

THREE ESSAYS IN THE ECONOMICS OF EDUCATION

By

Alex Johann

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics—Doctor of Philosophy

2023

ABSTRACT

The two arms of my dissertation research are on bringing the role of neighborhood peers and cohort exposure into the school context and investigating gender gaps in noncognitive skills. Both projects are united in expanding our understanding of human capital. Much of the research in the economics of education focuses on the efficacy of specific policies, which is valuable in its own right. However, my research looks upstream of existing policy frameworks to better understand human capital accumulation at a more fundamental level. Through this approach, my dissertation research helps devise new frameworks and tools that not only build upon our current successes but also expand the bounds of what we think is possible.

In my first chapter, **In the School, Down the Block: Achievement Effects of Peers in the School, Neighborhood, and Cohort**, I estimate the effect of mean peer ability on students' test scores using data on all Michigan public school students over thirteen years. I consider peers in the same cohort at school as well as peers in adjacent cohorts, and peers living on the same block. I contribute two novel findings to the literature. First, school peer effects are much stronger than block effects. For peers in the same cohort, the school effect is 10 times larger. Second, cohort membership plays a substantial role in determining peer influence in schools but not in neighborhoods. For students in the same school, the adjacent-cohort peer effect is 40-80% smaller than the same-cohort effect. Meanwhile, for students living on the same block, peer effects are similar, regardless of cohort. These results are robust to a regression discontinuity design focusing on students near the birthdate cutoff for entry into kindergarten. I also find evidence that peers in the older cohort matter more than peers in the younger cohort, particularly in the school context, suggesting that relative age also plays a role in determining peer influence.

My second and third chapters, **Equalizing Inputs, Enduring Gaps: Examining Changes in Levels and Correlates of Gender Gaps in Noncognitive Skills Over Time** and **Raising Boys, Raising Girls: Modeling Gender Differences in the Process of Early Childhood Skill Formation**, approach the issue of gender gaps in noncognitive skills with two alternative analytic frameworks. In both papers, I leverage the smaller but more in-depth ECLS-K datasets to

combine information on teacher-reported noncognitive skills in elementary school with background characteristics and extensive data on parental education activities. The second chapter, **Equalizing Inputs, Enduring Gaps: Examining Changes in Levels and Correlates of Gender Gaps in Noncognitive Skills Over Time**, takes a more descriptive and correlative approach by examining changes over time in boy-girl gender gaps between the two waves of the ECLS-K survey, finding that gender gaps in noncognitive skills remain large and persistent between the 1998-1999 and 2010-2011 nationally representative kindergarten cohorts. Additionally, I use factor analysis and Oaxaca-Blinder decomposition to examine changes in observable inputs to noncognitive skills over time and find that changes in these inputs would predict a narrowing of gender gaps between the two cohorts, despite no such change occurring.

My third chapter, **Raising Boys, Raising Girls: Modeling Gender Differences in the Dynamic Process of Early Childhood Skill Formation**, uses a more structural approach on the same datasets to attempt to provide an answer to the mystery raised in the first: if our usual predictors do not appear valid, what causes gender gaps in noncognitive skills? Differences in inputs? Or differences in production functions? Specifically, I apply the Technology of Skill Formation model proposed by Cunha and Heckman (2008) that takes a Markov chain approach and incorporates both cognitive and noncognitive skills to estimate parameters of skill investment over time. This approach leverages the panel structure of the ECLS-K datasets as well as the availability of both cognitive and noncognitive skill measures. Results are inconclusive, as correctly measuring and determining meaningful parental inputs in the investment process is tricky. I test the robustness of the Cunha and Heckman (2008) model to modeling assumptions and measurement of parental inputs, and find 1) the value-added model sufficiently captures the process of skill formation, relative to the cumulative model of Todd and Wolpin (2003), and 2) parental investment as captured by the measures available in the ECLS-K do not have a statistically detectable impact on the formation of noncognitive skills, regardless of the specification used.

Copyright by
ALEX JOHANN
2023

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Professor Scott Imberman, my committee member and DGS Professor Todd Elder, and the rest of my committee, Professors Ben Zou and Amanda Chuan. I would also like to thank my fiancée, Katlyn Hettinger, for supporting and putting up with me and my two classmates Elise Breshears and Joffré Leroux for our regular pandemic Zoom meetings, which kept me sane and focused in a very tough time.

TABLE OF CONTENTS

CHAPTER 1	IN THE SCHOOL, DOWN THE BLOCK: ACHIEVEMENT EFFECTS OF PEERS IN THE SCHOOL, NEIGHBORHOOD, AND COHORT	1
1.1	Abstract	2
1.2	Introduction	4
1.3	Bias and Identification Strategy	7
1.4	Data	10
1.5	Empirical Strategy	12
1.6	Results	16
1.7	Robustness Checks	19
1.8	Heterogeneity Analysis	24
1.9	Conclusion	25
	BIBLIOGRAPHY	27
APPENDIX A	FIGURES	32
APPENDIX B	TABLES	35
APPENDIX C	APPENDIX TABLES AND FIGURES	40
APPENDIX D	REFLECTION PROBLEM	49
APPENDIX E	CUTOFF IV	50
 CHAPTER 2	 EQUALIZING INPUTS, ENDURING GAPS: EXAMINING CHANGES IN LEVELS AND CORRELATES OF GENDER GAPS IN NONCOGNITIVE SKILLS OVER TIME	 54
2.1	Abstract	55
2.2	Introduction	56
2.3	Data	58
2.4	Results	62
2.5	Conclusion	70
	BIBLIOGRAPHY	72
APPENDIX A	FIGURES	77
APPENDIX B	TABLES	80
APPENDIX C	ONLINE APPENDIX	88
 CHAPTER 3	 RAISING BOYS, RAISING GIRLS: MODELING GENDER DIFFERENCES IN THE PROCESS OF EARLY CHILDHOOD SKILL FORMATION	 120
3.1	Abstract	121
3.2	Introduction	122
3.3	Cross-Model Validation of Human Capital Production Function	125
3.4	Model Description	128
3.5	Data	133
3.6	Results	137
3.7	Robustness	139
3.8	Conclusion	143

BIBLIOGRAPHY 145
APPENDIX A FIGURES 150
APPENDIX B TABLES 151

CHAPTER 1

IN THE SCHOOL, DOWN THE BLOCK: ACHIEVEMENT EFFECTS OF PEERS IN THE SCHOOL, NEIGHBORHOOD, AND COHORT

1.1 Abstract

I estimate the effect of mean peer ability on students' test scores using data on all Michigan public school students over thirteen years. I consider peers in the same cohort at school—as well as peers in adjacent cohorts, and peers living on the same block. I contribute two novel findings to the literature. First, school peer effects are much stronger than block effects. For peers in the same cohort, the school effect is 10 times larger. Second, cohort membership plays a substantial role in determining peer influence in schools but not in neighborhoods. For students in the same school, the adjacent-cohort peer effect is 40-80% smaller than the same-cohort effect. Meanwhile, for students living on the same block, peer effects are similar, regardless of cohort. These results are robust to a regression discontinuity design focusing on students near the birthdate cutoff for entry into kindergarten. I also find evidence that peers in the older cohort matter more than peers in the younger cohort, particularly in the school context, suggesting that relative age also plays a role in determining peer influence.

JEL classification: J24, I21, R23

This research result used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information and opinions of the author(s) and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof.

1.2 Introduction

When it comes to improving the economic mobility of children, it is hard to find stronger tools than the schools they attend and the neighborhoods they grow up in. Research on factors that relate to intergenerational and economic mobility finds that both schools and neighborhoods likely play a strong role in explaining how individuals and families improve their socioeconomic well-being and transmit that well-being to their children (Black and Devereux, 2010). On a more micro level, the causal case for schools and education as sources of economic mobility has been repeatedly made, with evidence finding that changes in schools (Deming et al., 2014), teachers (Chetty et al., 2014), and school resources (Jackson et al., 2016) have substantial impacts on earnings and employment in adulthood. However, this education literature often focuses solely on the effects of schools, neglecting neighborhoods as another important pillar of the economic mobility, education, and long-term welfare we are ultimately trying to foster.

School and neighborhood effects on children are deeply intertwined, particularly in the US. Although this is often understood to be because living in certain neighborhoods lends access to higher quality schools (such as in Laliberté (2018)), improvements in education outcomes from growing up in better neighborhoods are observed even in cases where there are no observable increases in school quality (Chyn, 2018). While schools themselves are certainly important for a child's education, it should not be surprising that neighborhoods influence a child's growth and educational development through channels other than schools themselves, as students spend only part of their lives at school. Indeed, we can conceive of neighborhood effects in general, and on education specifically, as a bundle of different effects, with local school quality as only one component of that bundle. Lower-poverty neighborhoods can mean less crime (Oreopoulos, 2003), better health (Sanbonmatsu et al., 2011), higher parental earnings (Baum-Snow et al., 2019), and better quality peers (Agostinelli et al., 2020), all of which have evidence supporting their positive impacts on education in their own right. The purpose of this study is to hone in on one particular component of this bundle of aspects that a child's neighborhood provides to their education and bring it into the school context: peer effects. That is, how do educational effects of the peers a child

is exposed to by living in their neighborhood influence and compare to the effects of the peers a child is exposed to in school?

While the school peer effects literature is broad (see Sacerdote (2014) for a thorough review), only a handful of papers in the economics literature have estimated the causal effects of neighborhood peers. In the most related paper to this one, Agostinelli (2018) uses variation in cohort ability levels within schools to estimate that a one percentile rank increase in peer ability at age 16 increases own ability by 0.63 percentile ranks. Fernández (2021) finds that having neighborhood peers enroll in a university increases the likelihood that students will enroll in university themselves. And finally, List et al. (2020) finds that nearby neighbors of children randomly assigned to a high-quality pre-K program also experience positive spillovers from this intervention. Collectively, these papers show that neighborhood peers do influence students' education in numerous impactful ways. Building on that work, this paper is the first in the economics literature, to the best of my knowledge, to estimate neighborhood peer effects jointly with school peer effects and examine their interaction with peer cohorts.

One difficulty with both peer effects and neighborhood effects research is the selection problem. That is, peer groups and neighborhoods both form endogenously, with agents selecting into neighborhoods and peer groups based on the characteristics of other agents. Research such as Heckman and Landersø (2021) shows how failing to account for this endogenous sorting can bias estimates of neighborhood effects. School peer effects papers, such as Carrell et al. (2013) and Imberman et al. (2012), often rely on experimental or quasi-experimental variation in classroom peer assignment to identify their estimates. To overcome this selection problem, my research design combines two approaches: 1) controlling for the abilities of students within the same school or block but adjacent cohorts and 2) instrumenting for actual school cohorts with assigned cohorts based on birthdate. The identifying assumption of this approach is that selection into schools and neighborhoods is not specific to the characteristic of peers in each cohort, but rather occurs on a broader set of neighborhood, school, and peer characteristics. In addition, the use of assigned cohorts based on birthdate eliminates the possibility of endogenous cohort selection itself, such

as through academic redshirting or grade retention. To complement this method, I also test the robustness of my main results by running a quasi-border discontinuity design with a bandwidth of one month around the cohort birthdate cutoff for kindergarten entry. This approach yields similar results and validates that my main model adequately deals with selection.

Applying this identification strategy to data on all public school students in the state of Michigan between the 2007-2008 and 2019-2020 school years, I show that 1) school peer effects are substantially stronger than neighborhood peer effects, and 2) the cohorts of one's peers play a strong role in influencing peers' effects on each other in the school context but a much weaker role in the neighborhood context. I estimate that own-cohort school peers increase students' test scores by 0.3 standard deviations for a one standard deviation increase in average peer ability, similar to school peer effect sizes found in previous literature (Sacerdote, 2014). The effect drops off by 0.15 standard deviations for school peers in the cohort above and 0.2 standard deviations for school peers in the cohort below. In contrast, block peers increase students' test scores by 0.04 standard deviations for a one standard deviation change in peer ability. Block peer effects vary by cohort by 0.02 standard deviations at most, indicating that the school cohort does not play nearly as strong of a role among peers in the neighborhood context as in the school context. Finally, I find the effect of school peers in the cohort above is 0.05 higher than for school peers in the cohort below, providing evidence that age also plays a role in peer influence, particularly in the school context.

