0.010)	(0.015)	(0.013)	(0.014)
Attentional Focus	0.121***	0.077**	0.046*	-0.002	0.156***	0.063*
	(0.023)	(0.024)	(0.022)	(0.032)	(0.029)	(0.030)
Inhibitory Control	-0.009	0.032*	-0.002	0.236***	0.218***	0.274***
	(0.014)	(0.014)	(0.013)	(0.020)	(0.018)	(0.018)
Background Characteristics			✓			
Parental Investment			✓			
School-Level Policies			✓			
Observations			6,817			
R^2	0.558	0.536	0.605	0.136	0.311	0.273

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. Panel A presents evidence from an OLS regression with controls for family and children's background characteristics, baseline skill measures and parental investment measures. Panel B presents evidence from an instrumental variables specification which incorporates school-level policies (presented in Table 1) as instruments. Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

investment is given by:

$$\theta_{S,i,t+1} = (\gamma_{C,S,t}^R \theta_{C,i,t}^{\rho_S^R} + \gamma_{NC,S,t}^R \theta_{NC,i,t}^{\rho_S^R} + \gamma_{P,S,t}^R \theta_{P,i}^{\rho_S^R} + \gamma_{I,S,t}^R \mathcal{I}_{S,i,t}^{\rho_S^R})^{1/\rho_s^R} + \eta_{i,S,t}^R \tag{1}$$

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

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

where $\gamma_{C,S,t}^R + \gamma_{NC,S,t}^R + \gamma_{P,S,t}^R + \gamma_{I,S,t}^R = 1$, $\eta_{i,S,t}^{R_i}$ represents an i.i.d. shock which is independent to contemporaneous skills, across skill dimensions ($S = \{C, NC\}$), by time and across retention status. Importantly, we allow for the parameters of the CES production function to vary across students' retention status R, thus allowing us to examine how retention affects the skill development process. $\mathbb{1}$ is an indicator function which equals one if true, $X_{i,t}$ includes observed characteristics measured at baseline. $\varepsilon_{i,t} \sim N(0, \sigma_{\varepsilon_{i,t}})$ is an error term which is independent of observed characteristics, latent factors, and of $\eta_{i,S,t}^{R_i}$. 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 the likelihood of retention.

We follow Cunha et al. (2010); Agostinelli and Wiswall (2016a); Attanasio et al. (2019) and present a CES production function of latent skill in period $t + 1$, where $\sigma_s^R = \frac{1}{1-\rho_s^R}$ 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 $t + 1$. 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-Douglass production function, as it places fewer restrictions on the possible interactions between the its inputs. 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.

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), Del Boca et al. (2016), 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., 2019; 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. $\mathcal{I}_{S,i,t}$ denotes parents' investments in children's skill dimension S, $v_{S,I,i,t}$ is a mean-zero error term which is independent of observed characteristics,

latent skills and investment factors.[11] Since parental investment choices may be endogenous to shocks in the skill development process ($I_{S,i,t} \not\perp \eta_{i,S,t}^{R}$), 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. (2019) and include household income, which represents a measure of household resources. Additionally, we incorporate a variable capturing whether child i is the oldest in the family, in light of recent evidence of higher parental investment for first-born children (Price, 2008; Pavan, 2016; Black et al., 2018; Lehmann et al., 2018; Breining et al., 2020) and a binary variable capturing whether the mother received WIC benefits during pregnancy.

3.2   Measurement System: Latent 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 $\theta_{S,\tau}$:

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

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.[12] 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, $v_{S,k,i,\tau}$ represent mean-zero measurement errors with variance $\sigma_{v_{S,k,\tau}}^2$, 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 (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

---

[11]$I_{S,i,t}$ depends on the same observed characteristics as in equation (1).

[12]We have alternative estimated a triangular measurement system in which the set of non-cognitive measures load 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 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:

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

where $\sigma_{\theta_{S,t}}^2$ indicates the variance of the latent factor S at time t. To describe its off-diagonal elements, first we split vector $\mathbf{T}_t$ into two blocks, where each block has measures dedicated to a different skill as follows:

$$\left[\begin{array}{c} \mathbf{T}_{NC,t} \\ \mathbf{T}_{C,t} \end{array}\right] = \left[\begin{array}{c} \mathbf{X}_{t,T}\beta_{NC,t} + \Lambda_{NC,t}\theta_{NC,t} + \mathbf{v}_{NC,t} \\ \mathbf{X}_{t,T}\beta_{C,t} + \Lambda_{C,t}\theta_{C,t} + \mathbf{v}_{C,t} \end{array}\right]$$

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 $COV\left(\mathbf{T}_t | \mathbf{X}_{t,T}\right)$, encompassing elements *within* a block:

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

and across blocks:

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

We use the $(\mathcal{K}_{NC} + \mathcal{K}_C)(\mathcal{K}_{NC} + \mathcal{K}_C - 1)/2$ off-diagonal elements of $COV\left(\mathbf{T}_t | \mathbf{X}_{t,T}\right)$ 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 $COV\left(\mathbf{T}_t | \mathbf{X}_{t,T}\right)$ 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 $COV\left(\mathbf{T}_t | \mathbf{X}_{t,T}\right)$ to identify $\sigma_{v_{S,k,t}}^2$.

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:

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

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

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

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

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: $COV\left(T_{NC,1,t}, T_{C,1,t} \,|\mathbf{X}_{t,T}\right) = \sigma_{\theta_t^{NC}\theta_t^{C}}$. 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 $\mathbf{T}_t$.[13]

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 3.4, 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.[14]

3.3 Parental Skills and Investment

Given our interest in understanding the impacts of retention on children's non-cognitive skill development, we additionally incorporate measures of parents' non-cognitive skills to our analysis. We take advantage of three baseline measures of parents' psychological well-being. The first variable measures parents' mental well-being, whereas the other two variables encompass measures of parental anxiety. Following the measurement system for latent skills, we posit that observed measures of parents' psychological well-being $T_{p,i,t}$ ($p \in \mathcal{P}$) measure

---

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

[14]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.

latent parental non-cgnitive skills ($\theta_{P,i}$) with error as follows:

$$T_{p,i,t} = \beta_{p,t}X_{p,i,t} + \alpha_P\theta_{P,i} + v_{p,i,t} \tag{7}$$

where $v_{p,i,t}$ are measurement errors which are independent across paerntal skill measures p, observed characteristics and latent parental non-cognitive skills. To recover the latent parental non-cognitive skills factor, we follow the identification argument outlined above for the children's skill measures.

We additionally capture two measures of parental investments in their children, encompassing dedicated investments in their kids' cognitive ($\mathcal{I}_{C,i,t}$) and non-cognitive ($\mathcal{I}_{NC,i,t}$) skill development.[15] We estimate the measures of parental investment following a principal components analysis, in which we posit that the number of books available in the household and weekly time parents spend reading to their children represent measures of cognitive investment, whereas the number/types of activities done with their children, the frequency with which they hit their kids, engagement in school activities along with their attitudes towards parenting and their relationship with their child denote parental investments in their children's non-cognitive skill development. Upon defining the two measures of parental investments in their children, we estimate equation (3) to account for the potential endogeneity of parental investment to their children's latent skills at time t. We follow Cunha et al. (2010); Agostinelli and Wiswall (2016a). We first estimate equation (3) for each skill component S and subsequently include predicted $\widehat{\mathcal{I}_{S,i,t}}$ as an input in the production function.

3.4 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.

3.4.1 Latent Factors

We estimate the distribution of the latent skill factors at time t ($\widehat{F_{\theta_t^{NC},\theta_t^C}}$), $t+1$ ($\widehat{F_{\theta_{t+1}^{NC},\theta_{t+1}^C}}$) and latent parental non-cognitive skills ($\widehat{F_{\theta_t^P}}$) each following a mixture of two normals, which imposes few restrictions on the underlying distribution of the factors. While Freyberger

---

[15]Reduced form results, presented in Table A.2, motivate the choice of dedicated investment measures, as we do not find evidence that investments in children's cognitive skills affect their non-cognitive skill outcomes in the endline survey round, and viceversa.

(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).[16] We additionally assume the error terms in the measurement system ($\boldsymbol{v}_{S,\tau}$), parental non-cognitive skill equations ($\boldsymbol{v}_P$), retention equation ($\boldsymbol{\varepsilon}_t$) and in the skills production functions ($\boldsymbol{\eta}_{S,t}^R$) are normally distributed.