The remainder of this paper is structured in the following way. Section 1.3 describes my model of selection bias as well as my identification strategy to overcome these biases. Section 1.4 describes the Michigan education dataset used for this study and defines the terms used for the remainder of the paper. Section 1.5 details the empirical methodology I use applying my identification strategy to obtain my estimates. Section 1.6 details my results. Section 1.7 runs through numerous robustness checks to my main empirical strategy. Section 1.8 details several heterogeneity analyses, and Section 1.9 concludes.

1.3 Bias and Identification Strategy

In this section, I lay out the potential sources of bias that may confound estimating block-cohort or school-cohort peer effects without random or quasi-random peer group assignment. For each source of bias, I also describe how parts of my identification strategy deal with this source of bias, and what assumptions need to be true to eliminate any remaining bias. For simplicity, I will be referring to school peers for the remainder of this section, but all of the concepts discussed generalize to block peers as well.

To formalize my model of unobserved selection, let $Y_{i,t}$ be the observed outcome of interest Y (e.g. test scores) of observation i at time t , $\bar{Y}_{-i,c,s,t}$ be the average of the same for the all other individuals in cohort c and school s at time t , following the standard linear-in-means model from the literature (Sacerdote, 2014). A simple model seeking to determine the peer effects of block-cohort peers would be:

$$\underbrace{Y_{i,t}}_{\text{Own Test Score}} = \beta_0 + \underbrace{\beta_1 \bar{Y}_{-i,c,s,t-1}}_{\text{Own Cohort Avg Test Score}} + v_{i,t} \quad (1.1)$$

Where $t - 1$ peer scores in place of t are used to mitigate the reflection problem, denoted as $\bar{Y}_{-i,c,s,t-1}$.¹ The issue of the reflection problem is described in further detail in Appendix Section D. My first potential source of bias in the equation above is selection into schools (or blocks). Let $\alpha_{s,t}$ be an unobserved time-varying school-specific (or block-specific) component that is correlated with both $Y_{i,t}$ and $\bar{Y}_{-i,c,s,t-1}$ in $v_{i,t}$. Because my main specifications include school (and block group) fixed effects, only time-varying unobserved effects are potentially confounding. Let $v_{i,t} = \eta_{i,t} + \alpha_{s,t}$. Rewriting equation 1.1, we have:

$$\underbrace{Y_{i,t}}_{\text{Own Test Score}} = \beta_0 + \underbrace{\beta_1 \bar{Y}_{-i,c,s,t-1}}_{\text{Own Cohort Avg Test Score}} + \underbrace{\alpha_{s,t}}_{\text{Unobserved Time-Varying School Effect}} + \eta_{i,t} \quad (1.2)$$

¹Results are also robust to alternative methods, such as using peer scores lagged by two years or from third grade.

$$\alpha_{s,t} = \gamma_0 + \underbrace{\gamma_1 \bar{Y}_{-i,c,s,t-1}}_{\text{Confounding Correlation}} + \nu_{s,t}, \quad \gamma_1 \neq 0 \quad (1.3)$$

Which shows that β_1 will be biased by $\alpha_{s,t}$.² In the selection story, $\alpha_{s,t}$ is the tendency of families to select into improving schools that have increasingly higher-ability students over time or to sort into similarly improving neighborhoods, as shown in Heckman and Landersø (2021). That is, $\alpha_{s,t}$ is the set of unobserved family characteristics that influenced observation i to live in school s with the other $-i$ peers that also affect student i 's outcomes.

Many papers in the peer effects literature attempt to overcome this unobserved selection problem with random or quasi-random variation in peer groups (Sacerdote, 2014). My solution to this selection problem, in the absence of random assignment to exploit, is to control for the outcomes of other school peers in adjacent cohorts:

$$\begin{array}{c} \text{Cohort Below Avg Test Score} \\ \overline{Y}_{c-1,s,t-1} \end{array}, \quad \begin{array}{c} \overline{Y}_{c+1,s,t-1} \\ \text{Cohort Above Avg Test Score} \end{array}$$

$$\alpha_{s,t} = \gamma_0 + \gamma_1 \bar{Y}_{-i,c,s,t-1} + \gamma_2 \overline{Y}_{c-1,s,t-1} + \gamma_3 \overline{Y}_{c+1,s,t-1} + \nu_{s,t} \quad (1.4)$$

$$Y_{i,t} = \beta_0 + \beta_1 \bar{Y}_{-i,c,s,t-1} + \beta_2 \overline{Y}_{c-1,s,t-1} + \beta_3 \overline{Y}_{c+1,s,t-1} + \alpha_{s,t} + \eta_{i,t} \quad (1.5)$$

Where $\overline{Y}_{c-1,s,t-1}$ and $\overline{Y}_{c+1,s,t-1}$ are average outcomes for peers in the cohort below and cohort above student i but still in school s , measured in $t - 1$. As equation 1.5 shows, β_1 is identified if $\gamma_1 = 0$. That is, if unobserved time-varying school (or block) effects only correlate with adjacent cohorts simultaneously such that no more correlation remains own cohort, then we interpret the own-cohort school (or block) effects as causal. Even if $\gamma_1 \neq 0$ in equation 1.4, the inclusion of $\overline{Y}_{c-1,s,t-1}$ and $\overline{Y}_{c+1,s,t-1}$ likely reduce the magnitude of γ_1 as long as the time-varying school selection occurs across cohorts over time rather than sharply diverging between adjacent cohorts.

²Substituting in equation 1.3 for $\alpha_{s,t}$ in equation 1.2 shows that the parameter estimated on $\bar{Y}_{-i,c,s,t-1}$ will be $\beta_1 + \gamma_1$.

We can also further relax the identification assumption of $\gamma_1 = 0$ in equation 1.4 by looking at relative effects between cohorts. To identify $\beta_1 - \beta_2$ and $\beta_1 - \beta_3$, the identification assumption instead becomes $\gamma_1 = \gamma_2 = \gamma_3$. That is, the relative peer effects between adjacent cohorts are identified if the remaining endogenous correlation is the same or similar magnitude. Like the argument over the reduction of the magnitude of γ_1 in the paragraph above, this identification assumption for relative cohort effects holds if time-varying selection into block or school does not vary sharply year-to-year. If parents are changing their degree of selection into schools or blocks over time, particularly in a way that correlates with peer ability (e.g. parents observe a school has started having higher ability students more recently and want to send their kids there), this assumes that this affect is only observed at the level of several cohorts, rather than based on the abilities of a single cohort. In my robustness check that uses a quasi-border discontinuity design that restricts the set of instrumenting peers to those born within one month of the school cutoff, I take this restriction further by lowering the likelihood that parents would be able to select into a block based on observed age differences ex-ante. Section 1.7.1 details this identification strategy further and shows the robustness of my main specification to this weaker identification assumption.

However, selection may not occur only at the block or school level. If families sort into cohorts based on the abilities of children in those cohorts,³ then the estimates in equation 1.5 will still be biased. Let $\eta_{i,t} = \epsilon_{i,t} + \alpha_{c,t}$, where $\alpha_{c,t}$ represents time-varying selection into cohort c that correlates with both $\bar{Y}_{-i,c,s,t-1}$ and $Y_{i,t}$. Equation 1.5 would then look like:

$$Y_{i,t} = \beta_0 + \beta_1 Y_{-i,c,s,t-1} \quad \overbrace{+\alpha_{c,t}}^{\text{Unobserved Time-Varying Cohort Effect}} \quad + \beta_2 Y_{c-1,s,t-1} + \alpha_{c-1,t} \\ + \beta_3 Y_{c+1,s,t-1} + \alpha_{c+1,t} + \alpha_{s,t} + \epsilon_{i,t}$$

This final issue is where the use of cohort entry dates comes in. By using only the abilities of peers assigned to their cohort based on month and year of birth in my main specification, α_c , selection into cohort, is eliminated.

³E.g. through policies such as redshirting or grade retention/promotion

1.4 Data

Before describing my empirical strategy in detail, I will describe my dataset and terms for this analysis to provide context to my methodology. The data for this analysis comes from the Michigan Education Data Center (MEDC).⁴ It contains data on the universe of public school students in Michigan between the 2007-2008 and 2019-2020 school years, comprising tens of millions of observations over those thirteen years. The dataset I created for this analysis comes from three source datasets: student enrollment data, which contains variables on student characteristics (birth date, race, gender, poverty status, IEP status, etc.), student assessment data, which contains test scores, and student geocode information, which contains student census block. All of these datasets are at the student-year level.

A student is considered to be living in the same neighborhood as another if they share the same census block. Although a census block is not a perfect match for a true city block in all cases, it is the smallest geographic unit available. Figure C.1 contains an example map of downtown East Lansing, near the Michigan State University Economics Department. The entire map in red is one census tract, and each small square is both one city block and one census block. One issue with the use of census blocks as neighborhoods is they extend beyond one city block in more sparsely populated areas. Because of this, I exclude any student living in a census block classified in the 2010 census as rural.

While dropping students in rural census blocks is one of my sample restrictions, I also make several others before running my OLS regressions. First, I drop students not in grades 5 through 8, since test scores are first produced in grade 3, peer scores are calculated in the previous year, and students in the cohort below grade 4 do not have lagged test scores. Next, I drop any observations missing control or outcome variables, either for themselves or peers, including demographics (race, gender, special education status), month and year of birth, and math and reading test scores. After that, I drop students who have zero students in their own or adjacent cohort in school or block. This

⁴I owe thanks to the Michigan Education Research Institute (MERI) for guiding me through the application process and making this data available. This data source is not identical to the data maintained by the Michigan Department of Education (MDE) and Michigan's Center for Educational Performance and Information (CEPI).

leaves 2,999,834 observations in my OLS math sample and 2,983,697 observations in my OLS reading sample.

To get from my OLS analysis sample to my main specification sample, which instruments for actual cohorts with assigned cohorts, I make two additional sample restrictions. First, I drop all students who do not have at least one student assigned to own or adjacent cohorts in either school or block.⁵ Second, I drop all students who are missing average lagged test score data or peer demographic data for any of the own or adjacent assigned cohorts in school or block. These two changes leave a final assigned cohort sample of 2,666,141 observations for math and 2,656,132 for reading.

In Tables B.1 and C.1, I compare my assigned cohort sample to the full sample of Michigan public school students in grades 5 through 8. Looking at Table B.1, the most striking difference between the two samples is the urbanicity of school attended, where students are 12.3 percentage points more likely to attend a suburban school and 13.1 percentage points less likely to attend a rural school. This is mostly due to dropping students in rural census blocks. Additionally, sample students are less likely to be white (4.7 percentage points) and are more likely to be Black (3.9 percentage points). Michigan's rural population is overwhelmingly white, so this likely reflects the rural/urban urbanicity sample restriction. Finally, students in the analysis sample live in slightly denser neighborhoods and attend slightly larger schools, with 36 more students per cohort in school (22%) and 0.3 more students per cohort in the neighborhood (19%). In Table C.1, I again compare my assigned cohort analysis and full samples, but this time using ACS 5-year data on block groups. Table C.1 confirms the racial composition changes shown in Table B.1, while also showing that my sample census blocks are more populous (56 more residents) and have higher household incomes (\$2,375 higher median household income on average). In total, the sample is more urban, slightly more advantaged and higher performing, and denser than the full Michigan grade 5-8 public school population.

⁵Counts of students assigned to school-cohorts are set to zero if there are no students in that actual school-cohort. For example, if a school is K-5 and the student of interest is in grade 5, they will have no cohort-above-assigned students in grade 6, even if there are students in their cohort who should be in grade 6 based on their birthday.

To help illustrate the motivation for the use of assigned cohort as an instrument, Figure A.1 shows compliance with cohort assignment in my analysis sample, both overall and by birth month. As the figure shows, almost all cohort non-compliance comes from students in a lower cohort than assigned by the statewide cutoff and their birthdate. 17% of students are in a lower cohort than assigned, while only 0.01% are in a higher cohort. The second panel of Figure A.1 separates the first panel's histogram by birth month and shows there is considerable heterogeneity throughout the year. Notably, the tendency to attend a lower cohort than assigned increases throughout the year, as students become younger relative to their assigned classmates. These two trends in assigned cohort compliance are likely a combination of three factors 1) academic redshirting (holding children back from entering kindergarten at age 5), 2) grade retention, and 3) individual districts' (unobserved) earlier entry cutoffs. Factors 1 and 2 are the most pressing for the case of endogeneity since they are more individual decisions. Factor 3 may also be endogenous, depending on the decision-making process that districts engage in. Regardless, as long as being in a district with an earlier cutoff data does not *increase* the likelihood of staying in a cohort the student is no longer assigned to, the IV estimation using the assigned cohort should still have a tractable and valid interpretation while dealing with remaining potential sources of endogeneity, such as grade retention/promotion and academic redshirting.

1.5 Empirical Strategy

To lay out my empirical strategy, I start by detailing my naive OLS regression, followed by my instrumental variables specification. The basic structure of my dependent variable and independent variables of interest is:

1.5.1 Naive OLS

$$\begin{aligned}
 \underbrace{Y_{i,t}}_{\text{Own Test Score}} &= \beta_0 + \underbrace{\beta_1 \bar{Y}_{-i,c,s,t-1}}_{\text{Own cohort, school}} + \underbrace{\beta_2 \bar{Y}_{c-1,s,t-1}}_{\text{Cohort below, school}} + \underbrace{\beta_3 \bar{Y}_{c+1,s,t-1}}_{\text{Cohort above, school}} \\
 &\quad + \underbrace{\beta_4 \bar{Y}_{-i,c,b,t-1}}_{\text{Own cohort, block}} + \underbrace{\beta_5 \bar{Y}_{c-1,b,t-1}}_{\text{Cohort below, block}} + \underbrace{\beta_6 \bar{Y}_{c+1,b,t-1}}_{\text{Cohort above, block}}
 \end{aligned}$$

Where $Y_{i,t}$ is outcome Y of student i in school year t . Each \bar{Y} is the mean standardized test score of peers in each relative school-by-cohort and block-by-cohort combination.⁶ $\bar{Y}_{-i,c,s,t-1}$ is the mean test score Y of all other ($-i$) students in cohort c and school s in year t , measured in $t-1$, and $\bar{Y}_{-i,c-1,s,t-1}$ and $\bar{Y}_{-i,c+1,s,t-1}$ are the mean test scores Y of all other students in cohorts $c-1$ and $c+1$ and school s . Each $\bar{Y}_{-i,c+j,b,t-1}$ is the same but for peers within census block b . The notable difference from equation 1.5 in Section 1.3 is that both block and school peer effects are estimated simultaneously. This has two benefits: 1) the resulting estimates show the value-added of peers in each block or school context, reflecting that both sources of skill formation occur together, rather than in isolation, and 2) estimating simultaneously further enhances my identification strategy by controlling for how selection into block may affect school peer estimates, and vice versa.

Adding in my control variables and fixed effects (and combining terms into summations), my full naive OLS equation is:

$$\begin{aligned}
 Y_{i,t} = & \beta_0 + \sum_{(j=-1)}^1 \beta_j \bar{Y}_{-i,c+j,s,t-1} + \sum_{(j=-1)}^1 \alpha_j \bar{Y}_{-i,c+j,b,t-1} \\
 & + \delta_1 X_{i,t} + \sum_{(k \in s,b)} \sum_{(j=-1)}^1 \delta_{j,k} \bar{X}_{-i,c+j,k,t} + \Gamma_{m,c,g,s,b_g,t} + \epsilon_{i,t}
 \end{aligned}$$

⁶When there are zero students in any of the groups, the mean is instead replaced with zero. To prevent this from causing the regression to conflate having zero peers with having average-ability peers, a dummy variable is included in all regressions for each peer group category that is equal to one if there are no students in each respective peer group.

All X variables are controls for individual, relative school-cohort, and relative block-cohort characteristics. $X_{i,t}$ is a vector of individual characteristics including race/ethnicity, gender, and special education status. Each $\bar{X}_{-i,c+j,k,t}$ is a vector of means of controls for the same characteristics, as well as number of students in group and average age of students in group, among the other peers ($-i$) in relative cohorts $c - 1$, c , and $c + 1$ and school s and block b . Finally, $\Gamma_{m,c,g,s,b_g,t}$ is a matrix of fixed effects for month of birth, cohort (year of birth), cohort, school, census block group, and school year, respectively.

As discussed in Section 1.3, this estimation strategy controls for any endogenous variation common between adjacent peer cohort groups. With the addition of school and block group fixed effects, any unchanging endogenous correlation between own ability and peer abilities should be taken care of as well. However, as also discussed in Section 1.3, this estimation strategy is vulnerable to sorting into cohorts based on peer abilities (or other factors correlated with both own and peer abilities, such as income and education levels).

1.5.2 Assigned Cohort IV

1.5.2.1 Reduced Form

To address this, I use an instrumental variables strategy that instruments for school-by-cohort peer ability and block-by-cohort peer ability with the abilities of students who are assigned to own or adjacent cohorts based on their birthdate and the Michigan school cohort entry cutoff date at age 5. The reduced form equation for this IV strategy is:

$$Y_{i,t} = \beta_0 \sum_{(j=-1)}^1 \beta_j \bar{Y}_{-i,ac+j,s,t-1} + \sum_{(j=-1)}^1 \alpha_j \bar{Y}_{-i,ac+j,b,t-1} \\ + \delta_1 X_{i,t} + \sum_{(r \in c,ac)} \sum_{(k \in s,b)} \sum_{(j=-1)}^1 \delta_{j,k,r} \bar{X}_{-i,r+j,k,t} + \Gamma_{m,c,g,s,b_g,t} + \epsilon_{i,t}$$

The main difference between the reduced form equation and the naive equation is the use of assigned cohort peer groups instead of actual cohort peer groups. Each \bar{Y} , X , and γ represent the same concepts as in the previous equation, mean peer group ability, controls for average peer group

characteristics, and fixed effects, but now \bar{Y} and \bar{X} are indexed by ac , which I am using to represent assigned cohort. The \bar{X} 's indexed by ac are the proportions of observable characteristics and the number of students for each assigned cohort, in addition to the controls for each actual cohort. For each Y and \bar{X} in each school ac , the value is set to zero if there are no members in the actual cohort at school. For example, if a school is K-5 and student i is in grade 5, there will be no assigned students in cohort $c + 1$, even if there are students in their cohort who should be in grade 6 based on their birthday. Instead, those students are not included in any of the ac measures and all $ac + 1$ values are set to zero, and an indicator variable for zero students in $c + 1$ is set to one.

1.5.2.2 Instrumental Variables

The first stage equations mirror the reduced form equation on the right-hand-side, but now have $\bar{Y}_{-i,c,s,t-1}$, $\bar{Y}_{c-1,s,t-1}$, $\bar{Y}_{c+1,s,t-1}$, $\bar{Y}_{-i,c,b,t-1}$, $\bar{Y}_{c-1,b,t-1}$, and $\bar{Y}_{c+1,b,t-1}$ on the left-hand-side instead as follows:

$$\begin{aligned} \bar{Y}_{-i,c,s,t-1} = & \beta_0 + \sum_{(j=-1)}^1 \beta_j \bar{Y}_{-i,ac+j,s,t-1} + \sum_{(j=-1)}^1 \alpha_j \bar{Y}_{-i,ac+j,b,t-1} \\ & + \delta_1 X_{i,t} + \sum_{(r \in c, ac)} \sum_{(k \in s, b)} \sum_{(j=-1)}^1 \delta_{j,k,r} \bar{X}_{-i,r+j,k,t} + \Gamma_{m,c,g,s,b_g,t} + \epsilon_{i,t} \end{aligned}$$

With the other five first stage equations having $\bar{Y}_{c-1,s,t-1}$, $\bar{Y}_{c+1,s,t-1}$, $\bar{Y}_{-i,c,b,t-1}$, $\bar{Y}_{c-1,b,t-1}$, and $\bar{Y}_{c+1,b,t-1}$ on the left-hand-side instead of $\bar{Y}_{-i,c,s,t-1}$. The second stage equation combines the naive OLS equation and the predicted values from the first stages:

$$\begin{aligned} Y_{i,t} = & \beta_0 + \sum_{(j=-1)}^1 \beta_j \widehat{\bar{Y}}_{-i,c+j,s,t-1} + \sum_{(j=-1)}^1 \alpha_j \widehat{\bar{Y}}_{-i,c+j,b,t-1} \\ & + \delta_1 X_{i,t} + \sum_{(r \in c, ac)} \sum_{(k \in s, b)} \sum_{(j=-1)}^1 \delta_{j,k,r} \bar{X}_{-i,r+j,k,t} + \Gamma_{m,c,g,s,b_g,t} + \epsilon_{i,t} \end{aligned}$$

where β_j through α_j now gives our exogenous peer effects for school and block peers.

1.6 Results

1.6.1 OLS Results

Before turning to the results of my assigned cohort IV estimation, I will briefly discuss the results of OLS estimation described in Section 1.5.1. Column 1 of Tables B.2 and B.3 display the results of the OLS estimation for math and reading scores, respectively. The top panel shows the estimated peer effects for school-by-cohort peers and the bottom panel shows the same for block-by-cohort peers. Each of the three rows within each panel shows the effects by each relative peer cohort: cohort below, $c - 1$, own cohort, c , and cohort above, $c + 1$. As with all regressions in this paper, robust standard errors are used, with clustering at the school level. Notably, the large sample size, displayed in the bottom row of the table, results in small standard errors, with almost all results in both tables significant at the 0.01 level.

For the math scores in column 1 of Table B.2, the results show peer effects of a 0.3 standard deviation increase in own test scores for a one standard deviation increase in the ability of peers in the same cohort at school. These estimates show an own-cohort school effect in line with the classroom peer effects literature (Sacerdote, 2014), a drop off of 0.2 standard deviations when compared to the cohort above, and an even larger drop in effect size of 0.25 standard deviations for the cohort below. Effect sizes for reading scores, in column 1 of Table B.3, follow a similar pattern of differences in relative school-cohort effects, with lower overall magnitudes. This latter result of varying adjacent-cohort drops in effect size suggests 1) relative cohort plays a substantial role in influencing peer effects in the school context, and 2) relative *age*, in addition to cohort, plays an important role as well.

The story for block peers is markedly different. First, the effect sizes for block peers are much smaller, ranging from a 0.04 to a 0.05 standard deviation increase in own score for a one standard deviation increase in peer score for math, and a 0.03 to 0.04 range for reading. In this OLS specification, we can also detect differences between the three adjacent cohorts among block peers, although the sizes of the differences, a maximum of 0.01, are much less substantial for block peers. Notably, and most robustly to specifications shown later in the paper, effect sizes are larger for

block-cohort peers in the cohort above, indicating the presence of a relative age effect in the block context as well.

As discussed in Section 1.3, we should expect these OLS estimates to deal with school-level and block-level bias, particularly for differences between adjacent cohort peer groups, but not address sorting into cohorts. For a cohort sorting example, if parents choose to engage in academic redshirting (delay Kindergarten entry by one year) in response to the increasing abilities over time of school or neighborhood peers, either to increase contact with high-performing peers or reduce contact with low-performing peers, then these estimates will be biased.

1.6.2 Main Results

Next, we turn to my preferred specification, which instruments for the abilities of peers in each school-by-cohort and block-by-cohort group with the abilities of peers *assigned* to each cohort group based on birthdate and cohort entry cutoff at age 5 (laid out in Sections 1.5.2). Estimates for the remainder of this section follow the same format as the OLS columns, with the addition of Kleibergen-Paap F Stats for the IV estimates.⁷

1.6.2.1 Reduced Form Results

To explore the assigned cohort IV, I first show reduced form results produced by regressing own ability on the average ability of students in assigned cohorts at school and in the block. Column 2 of Tables B.2 and B.3 show the results of this estimation, using the equation shown in Section 1.5.2.1. Importantly, all estimates in both tables are highly significant ($p < 0.01$), indicating that the weak instrument problem is likely not a concern. As a second precaution against weak instruments, all results will be reported with a Kleibergen-Paap Wald F statistic. Overall, estimates follow a similar pattern to the OLS estimates, especially for block-cohort peers.⁸

⁷Similar to a Cragg-Donald test for weak instruments with multiple endogenous variables, as proposed by Stock et al. (2002), a Kleibergen-Paap Wald test uses an appropriate cluster-robust degrees of freedom. See Andrews et al. (2019) for further discussion.