3.4.2 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:[17]

$$\widehat{\boldsymbol{\xi}_{S,t+1}} = \mathbf{T}_{S,t+1} - \widehat{\boldsymbol{\beta}_{S,t+1}}\mathbf{X}_{t+1} = \widehat{\boldsymbol{\alpha}_{S,t+1}}\theta_{S,t+1} + \boldsymbol{v}_{S,t+1} \tag{8}$$

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 (8) as retention dependent:

$$\widehat{\xi_{S,k,i,t+1}^0} = \widehat{\alpha_{S,k,t+1}}\theta_{S,i,t+1}^0 + v_{S,k,i,t+1} \qquad \text{if } R_{i,t} = 0 \tag{9}$$

$$\widehat{\xi_{S,k,i,t+1}^1} = \widehat{\alpha_{S,k,t+1}}\theta_{S,i,t+1}^1 + v_{S,k,i,t+1} \qquad \text{if } R_{i,t} = 1 \tag{10}$$

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 ($\boldsymbol{v}_{S,t+1} \perp \boldsymbol{v}_{S,t} \perp \boldsymbol{\eta}_{S,t}^R$), we replace $\theta_{S,t+1}^R$ by its production function $g_{S,t+1}^R(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_t)$ into equations (9) and (10), yielding:[18]

$$\widehat{\xi_{S,k,i,t+1}^0} = \widehat{\alpha_{S,k,t+1}}g_{S,t+1}^0(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_{S,t}) + \zeta_{S,k,i,t+1}^0 \qquad \text{if } R_{i,t} = 0 \tag{11}$$

$$\widehat{\xi_{S,k,i,t+1}^1} = \widehat{\alpha_{S,k,t+1}}g_{S,t+1}^1(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_{S,t}) + \zeta_{S,k,i,t+1}^1 \qquad \text{if } R_{i,t} = 1 \tag{12}$$

where $\boldsymbol{\zeta}_{S,t+1}^R = \widehat{\boldsymbol{\alpha}_{S,t+1}}\boldsymbol{\eta}_{S,t}^R + \boldsymbol{v}_{S,t+1}$ is a compounded error term with mean zero and variance $\Omega_{\zeta_{S,t+1}^R}$. Note that its off-diagonal elements are given by $\widehat{\alpha_{S,k,t+1}}\widehat{\alpha_{S,k',t+1}}\sigma_{\eta_{S,t}^R}^2 + \sigma_{v_{S,t+1}}^2$ and the diagonal elements follow $(\widehat{\alpha_{S,k,t+1}})^2\sigma_{\eta_{S,t}^R}^2 + \sigma_{v_{S,k,t+1}}^2$. Since the elements across both the

---

[16]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.

[17]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.

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

diagonal and off-diagonal are identified in the first estimation step, it is straight-forward to see that $\Omega_{\zeta_{S,t+1}^R}$ is identified as well.

3.4.3 Factor Normalizations

As discussed above, Agostinelli and Wiswall (2016b) have shown 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 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 possible differences in the scaling of the measures ($T_{S,1,t}$ and $T_{S,1,t+1}$) where normalizations were imposed. We 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 $E[T_{S,k,\tau}|\mathbf{X}_\tau]$ 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 so that they end up treating 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 qubic polynomial in the CES parameters ($\mathbf{P}_3(\gamma_S, \gamma_P, \gamma_I, \rho_S)$).[19] As a result, we use $\mathbf{P}_3(\gamma_{NC,S}^R, \gamma_{P,S}^R, \gamma_{I,S}^R, \rho_S^R)$ 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 $E[\xi_{S,k,t+1}^R] = -\alpha_{S,k,t+1}^R \mathbf{P}_3(\gamma_{NC}^R, \gamma_P^R, \gamma_I^R, \rho_S^R)$.

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

---

[19]See Figure B.1.

Table 3: Goodness of Fit of the Model

	Retained	Attention Focus	Inhibitory Control	SSRS	Math	Science	Reading
Means							
Actual	0.0615	4.8592	5.0616	0.0487	0.0095	0.0053	0.0081
Predicted	0.0618	4.8578	5.0615	0.0481	0.0113	0.0057	0.0093
Std. Devs.							
Actual	0.2402	1.2798	1.2231	0.9568	0.9949	0.9982	0.9954
Predicted	0.2408	1.3302	1.2862	1.0386	0.9904	1.0883	1.1100

Note: Table 3 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.

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 (12) becomes:

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

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.

3.4.4 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, parents' education, SES index and poverty status along with child-level characteristics, including gender, age and race. The retention decision additionally includes school-level characteristics along with repetition policies. The investment equations include income, an indicator of being the firstborn in the family, and an indicator of WIC receipt during pregnancy. 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.

Figure 1: Actual vs. predicted test scores cumulative distributions at $t = 1$
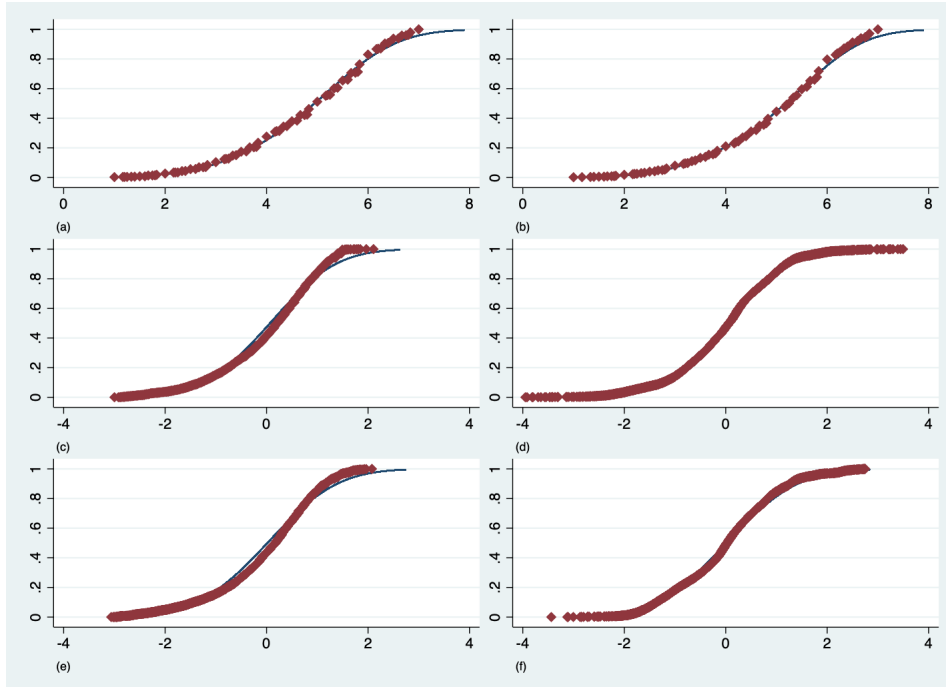
Note: Figure 1 presents the predicted and observed distribution of baseline test scores and non-cognitive skill measures. Predicted values come from simulations based on the estimated parameters of the model. Actual (diamond) and predicted (line) cumulative distributions plotted of the following test scores: (a) attention focus (b) inhibitory control (c) SSRS (d) math (e) science (f) reading.

4 Model Results

4.1 Goodness of Fit

We first examine the accuracy of the model in predicting retention and test score outcomes. In the first column of Table 3, we contrast students' observed and simulated retention outcomes, showing the model matches the likelihood of retention across survey rounds (6.15%). We further take advantage of estimated model parameters to simulate students' observed skill measures. As shown in the remaining columns of Table 3, the model closely matches the first and second moment of observed test scores and non-cognitive skill measures at baseline. In Figure 1, we confirm these results by showing that the distribution of observed and simulated skill measures largely mirror each other.

4.2 Variance Decomposition

Table B.2 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 2, 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). We find that latent cognitive ability explains an important share of the variance in test scores, ranging from 20% in science to upwards of 50% in math. Similarly, $\theta_{NC,t}$ accounts for 30-60% of the variance in the baseline non-cognitive skill measures. Lastly, 15-50% 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. These results thus show the limitation of empirical approaches as presented in Section 2.

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.3). In the second panel of Figure 2, 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 19% to 40%. 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 3, 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 3 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. These results thus remark substantial differences between observed test scores and latent abilities.

**Figure 2:** Variance Decomposition of Test Scores and Skills Distribution

(a) Variance Decomposition of Test Scores at t **(b)** Variance Decomposition of Test Scores at $t+1$
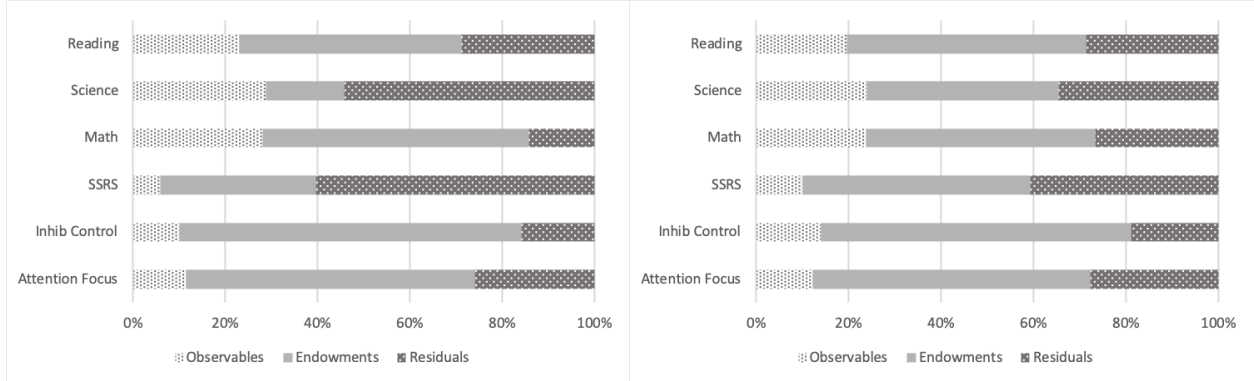
Note: The first panel of Figure 2 presents the share of the variance of test scores and non-cognitive skill measures explained by observable controls $\mathbf{X}_T$, latent cognitive and non-cognitive skills $\Theta_t = \begin{bmatrix} \theta_{C,t} & \theta_{NC,t} \end{bmatrix}$, and the share captured 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 $t+1$.