⁸However, there are some differences for school peers: lower magnitudes for their own cohort and higher magnitudes for adjacent cohorts. The first potential explanation for this is that most assigned cohort noncompliance is into the cohort below, not above, as shown in Figure A.1. This means that using assigned cohort peers is, for the most part, reassigning some peers to the cohort above their actual cohort. Because the OLS estimates suggest that actual own cohort peers have the strongest influence, assigned cohort own effects are weakened by adding actual below cohort peers and subtracting actual own cohort peers to the assigned own cohort peer pool. In turn, above effects are strengthened

1.6.2.2 IV Results

The Assigned Cohort IV results, shown in Figure A.2, with statistics listed in column 3 of Tables B.2 and B.3, are very similar to the OLS estimates. Across both math and reading scores and peer and block groups, assigned cohort point estimates are slightly larger, although not always to a statistically detectable ($p < 0.05$) degree. Math own-cohort peer effects range from 0.33 standard deviations for own cohort, with a 0.16 standard deviations decrease in effect size for the cohort above, and a 0.26 standard deviations decrease in effect size for the cohort below. The own-cohort school findings continue to match classroom peer sizes effects found in the literature, with the novel finding of substantial drop-offs in peer effects for students not in the same cohort combined with a somewhat-countervailing age-influence effect. My additional novel finding of substantially lower block peer effects holds as well. Math block-by-cohort effects range from 0.03 for own cohort to 0.05 for cohort below and 0.06 for cohort above, still showing little variation by cohort with a slight age-influence effect.

1.6.3 Summary

Across the two specifications, OLS and assigned cohort IV, there are four main takeaways: (1) peer effects vary substantially by relative cohort in the school context but not the block context, (2) math peer effects are stronger than reading peer effects, (3) own cohort school peer effects are substantially larger than block peer effects, about 0.3 standard deviations for a one standard deviation increase in peer ability and block-cohort peer effects are 0.04 standard deviations regardless of the relative cohort, and (4) cohort above school peer effects are higher than both cohort below and all block-cohort peer effects, at about 0.15 standard deviations. This own-cohort school peer effect is about the same effect size found in the quasi-experimental classroom peer effects literature of

by adding actual own cohort peers to the assigned above-cohort pool. Thus, relative cohort effects shown in the OLS estimation explain two of the three changes in school-cohort magnitude.

To help answer why assigned cohort-below peer effects are stronger than actual cohort below peer effects, we need an additional explanation. One potential cause of this is that cohort sorting is endogenous: lower-ability peers may be more likely to either be held back or enter kindergarten a year later. This effect would be offset for the assigned cohort above peers because of the stronger cohort exposure effect (the potentially lower ability peers are actually in own cohort), but not for the cohort below, which is receiving peers from the use of assigned cohorts who are two actual cohorts below, who likely have an even weaker cohort exposure effect. In short, using assigned cohort peers means also potentially reassigning endogenously cohort-switching peers, which can, as a result, boost peer effects.

about 0.3 standard deviations (Sacerdote, 2014). In contrast, this effect is much smaller than the effect of academic redshirting, which is about 0.7 standard deviations (Elder and Lubotsky, 2009; Cascio and Schanzenbach, 2016), though like academic redshirting and other education findings, the effect is stronger for math scores than reading scores.

1.7 Robustness Checks

Although all four conclusions⁹ are consistent across both OLS and assigned cohort IV specifications, there are still several specification changes that the results could be sensitive to. First, I will introduce my main robustness check, a birth cutoff discontinuity IV, explain the motivations, and show the results, which are similar to the main specification. Next, I will also show that results are robust to a handful of other alternative specification changes, including the inclusion of extra controls for low-income and English language learner status, the use of peer abilities measured in 3rd grade instead of $t - 1$, the addition of own ability in 3rd grade as an extra control, and dropping a school year where Michigan changed its standardized testing scheme.

1.7.1 Birth Cutoff Discontinuity IV

1.7.1.1 Cutoff Student IV Estimation

For this specification check, I restrict the set of peers used to instrument for school-cohort and block-cohort ability to peers born within one month of the cohort entry cutoff date when they are age five. This relaxes the assumption made for the assigned cohort IV that parents do not select into schools and blocks based on peer age as long as the peers are within several cohorts of each other. Now, the assumption is that parents may select based on the characteristics of peers in the same cohort, but because age is used by parents as a proxy for cohort, are unlikely to be able to distinguish the cohort of peers born within one month of each other. Appendix Section E goes into further detail about Michigan's cutoff policy, statistics for the cutoff groups, and additional sample restrictions used for this robustness check.

When using birth cutoffs, we have a subset of neighborhood peers who are plausibly otherwise

⁹Peer effect variation by cohort within school but not block, stronger math score effects, stronger school effects than block effects, and stronger cohort-above effects than cohort-below effects.

identical but differentiated only based on whether they are in the same cohort as the student of interest. To help illustrate, Figure A.3 shows a student in a census block with six other children of similar age, who are then sorted into cohorts at school by their birthdates.¹⁰ As Figure A.3 shows if the student of interest, i , is born in January 2005 and the cohort entry cutoff is December 1st the sorted peers are then separated into three separate groups when they attend school: the cohort below the student of interest, the same cohort as the student of interest, and the cohort above. The four neighborhood peers born in November and December 2004, as well as November and December 2005, are the two groups of cutoff peers in this example: plausibly similar in most characteristics except for their interaction with the student of interest at school as part of membership in the same cohort. This exercise can also be repeated for any different combination of neighborhood peers where at least one cutoff student in any of the four cutoff months¹¹ and at least one other non-cutoff student is present in the neighborhood within one cohort of the cutoff student.

The estimation strategy for the cutoff student IV is similar to a fuzzy border discontinuity. Figure A.4 helps illustrate the connection between the first stage equations above and the identification strategy and Figure A.3. The first two terms in the equation below use the average ability of cutoff school peers born just before or after the cutoff for being the oldest peers in student i 's cohort. The second two terms do the same, but with the average abilities of cutoff school peers born just before or after the cutoff for being the *youngest* peers in student i 's cohort. The next four terms repeat the process, but now for block cutoff peers, rather than school. Adding in controls for counts of students and proportions of observable characteristics in each cutoff group, the first stage equation looks like this:

¹⁰This example is an unusually population-dense block. The majority of students in the sample have only one cutoff student in their block-cohort or block-and-adjacent-cohort

¹¹November of cohort above, December of the same cohort, November of the same cohort, and December of cohort below.

$$\begin{aligned}
\bar{Y}_{-i,c,s,t-1} = & \beta_0 + \beta_1 \bar{Y}_{Dec,c,s,t-1} + \beta_2 \bar{Y}_{Nov,c+1,s,t-1} + \beta_3 \bar{Y}_{Nov,c,s,t-1} + \beta_4 \bar{Y}_{Dec,c-1,s,t-1} \\
& + \alpha_1 \bar{Y}_{Dec,c,b,t-1} + \alpha_2 \bar{Y}_{Nov,c+1,b,t-1} + \alpha_3 \bar{Y}_{Nov,c,b,t-1} + \alpha_4 \bar{Y}_{Dec,c-1,b,t-1} \\
& + \delta_1 X_{i,t} + \sum_{(m \in Nov, Dec)} \sum_{(k \in s, b)} \sum_{(j=-1)}^1 \delta_{j,k,m} \bar{X}_{-i,m,c+j,k,t} + \Gamma_{m,c,g,s,b_g,t} + \epsilon_{i,t}
\end{aligned}$$

With the other five first stage equations having $\bar{Y}_{c-1,s,t-1}$, $\bar{Y}_{c+1,s,t-1}$, $\bar{Y}_{-i,c,b,t-1}$, $\bar{Y}_{c-1,b,t-1}$, and $\bar{Y}_{c+1,b,t-1}$ on the left-hand-side instead of $\bar{Y}_{-i,c,s,t-1}$.

1.7.1.2 Cutoff IV Results

The results of this alternative estimation strategy are displayed in column 4 of Tables B.2 and B.3. These tables use the equations in Section 1.5.2.2 to estimate peer effects by instrumenting for actual school-by-cohort and block-by-cohort abilities with those of students born on either side of their cohort entry cutoff date when they entered kindergarten. Joint Kleibergen-Paap F Stats in column 4 of Tables B.2 and B.3 are 126 and 155, respectively, suggesting that weak identification is not biasing these results. Standard errors are slightly higher for school peer effects and up to 8 times larger for block peer effects. Although there is some loss of statistical power, there is still enough to validate the assigned cohort IV findings. For each peer group, school and block, in each panel, I also show a row for "Joint Test Pval". This is the p -value from an F test of whether all three group-cohort estimates are jointly zero.

Although the Cutoff IV point estimates are smaller across the board than the assigned cohort IV, the overarching story remains the same. For math, school-cohort peer effects for own cohort are now 0.25 (instead of 0.33), 0.07 for cohort above, and 0.003 for cohort below (cohort below not statistically significant at the 0.1 level). Math block-cohort peer effects are 0.02 for own cohort, 0.03 for the cohort above, and 0.01 for the cohort below, jointly significant at the 0.01 level. As in previous estimates, reading results are lower in magnitude. These results suggest that, while the assigned cohort specification may slightly overestimate the specific school peer effects point estimates, the main story of own-cohort estimates close to the literature, substantial effect size drop for adjacent cohorts, larger school peer effects than block peer effects, and higher effects for

cohort-above than cohort-below peers still hold. The results of this robustness check validate that, although there may be a small degree of selection bias in my main estimates, the four key takeaways still largely hold even under weaker identification assumptions on sorting.

1.7.2 Other Checks

Next, in Tables B.4 and B.5, I examine robustness to four different alternative specifications: including controls for low-income and English learner statuses, using grade 3 scores for peer ability, including a control for own ability in 3rd grade, and dropping the year of test type transition in Michigan, respectively. Unlike previous tables, the effects for each relative cohort are now in descending rows, instead of columns. Each column now represents each of the four robustness checks. All four regressions have similarly high Kleibergen-Paap F statistics, indicating that, even with lower sample sizes (and correspondingly larger standard errors), the first stages are still strong enough to avoid weak instrument bias.

The first column of Tables B.4 and B.5, "All Controls", includes additional controls for low-income and English learner statuses that are excluded from the main analysis due to lack of availability in the first five years of the data. Controls are added for student i 's own low-income and English learner status, the proportions of peers in cohort below, own, and cohort above school and block peer groups with low-income and English learner statuses, and the same proportions of peers in own and adjacent assigned cohorts. Point estimates are somewhat lower across the board, although the differences between individual coefficients and the main assigned cohort IV cannot be rejected at the 5% level. Although there is some loss in precision, the relatively small change in point estimates for the inclusion of additional controls suggests that not controlling for low-income and English learner statuses has a minimal effect on the estimates, if anything, and does not overall change the interpretation of the four takeaways.

"G3 Reflection Adj", the second column of Tables B.4 and B.5, measures peer ability using peer scores in grade 3, instead of in the year before, $t - 1$. As I discuss further in Appendix Section D, the reflection problem arises because peers affect each other's scores simultaneously. Although using peer scores measured $t - 1$ eliminates reflection bias from the present year, year

t , it is possible some reflection bias remains.¹² "G3 Reflection Adj" shows that most of the main conclusions of my preferred specification remain intact, with potentially lower estimates for own cohort. The conclusion that is not robust to this specification is above-cohort school peers' stronger effect size than below-cohort peers. However, this is likely a mechanical result of the robustness check itself: peers in the cohort above have third-grade scores further back in time as a direct result of being older, which mechanically reduces their 3rd-grade abilities' correlation with both their contemporaneous abilities and student i 's abilities.

Tables B.4's and B.5's third column, "Own Score Control", includes a control for student i 's ability as measured in third grade. This control is included to eliminate any possible remaining individual heterogeneity from before we first observe each student in grade 3. Because students in grade 3 are not included in the analysis sample, all of the fixed effects, including school and block group, do not account for unobserved heterogeneity from grade 3 and before. That is, students could have unobserved endogenous peer group sorting that occurs before the analysis sample begins and that is not fully ameliorated by my preferred specification. Then, controlling for 3rd-grade ability takes a more value-added approach, only showing the effects of changes in peer abilities after students enter my sample. The results suggest that this control may somewhat lower own-cohort school ability effects to 0.2 standard deviations and lower all block-cohort peer effects to 0.02 standard deviations from 0.04, but otherwise keeps the main takeaways intact: a large role for cohort among school peers, small or no role among block peers, substantially larger school peer effects than block peer effects, and some role for relative age effects.

Finally, the fourth column of Tables B.4 and B.5, shows the robustness of the results to dropping the school year where Michigan changed its standardized testing scheme: 2014-2015. In the 2014-2015 school year, Michigan changed both its statewide test from the MEAP to the M-STEP and the

¹²This is because student i 's own scores from $t - 1$ have reflection bias from the peer scores in $t - 1$, so the previous year's reflection bias may still artificially inflate peer effects estimates to a smaller degree. Using peer scores from grade 3, the earliest available year, is the most robust option available in this dataset to this potential source of bias because the further back in time peer scores are sourced, the weaker the correlation of student i 's ability with their own ability from the year of the peer scores, the less reflection bias will be present. However, this benefit is also its drawback: weakening correlation with reflection bias also means weakening correlation with peer ability in the current year, which is the year we are ultimately interested in proxying for.

timing of the standardized exams from the fall to the spring. Because peer abilities are measured in $t - 1$ to reduce reflection bias, estimates using the 2014-2015 school year would use peer abilities from a different exam taken at a different time of year as the exam used to produce own scores, which may bias my estimates if test score standardization and school-year fixed effects do not fully account for this change. Results are nearly identical to the main specification, showing that changes in test schemes are not unduly influencing my results.

1.8 Heterogeneity Analysis

To explore the possibility of other heterogeneous effects, I break down my sample by several different categories and rerun the cutoff IV analysis: grade, gender, race/ethnicity, economically disadvantaged status, and 3rd-grade test score. Results are displayed in Tables C.2 to C.8 in Appendix Section C.3. For the most part, the results are consistent across groups, including gender, race, and disadvantaged status. There are some potentially suggestive differences, such as higher point estimates for white students than black and non-economically-disadvantaged students than disadvantaged ones, but these differences are largely not statistically significant at the 0.05 level.

In contrast, Tables C.2, C.3, and C.8, which break down effects by grade and own 3rd grade test scores, have some stronger evidence of heterogeneity. For Tables C.2 and C.3, there is some evidence of heterogeneity by grade, especially as own-cohort school effects diminish between grades 5 and 7. However, because results increase again for grade 8, and both grade 5 and grade 8 lack cohort-above school peer controls,¹³ the results should be interpreted with caution, as this may be driven by the structure of the estimation method instead of reflecting true underlying heterogeneity. For Table C.8's math scores, displaying effects by ability in 3rd grade, above median students have stronger peer effects in both the school and the block from their own cohort peers. This provides suggestive evidence of nonlinear peer effects, as a strict interpretation of the linear-in-means model would predict there should be no differences in effect size by own ability.

In total, these heterogeneity effects demonstrate that the four takeaways, variation in cohort peer

¹³Grade 8 almost always lacks a cohort above in the same school, which is driving its low Kleibergen-Paap F Stat. Grade 5 has some cases with K-6 schools, but still has a large enough proportion of students with no cohort-above peers in the same school that the controlling effects of cohort-above ability on own-cohort effects may be diminished

effects for school and not block, own-cohort school peer effects of 0.3, block-cohort peer effects of 0.04, and cohort-above school peer effects 0.15 lower than own-cohort effects, are not driven by one group and are consistent across the sample. Like the robustness checks, we see some fluctuation in the point estimates, but most of these changes are either statistically insignificant or insubstantial. Suggestive evidence for nonlinear peer effects and fluctuation by grade provide potential avenues for heterogeneous effects without refuting the four takeaways, though both should be interpreted with caution.

1.9 Conclusion

In this paper, I bring the neighborhood context into the school and explore the role of peers' relative cohorts in influencing educational peer effects. Controlling for peers in adjacent cohorts and instrumenting for peer cohort with cohort assignment on the universe of Michigan public school students over thirteen years, I show that the cohorts of one's peers play a strong role in influencing peers' effects on each other in the school context. While own-cohort school peer effects of a 0.3 standard deviation increase for a one standard deviation increase in average peer ability are in the ballpark of school peer effect sizes found in previous literature (Sacerdote, 2014), I add the novel finding that the effect drops off by 0.15 standard deviations for school peers in the cohort above and by 0.2 standard deviations for school peers in the cohort below.

However, not only are block peer effects substantially lower than school peer effects, this relative cohort effect does not hold for block peers, indicating that peer group formation in the school environment is fundamentally different from the neighborhood environment. Neighborhood peers have a much smaller effect than school peers on educational outcomes. Finally, I provide evidence that school peers in the cohort above have an effect about 0.05 standard deviations higher than school peers in the cohort below, with a small increase for above cohort block peers as well, suggesting that age also plays a role in influencing peer influence in the peer group formation process. In total, these four takeaways broaden the literature by combining the school and block peer contexts, uncovering the role of relative cohorts, and suggesting the importance of age effects by other peers.

Collectively, these results add significant detail to our understanding of the role of peers in the education production function. My paper shows that focusing on only one cohort and ignoring adjacent cohorts leaves out substantial parts of the peer experience. Research designs that rely on year-by-year cohort fluctuation in peer characteristics for identification may need to consider the possibility of spillovers. Additionally, bringing neighborhood peers into the school context shows that neighborhoods play a smaller but still important role in the educational process and that peer group formation in this context may operate in unexpected ways compared to schools. In total, policymakers and researchers should take away the lesson that education comes from a broad variety of peers and environments that we may not traditionally focus on. So, if we wish to maximize human capital, economic mobility, and long-term well-being, we should take more opportunities to step back and evaluate students' education with a broader and more all-encompassing lens.

BIBLIOGRAPHY

- Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments. *Unpublished Manuscript, Arizona State University*.
- Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2020). It takes a village: the economics of parenting with neighborhood and peer effects. Technical report, National Bureau of Economic Research.
- Agostinelli, F. and G. Sorrenti (2018). Money vs. time: family income, maternal labor supply, and child development. *University of Zurich, Department of Economics, Working Paper* (273).
- Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11.
- 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.
- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics* 11(3), 338–81.
- Autor, D. H., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2020). Males at the tails: How socioeconomic status shapes the gender gap. *NBER Working Paper* (w27196).
- Baker, M. and K. Milligan (2016). Boy-girl differences in parental time investments: Evidence from three countries. *Journal of Human Capital* 10(4), 399–441.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences* 6(6), 248–254.
- Baron-Cohen, S. (2003). *The Essential Difference: Men, Women, and the Extreme Male Brain*. London: Allan Lane.
- Baum-Snow, N., D. A. Hartley, and K. O. Lee (2019). The long-run effects of neighborhood change on incumbent families.
- Becker, G. S., W. H. Hubbard, and K. M. Murphy (2010). The market for college graduates and the worldwide boom in higher education of women. *American Economic Review* 100(2), 229–33.
- 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.
- Bibler, A. (2018). Household composition and gender differences in parental time investments.

Available at SSRN 3192649.

- Black, S. E. and P. J. Devereux (2010). Recent developments in intergenerational mobility. *New This Week*, 2 – 90.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics* 93(2), 455–467.
- Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* 81(3), 855–882.
- Cascio, E. U. and D. W. Schanzenbach (2016). First in the class? age and the education production function. *Education Finance and Policy* 11(3), 225–250.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *The American Economic Review* 104(9), 2633 – 2679.
- Chetty, R., N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi (2016). Childhood environment and gender gaps in adulthood. *American Economic Review* 106(5), 282–88.
- Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review* 108(10), 3028–56.
- Cornwell, C., D. B. Mustard, and J. Van Parys (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources* 48(1), 236–264.
- 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.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources* 42(3), 528–554.
- Demaray, M. K., S. L. Ruffalo, J. Carlson, R. Busse, A. E. Olson, S. M. McManus, and A. Leventhal (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review* 24(4), 648–671.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics* 132(4), 1593–1640.
- Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School choice, school quality,

- and postsecondary attainment. *The American Economic Review* 104(3), 991 – 1013.
- DiPrete, T. A. and J. L. Jennings (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research* 41(1), 1–15.
- Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, et al. (2007). School readiness and later achievement. *Developmental psychology* 43(6), 1428.
- Elder, T. E. (2010). The importance of relative standards in adhd diagnoses: evidence based on exact birth dates. *Journal of Health Economics* 29(5), 641–656.
- Elder, T. E. and D. H. Lubotsky (2009). Kindergarten entrance age and children’s achievement impacts of state policies, family background, and peers. *Journal of Human Resources* 44(3), 641–683.
- Elliott, S. N., F. M. Gresham, T. Freeman, and G. McCloskey (1988). Teacher and observer ratings of children’s social skills: Validation of the social skills rating scales. *Journal of Psychoeducational Assessment* 6(2), 152–161.
- Fernández, A. B. (2021). Neighbors’ effects on university enrollment. *American Economic Journal: Applied Economics* (forthcoming).
- Fortin, N. M., P. Oreopoulos, and S. Phipps (2015). Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources* 50(3), 549–579.
- Gensowski, M., R. Landersø, B. Dorte, P. Dale, A. Højen, and L. Justice (2020). Public and parental investments and children’s skill formation. *The ROCKWOOL Foundation Research Unit* (155).
- Goldin, C., L. F. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives* 20(4), 133–156.
- Heckman, J. J. and R. Landersø (2021). Lessons from denmark about inequality and social mobility. Technical report, National Bureau of Economic Research.
- Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics* 6(1), 689–733.
- 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.
- Imberman, S. A., A. D. Kugler, and B. I. Sacerdote (2012). Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review* 102(5), 2048–82.

- Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes : Evidence from school finance reforms. *The Quarterly Journal of Economics* 131(1), 157 – 218.
- Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review* 21(6), 589–598.
- Johann, A. (2020). The increasing fragility of boys: Examining changes in levels and correlates of gender gaps in noncognitive skills over time. *Available at <https://sites.google.com/site/alwjohann/working-papers/gender-gaps-in-noncognitive-skills>*.
- Kling, J. R., J. Ludwig, and L. F. Katz (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics* 120(1), 87–130.
- Knickmeyer, R., S. Baron-Cohen, P. Raggatt, and K. Taylor (2005). Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology and Psychiatry* 46(2), 198–210.
- Laliberté, J.-W. P. (2018). Long-term contextual effects in education: Schools and neighborhoods. *University of Calgary, unpublished manuscript*.
- Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics* 3(1), 101–28.
- List, J. A., F. Momeni, and Y. Zenou (2020). The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods. Technical report, National Bureau of Economic Research.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3), 531–542.
- Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics* 92(3), 562–576.
- Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The quarterly journal of economics* 118(4), 1533–1575.
- Raver, C., P. W. Garner, and R. Smith-Donald (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In *School readiness and the transition to kindergarten in the era of accountability*, pp. 121–147. Paul H Brookes Publishing.
- Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics* 6(1), 253–272.

- Sanbonmatsu, L., L. F. Katz, J. Ludwig, L. A. Gennetian, G. J. Duncan, R. C. Kessler, E. K. Adam, T. McDade, and S. T. Lindau (2011). Moving to opportunity for fair housing demonstration program: Final impacts evaluation.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518 – 529.
- Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* 113(485), F3–F33.
- Tourangeau, K., J. Burke, T. Le, S. Wan, M. Weant, E. Brown, N. Vaden-Kiernan, E. Rinker, R. Dulaney, K. Ellingsen, B. Barrett, I. Flores-Cervantes, N. Zill, J. Pollack, D. Rock, S. Atkins-Burnett, and S. Meisels (2001). *ECLS-K Base Year Public-Use Data Files and Electronic Codebook*. Washington, DC: National Center for Education Statistics: U.S. Department of Education. (NCES 2001-029).