Figure 3: Joint Distribution of Latent Skill Factors

(a) Joint Distribution at t **(b)** Joint Distribution at $t+1$
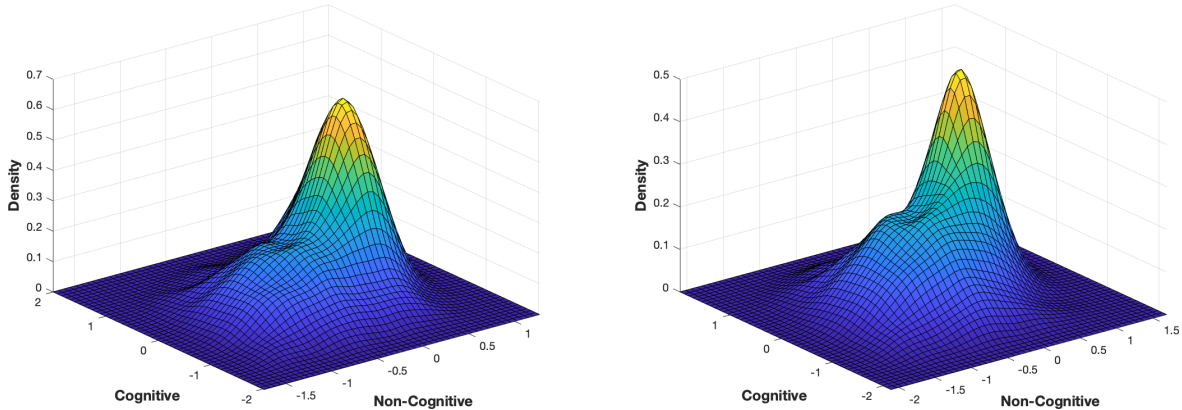
Note: The first panel of Figure 3 presents the estimated latent skills joint distribution at t $[f_{\theta_{C,t},\theta_{NC,t}}(\cdot,\cdot)]$. It was obtained from a random draw of the estimated parameters from the measurement system at time t (presented in Table B.2). The distribution is centered at $(0,0)$. 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.3). 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.

4.3   Parental Skills and Investment

The empirical model specified in Section 3 additionally incorporated the importance of parents' non-cognitive skills in the skill development process. We present the estimated param-

Table 4: Estimated Parameters of Parental Investment Equation

	Inv. Cognitive		Inv. Non-Cognitive
Oldest Child	-0.001	Oldest Child	-0.055*
	(0.026)		(0.030)
HH Income	0.190***	HH Income	0.106***
	(0.018)		(0.020)
WIC Pregnant	-0.283***	WIC Pregnant	-0.133***
	(0.033)		(0.039)
Parents NC Skills	0.020	Parents NC Skills	0.329***
	(0.022)		(0.024)
Non-Cognitive Skills	-0.047*	Non-Cognitive Skills	0.109***
	(0.027)		(0.031)
Cognitive Skills	0.177***	Cognitive Skills	0.088***
	(0.023)		(0.025)
Observations	6817	Observations	5851

Note: Table 4 presents estimated coefficients on parental investments. The first column presents evidence on parents' cognitive investment choices $(\mathcal{I}_{C,t})$, whereas the second column presents evidence on parents' non-cognitive investment choices $(\mathcal{I}_{NC,t})$. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

eters of equation (7) in Table B.4. Parents in two-parent household exhibit higher measures of psychological well-being, as do higher SES parents. Factor loadings across the two un-anchored measures are positive and significant.

Moreover, our empirical strategy allows for parental investment choices to respond endogenously to children's latent skills at time t. We present the estimated parameters of equation (3) in Table 4. First, higher income parents exhibit higher investment in their children's cognitive and non-cognitive skill development, fitting in with evidence presented by (Cunha et al., 2010; Attanasio et al., 2019). At the same time, being the oldest child is not associated with differential investment in the cognitive dimension, yet results in lower non-cognitive investment, conditional on other characteristics. We further note that having received WIC benefits during pregnancy is strongly associated with lower parental investment across the two skill dimensions considered in the model.

Our model further considers the importance of parents' latent non-cognitive skills to their investment choices. Interestingly, we find that higher θ_{NC} parents exhibit substantially higher investment in their children's non-cognitive skill development, yet this is not the case across the cognitive dimension. These results thus highlight the importance of considering different dimensions of parental investment choices in the skill development process. Furthermore, parental investments reinforce children's baseline cognitive skills, as a one σ increase in $\theta_{C,t}$ is associated with a 0.18 and 0.09 σ increase in cognitive and non-cognitive investments, respectively. Lastly, cognitive investments compensate for children with low non-cognitive skills, yet parental non-cognitive investments are reinforcing for higher $\theta_{NC,t}$ children.

5 Grade Retention and Dynamic Skill Development

5.1 Determinants of Grade Retention

In Table B.5, we present marginal effects in 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 (p-value $= 0.077$) are more likely to have repeated a grade.

Furthermore, our results remark the critical role played by students' latent abilities at time t in the retention process. First, a one σ increase in students' cognitive ability reduces the likelihood of retention by 4.4 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 1.3 pp. These findings are confirmed graphically in Figure 4, which shows that upwards of 12% 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 12% for students in the top $\theta_{NC,t}$ decile to 16.5% for their peers in the bottom decile. Similarly, Figure 5 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 non-cognitive skills have no discernible impact on the prevalence of early retention.

5.2 Skill Production Process

In Table 5, 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

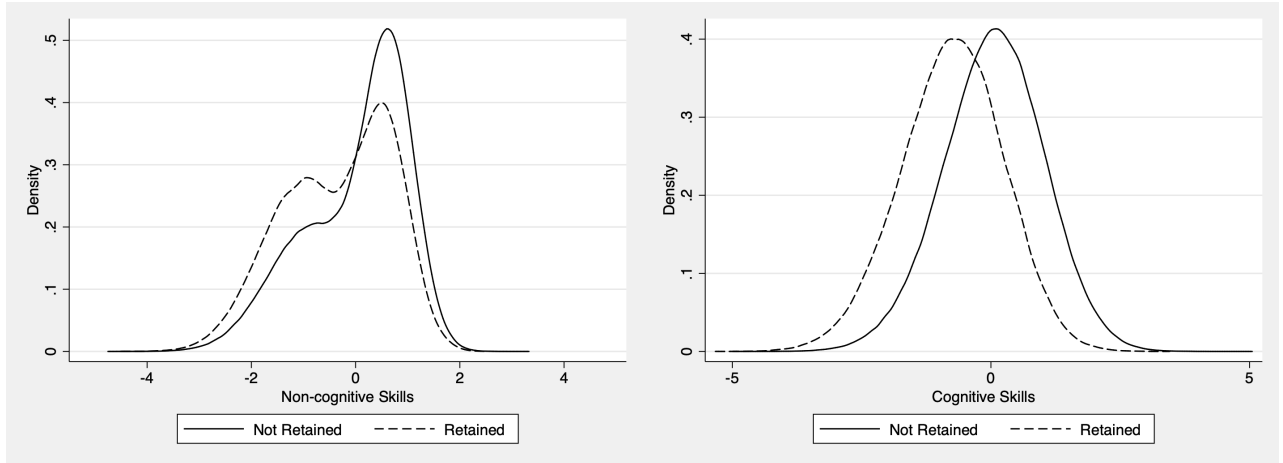**Figure 4:** Probability of Retention at t by Initial Level of Skills
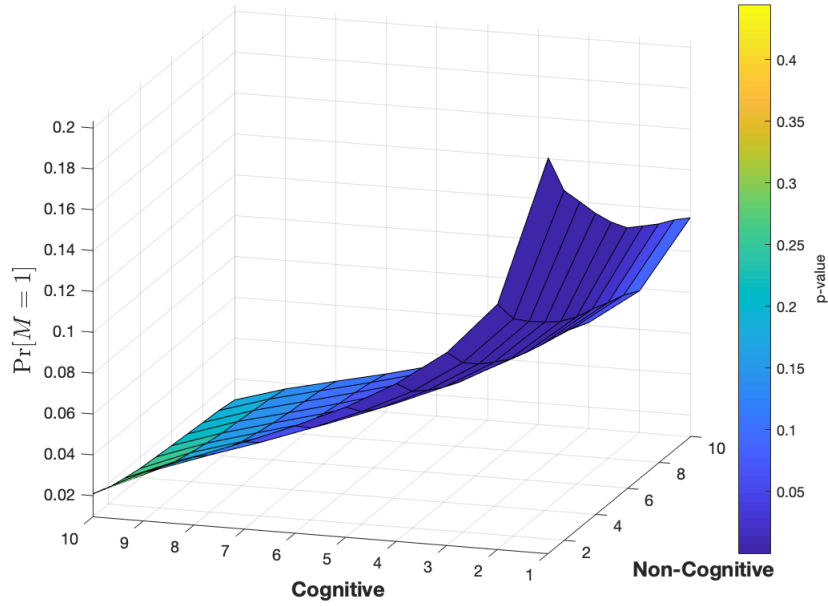
Note: Figure 4 presents the estimated probability of retention across joint deciles of the latent skills distribution in period t.

Figure 5: Sorting into Grade Retention by Skill Levels

(a) Non-cognitive Skills   **(b)** Cognitive Skills

Note: Figure 5 presents the marginal densities of latent non-cognitive and cognitive skills in period t by retention status.

σ_{NC}^1 equals 1.44. 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_C^1, equaling 0.75. We remark that despite normalizing the location of the factors ($E(\theta_{S,i,t+1}) = 0$), we recover elasticities of substitution which

Table 5: 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)
$\sigma_s^R = \frac{1}{1-\rho_s^R}$	0.980	1.440	1.057	0.748
	[0.950,1.010]	[1.233,1.646]	[1.025,1.088]	[0.643,0.853]
$\gamma_{C,t}$	0.034	0.058	0.869	0.645
	[0.005,0.064]	[-0.041,0.158]	[0.852,0.887]	[0.582,0.707]
$\gamma_{NC,t}$	0.737	0.695	0.048	0.077
	[0.708,0.764]	[0.572,0.795]	[0.030,0.074]	[0.031,0.179]
$\gamma_{I,t}$	0.255	0.212	0.090	0.323
	[0.207,0.311]	[0.079,0.458]	[0.064,0.125]	[0.204,0.470]
γ_P	-0.026	0.034	-0.007	-0.045
	[-0.056,0.003]	[-0.065,0.134]	[-0.024,0.010]	[-0.107,0.018]

Note: Table 5 presents the estimated parameters of equation (1) across retention status for the production of cognitive and non-cognitive skills at $t+1$. σ_s^R is the elasticity of substitution between skill endowments, parental skills and investments in retention status R. 95% confidence intervals are reported in brackets.

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.

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_{NC}^R parameter in the production function of non-retained and retained students equals 0.737 and 0.695, respectively. At the same time, period t cognitive skills have little *direct* influence on students' endline non-cognitive 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 (γ_C^1 exceeds 0.86). Unlike in the production function of non-cognitive skills, we find evidence of cross-productivity in the production function of cognitive skills among the retained.

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 0.21 in the production of non-cognitive skills. Meanwhile, parents' investments play a smaller role in the development of non-retained children's cognitive skills, while exhibiting a far larger impact for their 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 4. By considering different dimensions of parents' investments in their children's skill development, our results fit in with recent

evidence in Attanasio et al. (2020). The authors find that material investments contribute to the production of Colombian children's cognitive skills, whereas time investments improve children's non-cognitive skill development. Lastly, parents' non-cognitive skills (θ_P) do not directly impact their children's skill outcomes in period $t + 1$. Nonetheless, we had previously found (Table 4) that parents with higher θ_P engage in increased non-cognitive skill investments for their children. As such, the parents' non-cognitive skills indirectly improve their children's $\theta_{NC,t+1}$ outcomes through their investment decisions.

We present graphical evidence of the *total* productivity of latent skills in Figure 6. While self-productivity dominates the skill production process, initial-period 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 cognitive skills at $t + 1$ 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 their importance of initial inputs varies across retention status.[20] We next analyze the impacts of grade retention on students' cognitive and non-cognitive skill outcomes.

6 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:

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

(14)

---

[20]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.
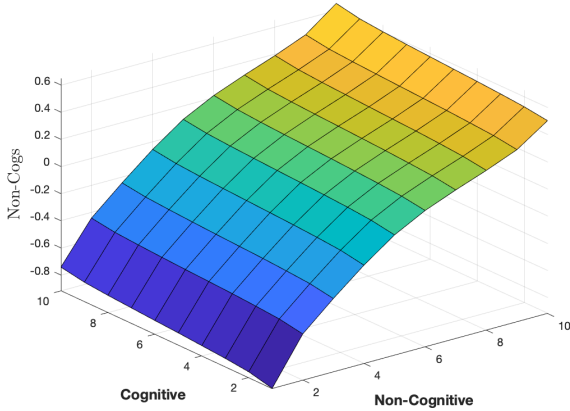
Figure 6: Latent Skills at $t+1$ as a function of $\boldsymbol{\theta_t}$
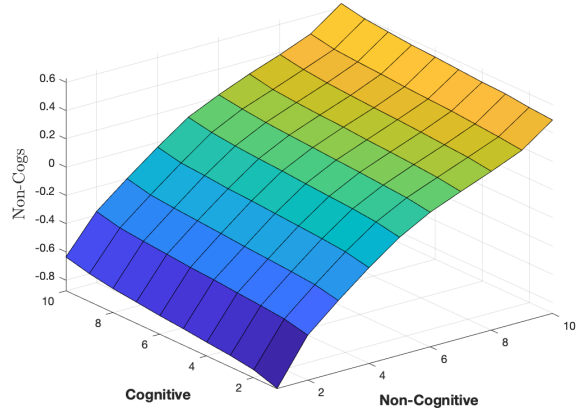
(a) $\theta_{C,t+1}$ as a function of $\boldsymbol{\theta_t}$: Non-Retained
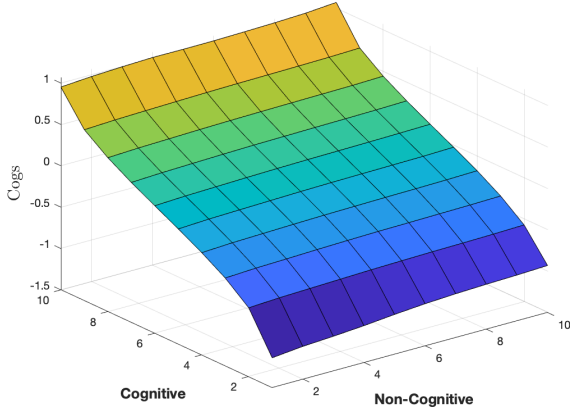
(b) $\theta_{C,t+1}$ as a function of $\boldsymbol{\theta_t}$: Retained
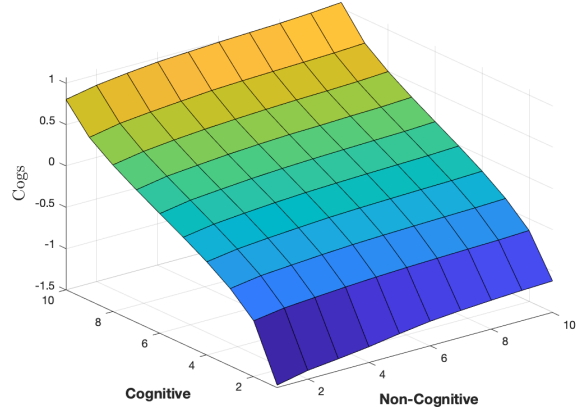
(c) $\theta_{NC,t+1}$ as a function of $\boldsymbol{\theta_t}$: Non-Retained

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

Note: The first two panels of Figure 6 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.

While equation (14) 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 as follows:

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

(15)

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:

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

6.1 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 6. 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 4). 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 0.082 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.[21] In this specification, grade retention would lower students' math, reading and science test scores by 0.197, 0.347 and 0.026 standard deviations, respectively (Figure B.2). The estimated (negative) impacts of retention would far exceed those estimated under our preferred model, which diferentiates test scores from latent skills. Following the evidence presented in Figure 2, these results further highlight the extent to which test scores

---

[21]We take advantage of model parameters to estimate a Roy model with unobserved heterogeneity as in Heckman et al. (2006, 2018).

Table 6: Estimated Impacts of Grade Retention

|	$\theta_{C,t+1}$	$\theta_{NC,t+1}$
	(1)	(2)
ATE	-0.019	0.014
	(0.009)**	(0.010)
TT	-0.082	0.020
	(0.009)***	(0.010)**

Note: Table 6 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 (14)). The second row presents the treatment on the treated parameter (equation (15)). Standard errors reported in parenthesis: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

measure latent abilities with error.

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 7. 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 0.22 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 0.12 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 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. We provide evidence on this front in Section 6.3

Figure 7: Average Treatment Effect of Retention on $t+1$ Cognitive Skills
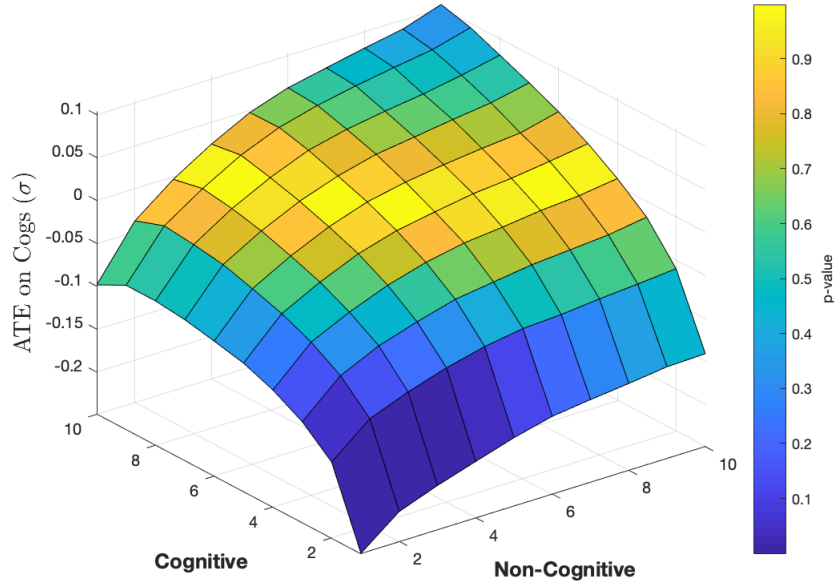
Note: Figure 7 presents the estimated impacts of grade retention on students' cognitive skill development ($\theta_{C,t+1}$) across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

6.2 Effects on Non-Cognitive Skill Development

As discussed above, grade retention implies that students will be separated from their class-mates, potentially resulting in stigmatization and thus affecting their non-cognitive skill development. In the second panel of Table 6, 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 corresponds to 0.014 σ of $\theta_{NC,t+1}$. The estimated ATE follows the same sign as the OLS estimates presented in Table 2, yet is smaller in magnitude vis-a-vis reduced form estimates. We additionally present evidence on the treatment on the treated parameter, which indicates that retained students also benefited from being held back, reaching 0.02 standard deviations.

Figure 8 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 0.15 standard deviations for students in the joint

bottom decile of the initial 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. Solis (2017) finds that first grade repetition yields mixed impacts on students' self-reported self-esteem and Eren et al. (2018) find that retention for eighth graders in Louisiana increases the number of absences and the prevalence of disciplinary incidents.[22] 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 2b. 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.3), 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.

6.3   Importance of Parental Investment

An important consideration in our context is whether parents respond to the retention event by changing their investment decisions. In our context, parents could reinforce or reduce their investment on retained children, thus exacerbating or muting the impacts of the event on the skill development process. In fact, investment decisions may differ across cognitive and non-cognitive skills, as, for instance, parents may choose to focus on boosting their children's self-confidence in light of retention, or they may focus on ensuring the child learns the academic content she failed to master before being retained. We take advantage of investment measures available in the endline survey of the ECLS-K, encompassing investment across both skill

---

[22]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 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.