APPENDIX A FIGURES

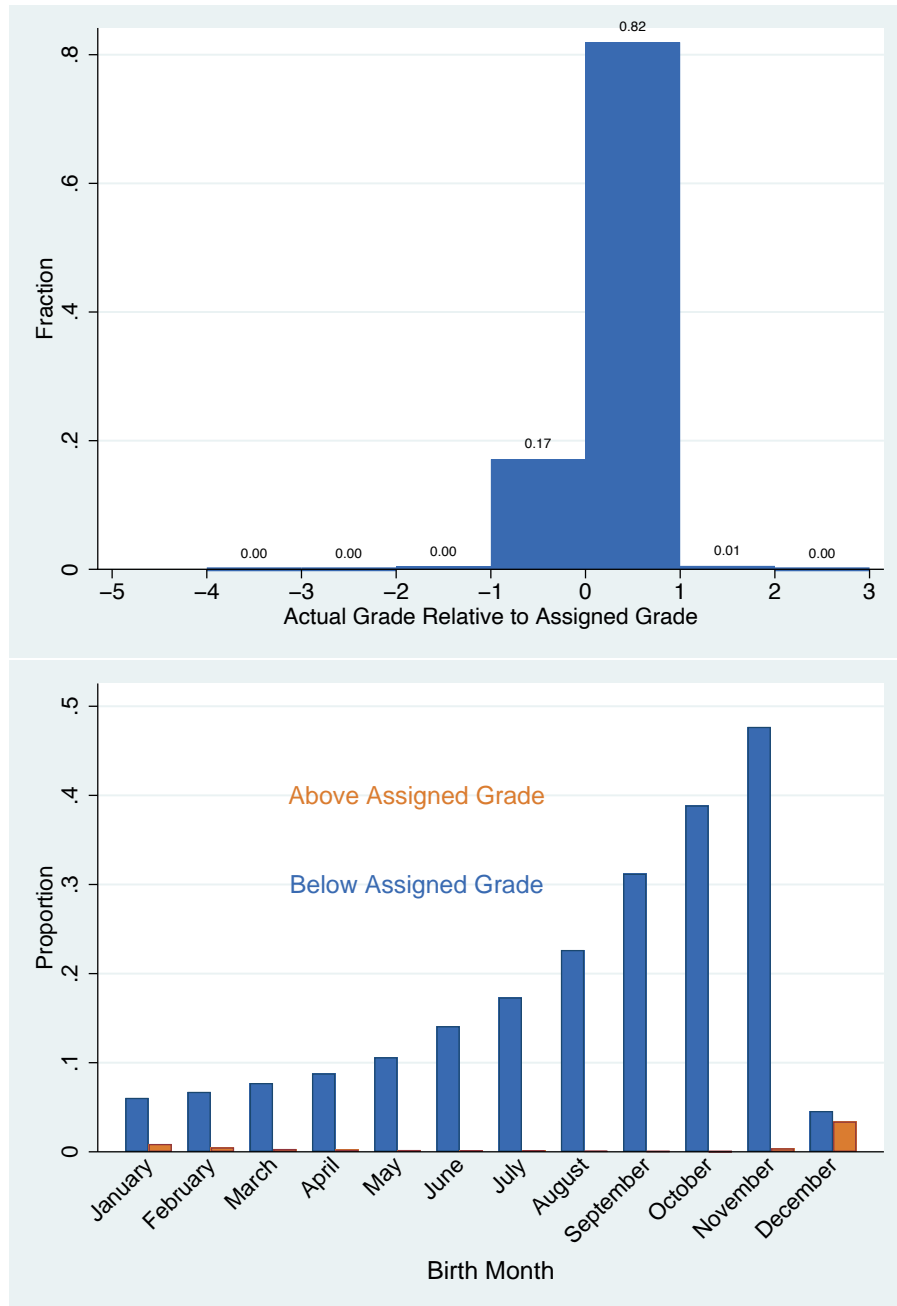


Figure 1.A.1 Cohort Compliance, Overall and by Birth Month

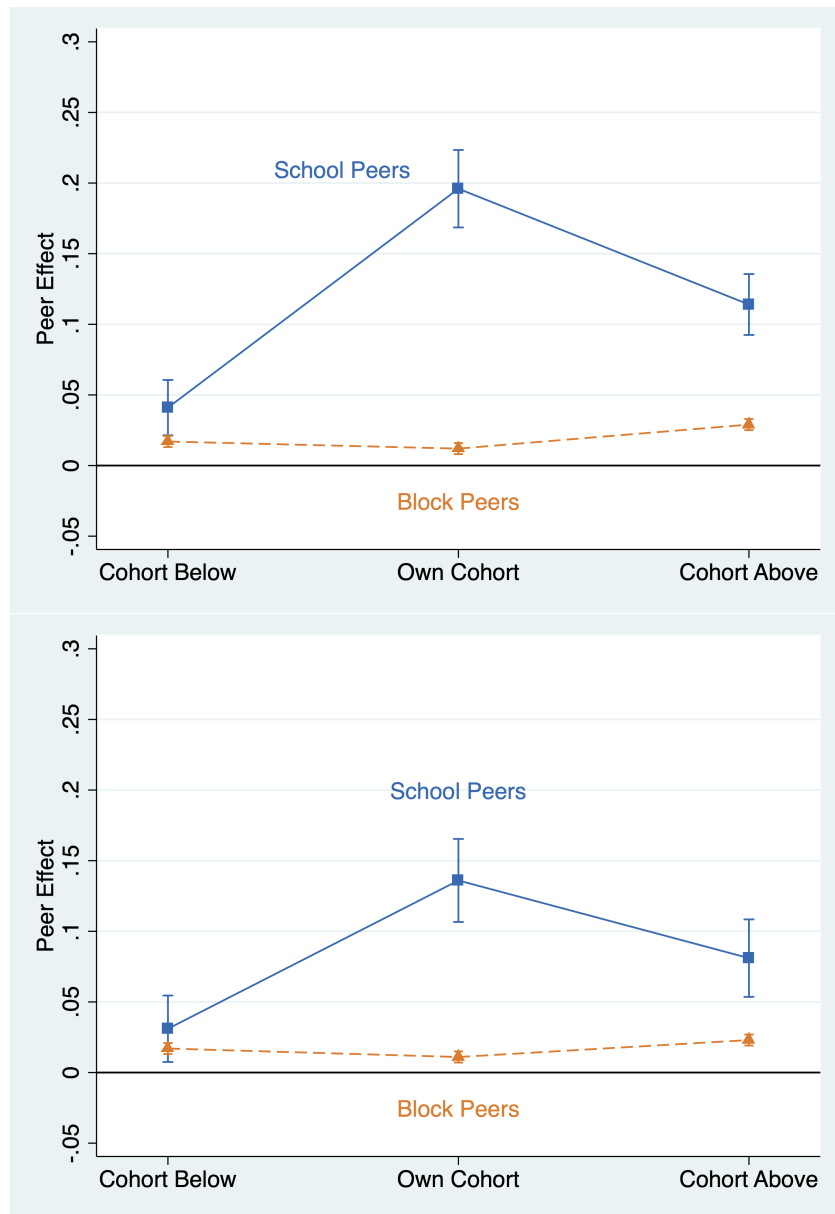


Figure 1.A.2 Assigned Cohort IV Peer Effects, Math and Reading Scores

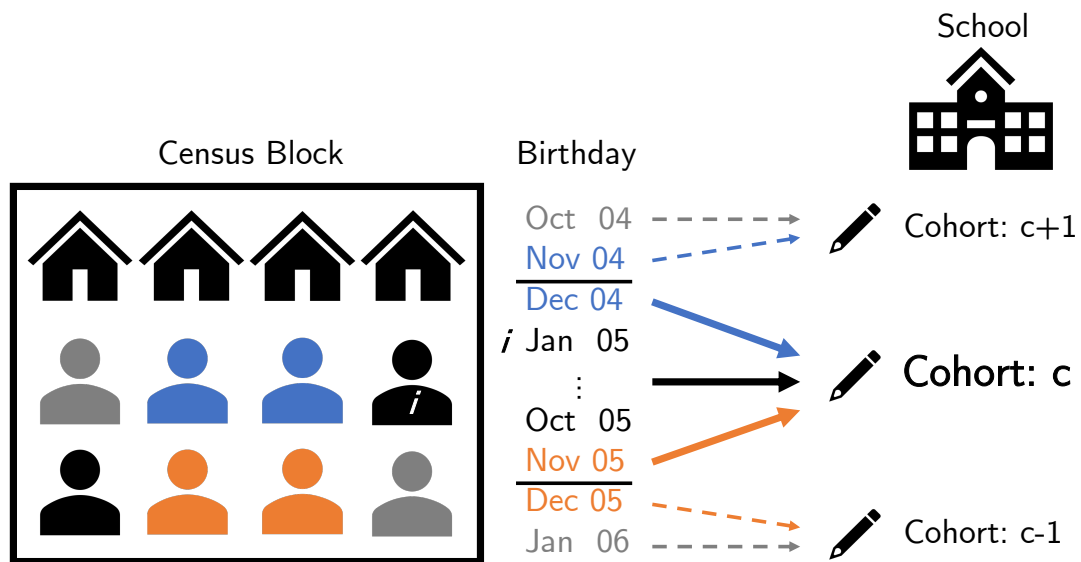


Figure 1.A.3 Sorting Neighborhood Peers into School Cohorts by Birth Month and Year

$$\bar{Y}_{-i,c} = \beta_1 \bar{Y}_{Dec,c} + \beta_2 \bar{Y}_{Nov,c+1} + \beta_3 \bar{Y}_{Nov,c} + \beta_4 \bar{Y}_{Dec,c-1} + \dots$$

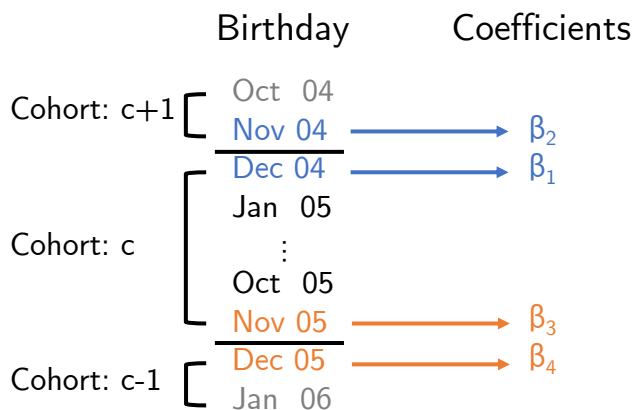


Figure 1.A.4 Connecting Identification Strategy to First Stage Equations

APPENDIX B TABLES

Table 1.B.1 Summary Statistics, Full versus Analysis Samples

Variable	Analysis Sample	Full Sample (Grades 5-8)	Differences
Math scores	0.038 (1.026)	0.000 (1.000)	0.038** [0.000]
Reading Scores	0.026 (1.006)	-0.000 (1.000)	0.026** [0.000]
# Students in School-Cohort	198.0 (124.7)	162.4 (118.0)	35.6** [0.1]
# Students in Block-Cohort	2.589 (3.724)	2.142 (3.282)	0.447** [0.002]
Female	49.8	48.7	1.1**
White	64.7	69.5	-4.7**
Black	21.9	18.0	3.9
Hispanic	6.6	6.2	0.3**
Asian/PI	4.0	3.0	0.9**
Other race	2.9	2.9	0.0
Eligible for Special Ed services	10.3	13.3	-3.0
Free or Reduced-Price Lunch	48.9	51.9	-3.1**
Limited English Proficiency	6.6	6.1	0.4**
Locale of student's school: city	26.8	22.1	4.7**
Locale of student's school: suburb	55.5	43.2	12.3**
Locale of student's school: town	8.3	12.1	-3.8**
Locale of student's school: rural	9.4	22.6	-13.2**
Observations	2,722,950	7,864,719	

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets, standard deviations in parentheses.