Figure 8: Average Treatment Effect of Retention on $t+1$ Non-cognitive Skills
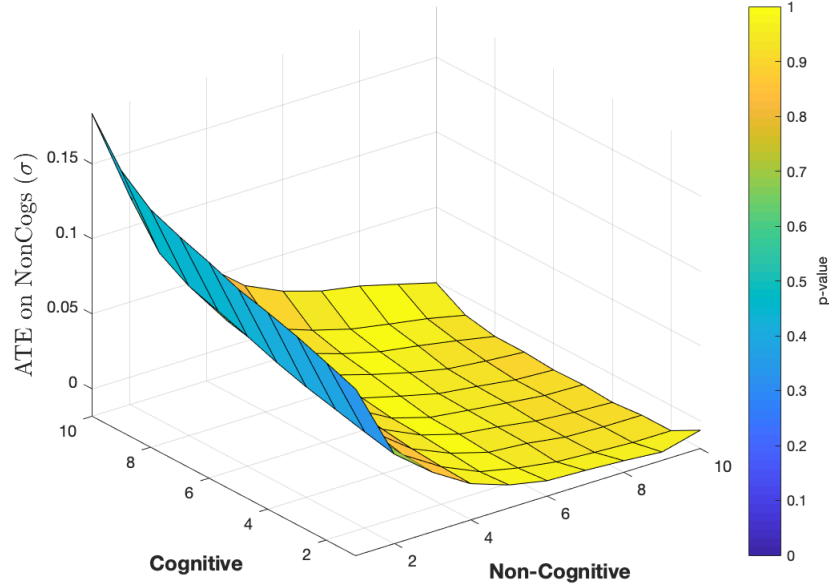
Note: Figure 8 presents the estimated impacts of grade retention on students' non-cognitive skill development ($\theta_{NC,t+1}$) across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

dimensions.[23] We take advantage of our empirical framework to recover potential investment outcomes ($\mathcal{I}_{S,i,t+1}^{R}$) across children's retention status (R_i) as follows:

$$\mathcal{I}_{S,i,t+1}^{R} = \beta_{X,R}X_{i,t+1,R} + \alpha_{P,R}\theta_{P,i} + \alpha_{C,t,R}\theta_{C,i,t} + \alpha_{NC,t,R}\theta_{NC,i,t} + \alpha_{P}\theta_{P,i} + e_{S,i,t+1}^{R} \qquad (16)$$

where $e_{S,i,t+1}^{R}$ is an error term which is independent of observed characteristics and latent initial skills. Equation (16) allows for parental investments at $t+1$ to directly depend on child i's retention status, thus allowing us to recover the impact of retention on investment decisions.

We present the estimated impacts of retention on parents' investments at $t+1$ in Figure 9. On average, grade retention leads parents to increase their investments in their children's cognitive skill development by 0.09 standard deviations. We find strong heterogeneity across the initial skill distribution, as compensating parental investments are strongly increasing in students' period t non-cognitive skills, while decreasing in the cognitive dimension. As

---

[23]Our measure of endline cognitive investments includes the frequency with which they read to their children, how often they practice writing with their kids and the number of books in the household. The measure of non-cognitive investment includes variables regarding parents' school engagement, their relationship with their children as well as the number/frequency of activities such as telling stories, playing games and talking about nature. We construct the two investment measures using principal components analysis.

Figure 9: Average Treatment Effect of Retention at t on Parental Investment in $t+1$

(a) Effects on Non-Cognitive Investment

(b) Effects on Cognitive Investment
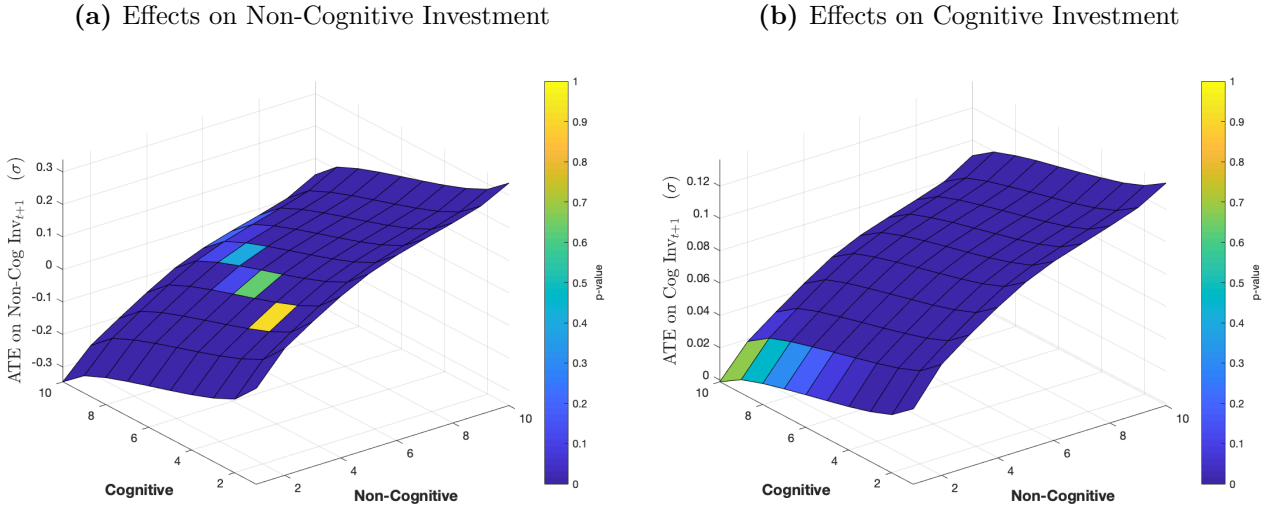
Note: Figure 9 presents the estimated impacts of grade retention on parental non-cognitive skill investment (Panel A) and cognitive investment (Panel B) in period $t+1$ across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

result, retention drives parents with high $\theta_{C,t}$ and low $\theta_{NC,t}$ to reduce their investments in their children's non-cognitive skill development, in the range of 0.3 σ. The extent of compensating investment in the non-cognitive dimension is smaller, as the estimated ATE equals 0.035 standard deviations. However, the extent of heterogeneous investments across students' initial skills follows in the same direction as for $\mathcal{I}_{NC,t}$. All in all, these results indicate that parents' directly respond in light of their children being held back. On average, they increase their investments in their kids' multidimensional skill development, yet their responses exhibit important differences across children's initial skills.

7 Policy Implications

7.1 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 $\{R_i(p) = 1, R_i(p') = 0\}$.[24] Our empirical framework incorporates school-level retention policies as drivers of the retention process. As such, we follow Fruehwirth et al. (2016) and examine the impacts of not allowing schools to retain children simply at their parents' request.[25] 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:

$$PRTE_S(p, p') \equiv \int \int E[\theta_{S,t+1}(p') - \theta_{S,t+1}(p)|R(p) = 1, R(p') = 0]dF_{\boldsymbol{\theta}}(\underline{\boldsymbol{\theta}}|R(p) = 1, R(p') = 0) \tag{17}$$

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

$$W_{p'}(\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}}) = \frac{E[R(p) = 1, R(p') = 0|\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}}] \times P(\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}})}{E[R(p) = 1, R(p') = 0]} \tag{18}$$

where $W_{p'}(\boldsymbol{\theta_t})$ measures the share of compliers across different deciles of the joint skill distribution at period t.[26]

Empirical Evidence. A policy change which limits parents' capacity to request their children to be retained would reduce the prevalence of early retention by 1.05 percentage points — corresponding to a 17% drop in baseline retention rates in the ECLS sample. Relative to the full sample, policy compliers are more likely to be male, black and Hispanic (Table B.6). Moreover, these students are more likely to come from single parent households and from lower SES-families. 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 0.37 and 0.17 standard deviations, respectively. In Figure B.4, we present evidence on the complier weights introduced in equation (18). Policy compliers are largely drawn from the bottom of the period t latent skills distribution: for instance, 20.9% of compliers come

---

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

[25]The estimated parameters of equation (2) (Table B.5) indicate that this policy change would yield the largest shift in students' retention outcomes among the various school-level policies considered in the model.

[26]$P(\boldsymbol{\theta_t} = \underline{\boldsymbol{\theta_t}})$ 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 inF that decile who are policy compliers.

Figure 10: Policy Relevant Treatment Effect: Changing Parental Request Retention Policy

(a) Effects on Non-Cognitive Skills   **(b)** Effects on Cognitive Skills
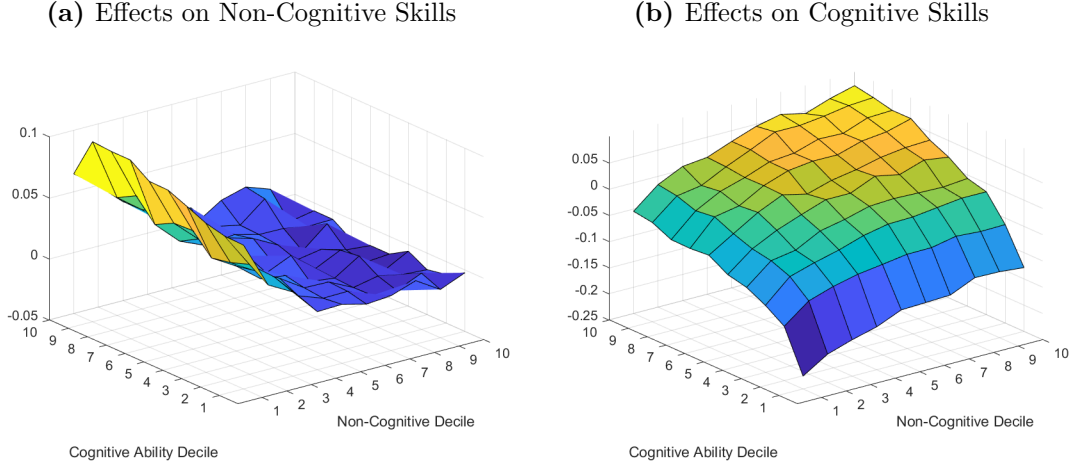
Note: Figure 10 presents the estimated PRTE of a change in school's ability to retain students based on parental requests. The first and second panels present the estimated PRTE on students' non-cognitive skill development ($\theta_{NC,t+1}$) and cognitive skill development ($\theta_{C,t+1}$) across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

from the bottom decile of the $\theta_{C,t}$ distribution, compared to just 3% of their peers in the top decile. Similar differences emerge in the non-cognitive dimension, as well.

Figure 10 presents the estimated PRTE across the period t latent ability distribution, which follows the shape of the average treatment effect (Figures 7-8). The average PRTE directly follows from complier weights presented in Figure B.4 and the heterogeneous impacts presented above. The policy change would thus improve students' period $t+1$ cognitive skills by 0.065 σ, while reducing their non-cognitive abilities by 0.018 standard deviations, though the latter result is not significant.

7.2 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.[27] 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 importance of parental income on their cognitive and non-cognitive investment choices (equation

---

[27]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.

(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, and impose and even 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:

$$ATE_S^Y \equiv \int \int E[g_{S,t+1}^0(\theta_{NC,t}, \theta_{C,t}, \theta_P, \mathcal{I}_{S,t}(\Delta Y)) - g_{S,t+1}^1(\cdot)]dF_{\theta_C,\theta_{NC}}(\underline{\theta}_{C,t}, \underline{\theta}_{NC,t}) \qquad (19)$$

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 7-8) and the productivity of investments in the skill formation process (Table 5). In our setting, the compensating income transfer would increase parents' cognitive and non-cognitive investments in period t by 0.52 and 0.35 standard deviations, respectively.

We present the estimated effects of the compensating income transfer in Figure 11. This policy change would modestly impact children's non-cognitive skill outcomes at endline (Panel A), with no discernible average impacts (0.009 σ). 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 8. On the other hand, the compensating income transfer would positively affect children's cognitive skill development. Implementing the compensating income transfer would increase their cognitive skills by 0.028 σ, on average. 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 0.2 σ 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 0.092 σ 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., 2019). 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 with direct income transfers could yield small improvements in children's skill development,

Figure 11: Estimated Impacts of Compensating Income Transfer on $\boldsymbol{\theta_{t+1}}$

(a) Effects on $\theta_{NC,t+1}$ **(b)** Effects on $\theta_{C,t+1}$
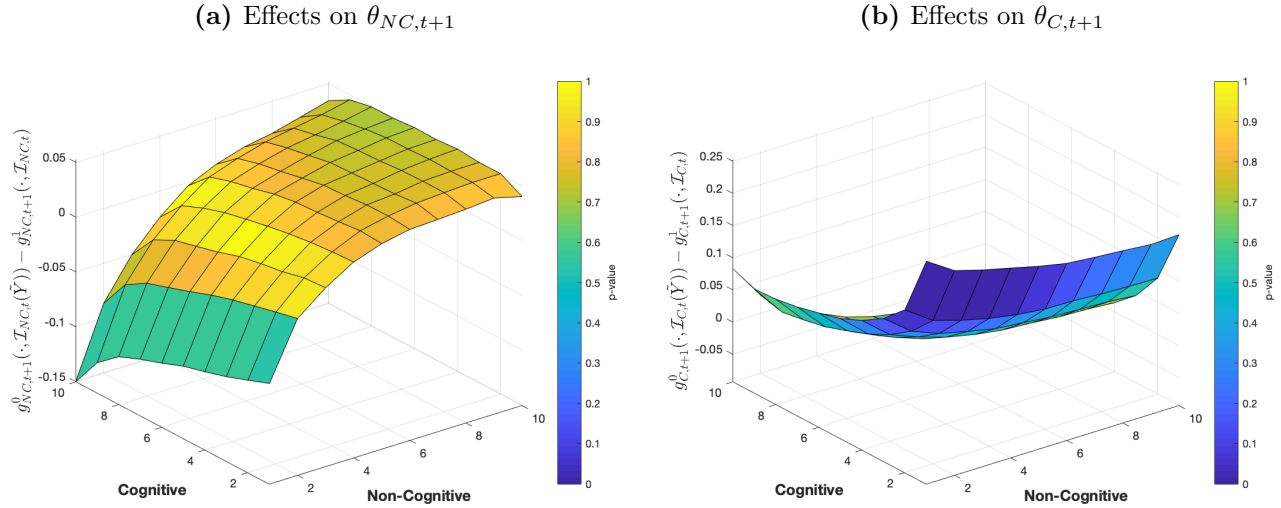
Note: Figure 11 presents the estimated impacts of compensating parents with the estimated cost of retention vis-a-vis holding their children back. We present the estimated impacts on non-cognitive (Panel A) and cognitive skills (Panel B) in period $t+1$ across the joint distribution of latent skills in period t. Results follow from 40,000 simulations based on the estimated parameters of the dynamic model.

depending on the outcome in consideration and their initial skill level.

8   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.

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 proces, 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. At the same time, parents engage in compensating investments across both skill dimensions, which may lessen the negative effects of being held back. We lastly take advantage of the estimated model parameters to examine how policy changes can affect children's

skill development outcomes. We find that changing school policies to reduce the prevalence of retention can improve children's skill outcomes. Similarly, investment-targeted income transfers in lieu of retention positively impact 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 (2019). Human capital development and parental investment in India. *Review of Economic Studies. Forthcoming.*

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.

Black, S. E., E. Grönqvist, and B. Öckert (2018). Born to lead? The effect of birth order on noncognitive abilities. *Review of Economics and Statistics 100*(2), 274–286.

Breining, S., J. Doyle, D. N. Figlio, K. Karbownik, and J. Roth (2020). Birth order and delinquency: Evidence from Denmark and Florida. *Journal of Labor Economics 38*(1), 95–142.

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 Boca, D., C. Flinn, and M. Wiswall (2016). Transfers to households with children and child development. *The Economic Journal 126*(596), F136–F183.

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. E. and Y. Zhou (2018). The black-white gap in non-cognitive skills among elementary school children. *Available at SSRN 3277607*.

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. Gouss, 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., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

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

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica 73*(3), 669–738.

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.

Lehmann, J.-Y. K., A. Nuevo-Chiquero, and M. Vidal-Fernandez (2018). The early origins of birth order differences in childrens outcomes and parental behavior. *Journal of Human Resources 53*(1), 123–156.

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.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2019). Identification of causal effects with multiple instruments: Problems and some solutions. Technical report, National Bureau of Economic Research.

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.

Pavan, R. (2016). On the production of skills and the birth-order effect. *Journal of Human Resources 51*(3), 699–726.

Price, J. (2008). Parent-child quality time does birth order matter? *Journal of human resources 43*(1), 240–265.

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.

Solis, A. (2017). The effects of grade retention on human capital accumulation. Technical report, Working Paper.

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

Appendix

A   Reduced Form Results

Table A.1: Determinants of Grade Retention

	(1)
Math Test Score	-0.035***
	(0.005)
Reading Test Score	-0.009*
	(0.005)
Science Test Score	-0.005
	(0.004)
Composite Social Rating Scales	0.004
	(0.004)
Attentional Focus	-0.023***
	(0.005)
Inhibitory Control	0.001
	(0.005)
Male	0.023***
	(0.006)
Age (R1)	-0.061***
	(0.009)
White	-0.002
	(0.009)
Black	-0.018
	(0.013)
Hispanic	-0.042***
	(0.011)
Both Parents	-0.014*
	(0.007)
Parents' Education	-0.002
	(0.002)
Log HH Income	0.005
	(0.006)
SES Index	0.023***
	(0.008)
Poverty Status	0.033***
	(0.010)
Parental Investment	Yes
School-Level Policies	Yes
Observations	6,817
R^2	0.073

Source: ECLS-K:2011. Note: Table A.1 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. Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Parental Investment and Skill Development

	(1)	(2)	(3)	(4)	(5)	(6)
	Attentional Focus	Inhibitory Control	SRS	Math	Reading	Science
Non-Cognitive Skills Index (R1)	0.318***	0.257***	0.267***	0.045***	0.071***	0.004
	(0.009)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)
Cognitive Skill Index (R1)	0.112***	0.031***	0.017**	0.448***	0.414***	0.453***
	(0.009)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
Parental NC Skills	0.015*	0.012*	0.039***	0.013**	0.003	0.003
	(0.009)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)
Parental Cog Investment	-0.016	-0.007	-0.028**	0.016*	0.019**	0.076***
	(0.012)	(0.009)	(0.011)	(0.009)	(0.009)	(0.009)
Parental NC Investment	0.019*	0.025***	0.045***	-0.003	0.001	-0.007
	(0.012)	(0.009)	(0.011)	(0.008)	(0.008)	(0.008)
Observations	5838	5838	5838	5838	5838	5838
R^2	0.261	0.243	0.196	0.522	0.503	0.529