Table 1.B.2 School and Block Peer Effects by Cohort, Math Scores

Specification Relative Cohort	OLS	Assigned Cohort Reduced Form	Assigned Cohort IV	Cutoff Student IV
School Peers				
Grade Below	0.047** [0.008]	0.068** [0.007]	0.066** [0.010]	0.046** [0.015]
Own Grade	0.303** [0.011]	0.269** [0.010]	0.327** [0.013]	0.344** [0.019]
Grade Above	0.110** [0.009]	0.160** [0.008]	0.163** [0.010]	0.128** [0.015]
Block Peers				
Grade Below	0.036** [0.001]	0.035** [0.001]	0.043** [0.002]	0.020** [0.007]
Own Grade	0.037** [0.001]	0.035** [0.002]	0.031** [0.003]	0.049** [0.010]
Grade Above	0.049** [0.001]	0.044** [0.001]	0.056** [0.002]	0.028+ [0.016]
Kleibergen-Paap F Stat			1621	126
Observations	2,999,834	2,676,385	2,666,141	2,612,911

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.B.3 School and Block Peer Effects by Cohort, Reading Scores

Specification Relative Cohort	OLS	Assigned Cohort Reduced Form	Assigned Cohort IV	Cutoff Student IV
Grade Below	0.041** [0.006]	0.061** [0.008]	0.058** [0.012]	0.029 [0.019]
Own Grade	0.244** [0.013]	0.196** [0.011]	0.256** [0.015]	0.244** [0.025]
Grade Above	0.065** [0.008]	0.126** [0.009]	0.130** [0.011]	0.115** [0.019]
Block Peers				
Grade Below	0.030** [0.001]	0.030** [0.001]	0.037** [0.002]	0.016* [0.007]
Own Grade	0.031** [0.001]	0.029** [0.001]	0.027** [0.002]	0.047** [0.009]
Grade Above	0.038** [0.001]	0.034** [0.001]	0.044** [0.002]	0.005 [0.016]
Kleibergen-Paap F Stat			1653	155
Observations	2,983,697	2,662,688	2,656,132	2,593,933

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.B.4 Robustness Checks, Math Scores

Robustness Check Relative Cohort	All Controls	G3 Reflection Adj	G3 Own Score Control	Test Change
School Peers				
Cohort Below	0.026 [0.017]	0.101** [0.012]	0.037** [0.010]	0.071** [0.009]
Own Cohort	0.230** [0.018]	0.184** [0.018]	0.195** [0.014]	0.350** [0.013]
Cohort Above	0.127** [0.014]	0.062** [0.011]	0.104** [0.011]	0.172** [0.010]
Block Peers				
Cohort Below	0.029** [0.003]	0.046** [0.003]	0.018** [0.002]	0.043** [0.002]
Own Cohort	0.019** [0.003]	0.026** [0.003]	0.012** [0.002]	0.031** [0.003]
Cohort Above	0.037** [0.003]	0.032** [0.002]	0.030** [0.002]	0.057** [0.002]
Kleibergen-Paap F Stat	1451	99	1592	1620
Observations	1,385,584	1,737,958	1,879,759	2,444,208

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets, clustered at the school level.

Table 1.B.5 Robustness Checks, Reading Scores

Robustness Check Relative Cohort	All Controls	G3 Reflection Adj	Own Score Control	Test Change
School Peers				
Cohort Below	-0.005 [0.020]	0.063** [0.012]	0.031** [0.012]	0.079** [0.010]
Own Cohort	0.175** [0.021]	0.116** [0.018]	0.133** [0.015]	0.272** [0.014]
Cohort Above	0.081** [0.019]	0.071** [0.012]	0.078** [0.012]	0.144** [0.012]
Block Peers				
Cohort Below	0.021** [0.003]	0.040** [0.002]	0.017** [0.002]	0.038** [0.002]
Own Cohort	0.017** [0.003]	0.019** [0.003]	0.011** [0.002]	0.026** [0.002]
Cohort Above	0.028** [0.003]	0.030** [0.002]	0.023** [0.002]	0.045** [0.002]
Kleibergen-Paap F Stat	239	70	747	1695
Observations	1,381,070	1,726,584	1,874,032	2,429,319

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

APPENDIX C

APPENDIX TABLES AND FIGURES

C.1 Appendix Figures

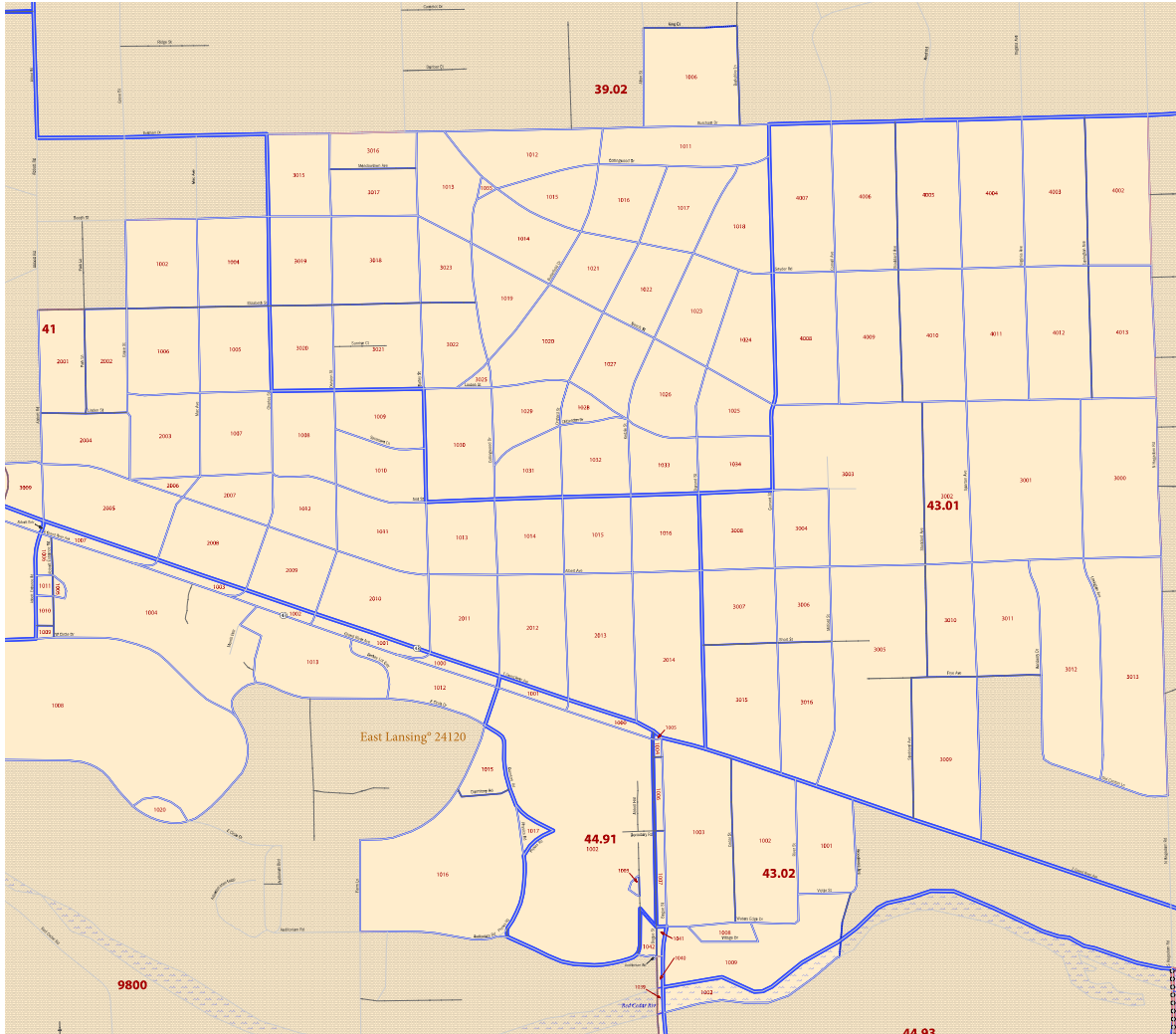


Figure 1.C.1 Census Map of Downtown East Lansing, MI

C.2 Appendix Tables

Table 1.C.1 Block Group Summary Statistics, Full versus Analysis Samples

Variable	Analysis Sample	Full Sample (Grades 5-8)	Differences
Block group total population	1,626 (1,011)	1,569 (911)	56** [0]
Female	51.5	51.0	0.5**
White	71.2	75.8	-4.6**
Black	17.3	13.7	3.5**
Hispanic	5.2	5.0	0.2**
Asian/PI	3.4	2.7	0.8**
Other race	2.9	2.8	0.1**
Median household income	\$66,494 (35,336)	\$64,120 (31,767)	\$2,375** [16]
Per-capita income	\$30,880 (14,811)	\$29,883 (13,432)	\$997** [7]
# of Census Block Groups	6,645	8,129	

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets, standard deviations in parentheses.

C.3 Heterogeneity Tables

Table 1.C.2 Assigned Cohort IV, By Grade, Math Scores

Grade/Relative Cohort	5	6	7	8
School Peers				
Cohort Below	0.094** [0.021]	0.049 [0.030]	0.083** [0.021]	0.104** [0.022]
Own Cohort	0.390** [0.019]	0.344** [0.024]	0.199** [0.021]	0.264** [0.026]
Cohort Above	0.108** [0.024]	0.265** [0.025]	0.231** [0.020]	0.049* [0.022]
Block Peers				
Cohort Below	0.035** [0.003]	0.047** [0.003]	0.048** [0.003]	0.044** [0.003]
Own Cohort	0.028** [0.004]	0.029** [0.004]	0.029** [0.005]	0.030** [0.004]
Cohort Above	0.057** [0.003]	0.056** [0.003]	0.047** [0.004]	0.030** [0.003]
Kleibergen-Paap F Stat	1484	810	508	11
Observations	689,118	720,788	714,241	498,924

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.C.3 Assigned Cohort IV, By Grade, Reading Scores

Grade/Relative Cohort	5	6	7	8
School Peers				
Cohort Below	0.069** [0.020]	0.064+ [0.037]	0.024 [0.027]	0.093** [0.027]
Own Cohort	0.281** [0.020]	0.210** [0.031]	0.197** [0.026]	0.240** [0.023]
Cohort Above	0.086* [0.037]	0.151** [0.029]	0.169** [0.022]	0.028 [0.034]
Block Peers				
Cohort Below	0.030** [0.003]	0.044** [0.003]	0.039** [0.004]	0.038** [0.003]
Own Cohort	0.025** [0.004]	0.025** [0.004]	0.023** [0.004]	0.023** [0.004]
Cohort Above	0.046** [0.003]	0.042** [0.003]	0.042** [0.003]	0.024** [0.003]
Kleibergen-Paap F Stat	1427	837	131	7
Observations	684,910	719,792	711,246	496,478

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.C.4 Assigned Cohort IV, By Gender

Subject Gender/Relative Cohort	Math		Reading	
	Female	Male	Female	Male
School Peers				
Cohort Below	0.066** [0.011]	0.066** [0.011]	0.069** [0.013]	0.051** [0.012]
Own Cohort	0.324** [0.014]	0.330** [0.013]	0.233** [0.016]	0.268** [0.016]
Cohort Above	0.157** [0.011]	0.164** [0.011]	0.151** [0.013]	0.137** [0.012]
Block Peers				
Cohort Below	0.041** [0.003]	0.045** [0.003]	0.037** [0.002]	0.037** [0.003]
Own Cohort	0.032** [0.003]	0.029** [0.004]	0.031** [0.003]	0.022** [0.003]
Cohort Above	0.055** [0.003]	0.057** [0.003]	0.043** [0.002]	0.045** [0.003]
Kleibergen-Paap F Stat	1847	1284	624	1271
Observations	1,327,525	1,338,463	1,320,717	1,328,745

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.C.5 Assigned Cohort IV, By Race/Ethnicity, Math Scores