Source: ECLS-K:2011. Note: Table A.2 presents evidence on the importance of parental investments on children's test score and non-cognitive skill outcomes at endline. The non-cognitive skills index follows from principal components analysis of the three teacher-reported nonn-cognitive measures available at baseline. The cognitive skills index follows from PCA from the three baseline test scores. Parental NC skills follow from measures of parents' sadness and anxiety at baseline. Parental cognitive and non-cognitive investments are defined in Section 3.3. 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	✓	✓	✓		✓		
SES Index	✓	✓	✓		✓		
Instruments							
HH Income				✓			
Birth Order				✓			
WIC at Pregnancy				✓			
Retention Policies					✓		
Latent Factors							
$\theta_{C,t}$	✓	✓	✓	✓	✓	✓	✓
$\theta_{NC,t}$	✓	✓	✓	✓	✓	✓	✓
θ_P			✓	✓	✓	✓	✓
$\mathcal{I}_{C,t}$				✓		✓	
$\mathcal{I}_{NC,t}$				✓			✓

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

Table B.2: Estimated Parameters of Measurement System at t

	Att. Focus	Inhib Cont.	SSRS	Math	Science	Reading
Male	-0.400***	-0.399***	-0.228***	0.063***	0.025	-0.080***
	(0.026)	(0.025)	(0.022)	(0.020)	(0.020)	(0.021)
Age in Months	0.407***	0.294***	0.153***	0.712***	0.418***	0.535***
	(0.038)	(0.036)	(0.031)	(0.029)	(0.030)	(0.031)
Black	-0.053	-0.019	0.019	-0.273***	-0.621***	0.032
	(0.048)	(0.045)	(0.039)	(0.036)	(0.037)	(0.037)
Hispanic	0.025	0.023	-0.027	-0.220***	-0.649***	-0.095***
	(0.036)	(0.033)	(0.029)	(0.028)	(0.028)	(0.029)
Other Race	0.070*	0.046	-0.016	0.128***	-0.338***	0.173***
	(0.040)	(0.037)	(0.034)	(0.033)	(0.033)	(0.034)
Lives BothParents	0.222***	0.241***	0.163***	0.122***	0.033	0.121***
	(0.033)	(0.031)	(0.027)	(0.026)	(0.026)	(0.027)
Parental Edu	-0.005	-0.011	0.000	0.041***	0.065***	0.043***
	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)	(0.008)
SES	0.212***	0.135***	0.086***	0.334***	0.206***	0.353***
	(0.029)	(0.028)	(0.024)	(0.023)	(0.023)	(0.024)
Non-Cogn Skills	1.821***	1.893***	1			
	(0.037)	(0.036)	■			
Cognitive Skills				1.095***	0.593***	1
				(0.018)	(0.015)	■
Constant	2.648***	3.562***	-0.826***	-4.669***	-3.094***	-3.700***
	(0.256)	(0.241)	(0.210)	(0.199)	(0.201)	(0.207)
Observations			6817			

Note: Table B.2 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.3: Estimated Parameters of Measurement System at $t+1$

	Att. Focus	Inhib Cont.	SSRS	Math	Science	Reading
Male	-0.423***	-0.386***	-0.302***	0.206***	0.134***	-0.163***
	(0.024)	(0.018)	(0.022)	(0.020)	(0.020)	(0.021)
Age in Months	0.222***	0.127***	0.091***	0.265***	0.304***	0.208***
	(0.033)	(0.025)	(0.030)	(0.030)	(0.029)	(0.030)
Black	-0.014	-0.082***	-0.132***	-0.597***	-0.568***	-0.218***
	(0.041)	(0.030)	(0.037)	(0.037)	(0.037)	(0.037)
Hispanic	0.104***	0.068***	0.069**	-0.267***	-0.367***	-0.143***
	(0.032)	(0.023)	(0.028)	(0.028)	(0.027)	(0.028)
Hispanic	0.174***	0.106***	0.107***	0.075**	-0.070**	0.024
	(0.036)	(0.027)	(0.033)	(0.033)	(0.033)	(0.033)
Lives BothParents	0.267***	0.213***	0.276***	0.129***	0.053**	0.109***
	(0.030)	(0.023)	(0.027)	(0.026)	(0.026)	(0.026)
Parental Edu	-0.001	-0.002	-0.007	0.020***	0.048***	0.036***
	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
SES	0.154***	0.085***	0.116***	0.310***	0.251***	0.313***
	(0.026)	(0.019)	(0.023)	(0.023)	(0.023)	(0.023)
Non-Cogn Skills	1.268***	1.009***	1			
	(0.021)	(0.015)	■			
Cognitive Skills				0.990***	0.896***	1
				(0.017)	(0.016)	■
Constant	2.294***	3.075***	-0.453**	-1.890***	-2.367***	-1.642***
	(0.225)	(0.165)	(0.203)	(0.198)	(0.197)	(0.200)
Observations			6817			

Note: Table B.3 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.4: Estimated Parameters of Parental Skill Measurement System

	Sadness	Anxiety I	Anxiety II
Male	-0.003	0.001	0.000
	(0.008)	(0.011)	(0.000)
Age (R1)	-0.008	-0.018	0.000
	(0.012)	(0.016)	(0.000)
Black	0.025*	0.036*	-0.001
	(0.015)	(0.020)	(0.001)
Hispanic	0.007	0.139***	0.001
	(0.012)	(0.015)	(0.000)
Other	0.024*	0.093***	0.000
	(0.013)	(0.018)	(0.001)
Both Parents	0.036***	0.046***	0.000
	(0.011)	(0.014)	(0.000)
Parents' Education	0.003	-0.012***	0.000
	(0.003)	(0.004)	(0.000)
SES Index	0.015	0.040***	0.000
	(0.009)	(0.013)	(0.000)
Parental NC Skills	0.719***	0.835***	1
	(0.011)	(0.015)	■
Observations		6641	

Note: Table B.4 presents the estimated parameters from the measurement system of parental non-cognitive skill outcomes. The first column presents evidence from the parental sandess 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.5: Determinants of Grade Retention (Marginal Effects at the Mean)

	(1)
Male	0.016***
	(0.004)
Age (R1)	-0.076***
	(0.007)
Black	0.013
	(0.008)
Hispanic	-0.007
	(0.007)
Other	0.004
	(0.007)
Both Parents	-0.018***
	(0.005)
Parents' Education	-0.003*
	(0.002)
SES Index	-0.003
	(0.005)
Retained for Immaturity	0.004
	(0.006)
Retained at Parents' Request	0.012*
	(0.006)
Retained for Academic Deficiency	-0.011
	(0.007)
Retained if Failed Test	0.006
	(0.008)
Retained More than Once	0.005
	(0.008)
Retained Without Parents' Consent	0.009
	(0.005)
School Size	-0.000*
	(0.000)
% Free and Reduced Price Lunch	-0.000**
	(0.000)
% Non-White	-0.000
	(0.000)
Parents' NC Skills	-0.001
	(0.003)
Non-Cognitive Skills	-0.013**
	(0.004)
Cognitive Skills	-0.044***
	(0.004)
Observations	6817

Note: Table B.5 presents estimated marginal effects from the retention equation (2). Standard errors in parentheses. ***$p <$ 0.01, **$p < 0.05$, *$p < 0.1$.

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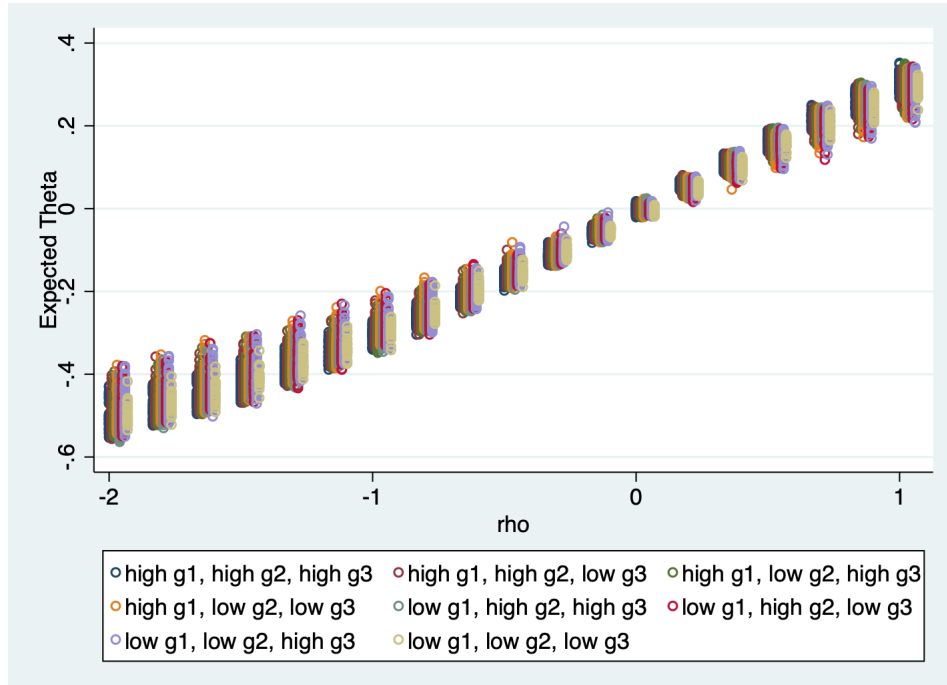
Note: Figure B.1 presents the results of 1,400 different combinations of γ_1, γ_2, γ_3 and ρ parameters in the CES production function (equation (2): $\theta_{t+1} = [\gamma_1 x^\rho + \gamma_2 y^\rho + \gamma_3 z^\rho + (1 - \gamma_1 - \gamma_2 - \gamma_3)n^\rho]^{(1/\rho)}$) where x, y, z and n come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

Table B.6: Characteristics of Policy Compliers for Parental Request Retention Policy

	Mean	SD
	(1)	(2)
Male	0.554	0.017
Age (R1)	5.476	0.013
Black	0.120	0.011
Hispanic	0.159	0.015
Other	0.124	0.013
Both Parents	0.696	0.017
Parents' Education	14.091	0.093
SES Index	0.024	0.027
Non-Cognitive Ability	-0.168	0.023
Cognitive Ability	-0.372	0.031
Complier Share	1.05%	

Note: Table B.6 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.
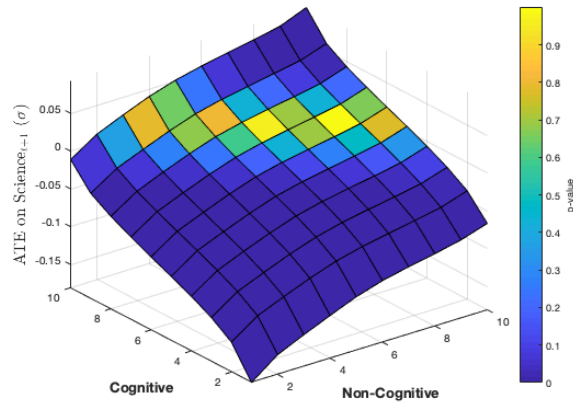
Figure B.2: Estimated Effects of Retention on Test Scores at $t + 1$

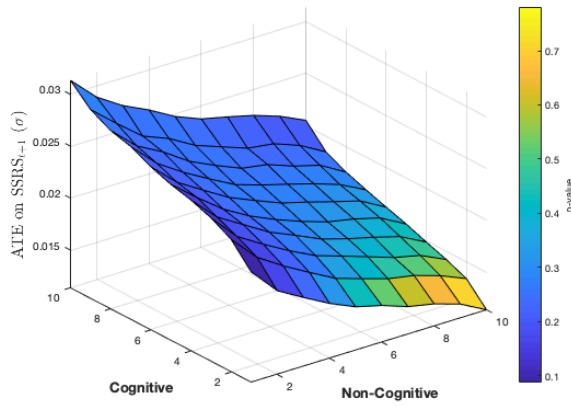**(a)** Math Test Score

(b) Reading Test Score
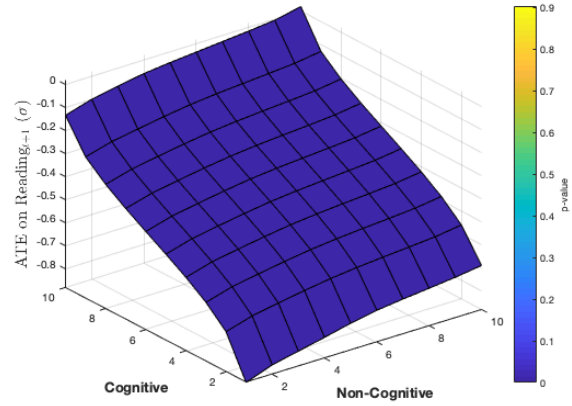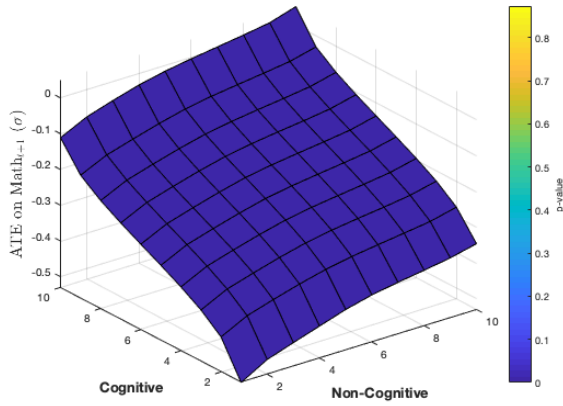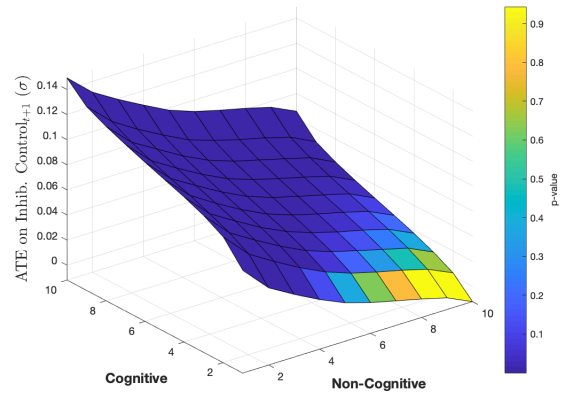

(c) Science Test Score

Note: Figure B.2 presents the estimated impacts of grade retention on endline test scores following a Roy model of retention using observed test scores as outcomes.

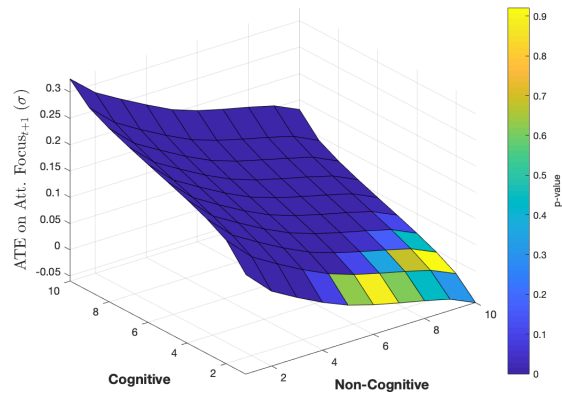**Figure B.3:** Estimated Effects of Retention on Non-Cognitive Skills at $t+1$

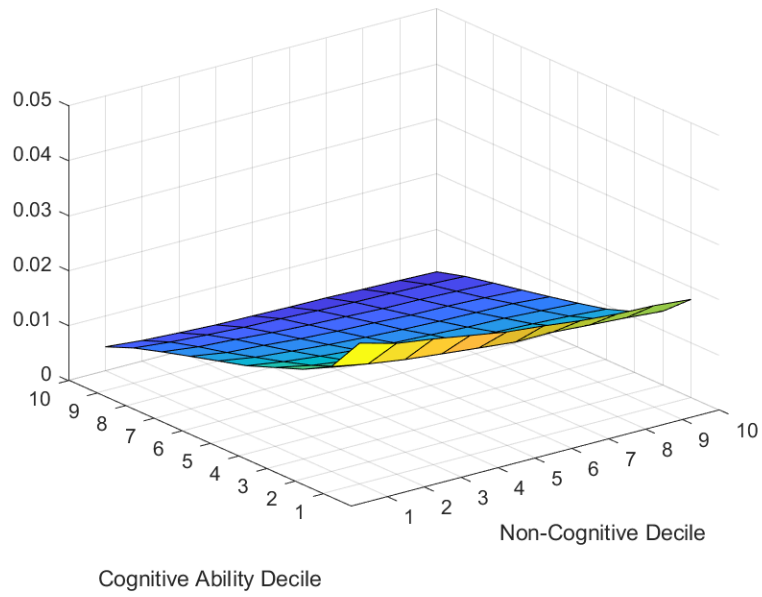**(a)** Social Rating Scale

**(b)** Inhibitory Control

(c) Attentional Focusing

Note: Figure B.3 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.4: PRTE: Complier Weights

Note: Figure B.4 presents estimated complier weights across the latent skill distribution in period t (equation (18)). The simulated policy change limits schools' ability to retain students solely at their parents' request.

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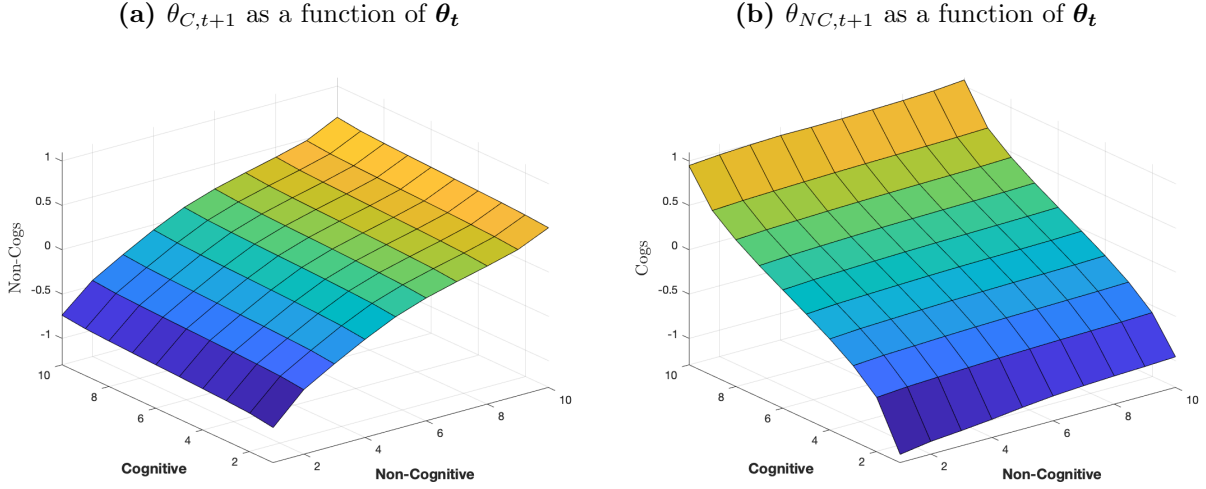
Note: The first panel of Figure C.1 presents the relationship between latent skills at t and cognitive skills at $t+1$. The second panel presents corresponding evidence for the relationship between $\theta_{C,t}$, $\theta_{NC,t}$ and $\theta_{NC,t+1}$.

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:

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

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 $\mathcal{I}_{S,t}$ is associated with increased cognitive and non-cognitive skills at $t+1$ by 0.21 and 0.32 standard deviations, respectively.