Race/Relative Cohort	White	Black	Hispanic	Asian/PI
School Peers				
Cohort Below	0.061** [0.010]	0.068** [0.017]	0.030 [0.020]	0.086* [0.033]
Own Cohort	0.314** [0.013]	0.287** [0.029]	0.340** [0.023]	0.350** [0.044]
Cohort Above	0.177** [0.011]	0.127** [0.017]	0.110** [0.022]	0.121** [0.028]
Block Peers				
Cohort Below	0.045** [0.002]	0.019** [0.003]	0.032** [0.007]	0.042** [0.009]
Own Cohort	0.033** [0.003]	0.007+ [0.004]	0.009 [0.008]	0.035** [0.010]
Cohort Above	0.058** [0.003]	0.030** [0.004]	0.039** [0.008]	0.042** [0.009]
Kleibergen-Paap F Stat	1087	1326	754	639
Observations	1,731,127	576,896	173,698	105,810

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.C.6 Assigned Cohort IV, By Race/Ethnicity, Reading Scores

Race/Relative Cohort	White	Black	Hispanic	Asian/PI
School Peers				
Cohort Below	0.031* [0.013]	0.089** [0.019]	0.044+ [0.023]	0.042 [0.037]
Own Cohort	0.227** [0.016]	0.246** [0.030]	0.255** [0.035]	0.199** [0.047]
Cohort Above	0.139** [0.011]	0.126** [0.022]	0.106** [0.025]	0.094** [0.032]
Block Peers				
Cohort Below	0.040** [0.002]	0.019** [0.003]	0.023** [0.007]	0.016+ [0.009]
Own Cohort	0.028** [0.003]	0.012** [0.004]	0.026** [0.009]	0.032** [0.009]
Cohort Above	0.048** [0.003]	0.024** [0.003]	0.022** [0.008]	0.029** [0.009]
Kleibergen-Paap F Stat	643	381	722	158
Observations	1,720,828	575,812	171,523	103,109

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.C.7 Assigned Cohort IV, By Economic Disadvantage Status

Subject Economic Disadvantage/ Relative Cohort	Math		Reading	
	Yes	No	Yes	No
School Peers				
Cohort Below	0.030* [0.014]	0.055** [0.015]	0.015 [0.018]	0.009 [0.018]
Own Cohort	0.235** [0.020]	0.276** [0.018]	0.194** [0.023]	0.184** [0.021]
Cohort Above	0.107** [0.014]	0.124** [0.013]	0.096** [0.016]	0.091** [0.015]
Block Peers				
Cohort Below	0.033** [0.003]	0.028** [0.003]	0.027** [0.003]	0.018** [0.003]
Own Cohort	0.011** [0.004]	0.028** [0.004]	0.017** [0.004]	0.022** [0.004]
Cohort Above	0.039** [0.003]	0.043** [0.004]	0.037** [0.003]	0.030** [0.003]
Kleibergen-Paap F Stat	1093	1132	579	174
Observations	783,234	831,175	777,673	827,633

** p<0.01, * p<0.05, + p<0.1.

Robust standard errors in brackets, clustered at the school level.

Table 1.C.8 Assigned Cohort IV, By 3rd Grade Score

Subject Initial Score/Relative Cohort	Math		Reading	
	Above Median	Below Median	Above Median	Below Median
School Peers				
Cohort Below	0.049** [0.010]	0.057** [0.010]	0.033** [0.012]	0.061** [0.012]
Own Cohort	0.266** [0.013]	0.224** [0.013]	0.160** [0.013]	0.186** [0.016]
Cohort Above	0.147** [0.012]	0.140** [0.010]	0.105** [0.011]	0.116** [0.011]
Block Peers				
Cohort Below	0.027** [0.002]	0.024** [0.002]	0.022** [0.002]	0.023** [0.002]
Own Cohort	0.026** [0.003]	0.006* [0.002]	0.019** [0.002]	0.014** [0.002]
Cohort Above	0.043** [0.003]	0.032** [0.002]	0.025** [0.002]	0.033** [0.002]
Kleibergen-Paap F Stat	1433	1530	277	1285
Observations	1,333,424	1,332,550	1,267,636	1,381,796

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets, clustered at the school level.

APPENDIX D REFLECTION PROBLEM

Let $Y_{i,t}$ be the observed outcome of interest Y (e.g. test scores) of observation i at time t , $\bar{Y}_{-i,c,b,t}$ be the average of the same for the all other individuals in cohort c and block b at time t .¹ A simple model seeking to determine the peer effects of block-cohort peers would be:

$$Y_{i,t} = \beta_0 + \beta_1 \bar{Y}_{-i,c,b,t} + \nu_{i,t} \quad (\text{D.1})$$

Which formalizes the effects of block-cohort as a linear combination of observed peer outcomes $\bar{Y}_{-i,c,b,t}$ and other unobserved factors. The first issue with equation D.1 is the reflection problem, as described by Manski (1993). That is, if we believe equation D.1, then it also follows that:

$$Y_{j,t} = \beta_0 + \beta_1 \bar{Y}_{-j,c,b,t} + \nu_{j,t} \quad \forall j \in c, b, t$$

If we plug this equation back into equation D.1 for each $\bar{Y}_{-i,c,b,t}$, we can see that, because $Y_{i,t} \in \bar{Y}_{-j,c,b,t}$, estimating equation D.1 directly would bias β_1 upward because $Y_{i,t}$ is mechanically correlated with itself. Intuitively, if student i 's block-cohort peers affect i 's outcomes and student i 's outcome affects their block-cohort peers simultaneously, then any estimation of equation D.1 would overestimate the peer effects of block-cohort peers on i . To correct this, I use peer scores lagged by one year², denoted as $\bar{Y}_{-j,c,b,t-1}$.³ Because $Y_{i,t}$ has changed in the year following $t-1$, the dependence of $\bar{Y}_{-j,c,b,t-1}$ on $Y_{i,t}$ should be more limited as it only exists to the extent that $Y_{i,t}$ is still dependent on $Y_{i,t-1}$, observation i 's own one year lagged outcome. While this likely does not eliminate the reflection problem entirely, it does reduce it to an extent that should limit the issue of overestimation. Now, rewriting equation D.1, we have:

¹This generalizes for any function f of $Y_{-i,c,b,t}$ where $\frac{\partial f(f(Y_{i,t}, Y_{-i,c,b,t}))}{\partial Y_{i,t}} \neq 0$.

²Results are also robust to alternative methods, such as using peer scores lagged by two years or from third grade.

³The best solution is to use pre-measures from before the block-cohort peers interacted with each other, as is done in Imberman et al. (2012). However, for test scores, students are not tested until grade 3, and some block-cohort students have been interacting since kindergarten or before.

APPENDIX E CUTOFF IV

E.1 Michigan School Cohort Entry Policy

Historically, Michigan had one of the latest birthdate cutoffs in the nation. As of the 2012-2013 school year, a child had to be age 5 by 12/1¹ in order to be old enough to enter kindergarten. In contrast, by 2018, no other state had an entry cutoff date later than 10/15, and most had cutoff dates in September or earlier.² However, in June 2013, Michigan revised its school code so that the cohort entry cutoff date would move up by one month each school year³ until the 2015-2016 school year, when the cohort entry cutoff date would be 9/1. Parents may still enroll their child in kindergarten if they do not meet this cutoff and their child is born before 12/1 and they submit a waiver to the school district, but the school district may make recommendations against enrolling this child “due to age or other factors”.⁴ What this means for the use of cohort entry cutoffs as an instrument is twofold. First, parents’ ability to waive their child even past the cutoff means there will likely not be strict compliance, requiring the use of assigned cohort based strictly on age and month of birth as an instrument for actual cohort to capture estimates of treatment-on-treated. For almost all of the analysis sample, the relevant legal cutoff date was 12/1 when the students were entering kindergarten. Because of this, I will be referring to 12/1 as the cutoff date and November and December as the cutoff months for the remainder of the paper.

In Tables E.1 and E.2, I examine whether students born on either side of the school entry cutoff are valid comparison groups by comparing their observable characteristics. All statistics presented, except number of peers, are conditional on having at least one student in the group. These peer groups are defined based strictly based on *assigned* cohort, which is determined by their month of birth, year of birth, and kindergarten entry cutoff date when they were age 5. If birth timing is random, following my identification assumption, then these two groups should be very similar on both observable and unobservable characteristics. As we can see in Tables E.1 and E.2, examining

¹There is anecdotal evidence that some districts implemented earlier recommended entry dates closer to the national norm than this later date.

²Source: <https://nces.ed.gov/programs/statereform/tab5 3.asp>

³Cutoff in 2013-2014: 11/01. Cutoff in 2014-2015: 10/01

⁴Source: <https://www.michigan.gov/documents/kindergarten 122554 7.pdf>

Table 1.E.1 School Cutoff Peers Comparison

Variable	Own-Above Cohort			Own-Below Cohort		
	Dec-Own	Nov-Above	Own-Above-Diff	Nov-Own	Dec-Below	Own-Below-Diff
# of students	15.20 (10.55)	9.64 (10.87)	5.56** [0.01]	12.69 (9.22)	10.06 (10.27)	2.63** [0.01]
Lagged math scores	0.101 (0.603)	-0.024 (0.616)	0.125** [0.000]	0.015 (0.616)	0.105 (0.617)	-0.090** [0.000]
Lagged reading scores	0.096 (0.551)	-0.039 (0.569)	0.135** [0.000]	-0.005 (0.569)	0.097 (0.569)	-0.103** [0.000]
Female	49.0	48.1	0.9**	51.0	49.3	1.7**
White	64.6	63.0	1.6**	63.4	62.2	1.2**
Black	21.5	23.3	-1.7**	21.6	23.6	-2.0**
Hispanic	6.7	6.8	-0.1**	7.1	6.9	0.2**
Asian/PI	4.0	3.9	0.1**	4.6	4.1	0.5**
Other Race	3.0	2.8	0.1**	3.1	3.1	0.0**
Eligible for Special Ed services	12.2	13.7	-1.4**	12.0	11.8	0.2**
Free or Reduced-Price Lunch	49.1	50.2	-1.1**	49.6	50.7	-1.1**
Limited English Proficiency	5.8	6.7	-1.0**	7.1	6.6	0.5**
Observations	1,864,041	1,185,126		1,902,919	1,419,362	

** p<0.01, * p<0.05, + p<0.1. Standard errors in brackets, standard deviations in parentheses.

Table 1.E.2 Block Cutoff Peers Comparison

Variable	Own-Above Cohort			Own-Below Cohort		
	Dec-Own	Nov-Above	Own-Above-Diff	Nov-Own	Dec-Below	Own-Below-Diff
# of students	0.19 (0.51)	0.18 (0.50)	0.01** [0.00]	0.18 (0.50)	0.19 (0.51)	-0.01** [0.00]
Lagged math scores	0.181 (1.022)	0.089 (1.025)	0.092** [0.003]	0.082 (1.021)	0.196 (1.013)	-0.114** [0.003]
Lagged reading scores	0.154 (0.981)	0.051 (0.987)	0.104** [0.003]	0.043 (0.986)	0.164 (0.982)	-0.121** [0.003]
Female	49.8	49.6	0.2	49.6	49.7	-0.1
White	64.6	65.1	-0.5**	64.6	64.1	0.5**
Black	19.6	19.1	0.5**	19.0	19.6	-0.6**
Hispanic	7.4	7.1	0.3**	7.5	7.6	-0.1
Asian/PI	5.3	5.5	-0.2**	5.7	5.4	0.2**
Other race	2.9	2.9	-0.0	3.0	3.1	-0.1+
Eligible for Special Ed services	10.2	11.2	-1.0**	11.4	10.3	1.1**
Free or Reduced-Price Lunch	45.9	44.9	1.0**	45.7	46.4	-0.7**
Limited English Proficiency	6.4	7.5	-1.1**	8.1	7.2	0.8**
Observations	282,636	273,213		272,653	280,656	

** p<0.01, * p<0.05, + p<0.1. Standard errors in brackets, standard deviations in parentheses.

school and block cutoff peers, respectively, the two groups are largely similar in terms of gender, race, special education, free/reduced-price lunch, and limited english proficiency status. The two areas of observable characteristics where they differ are number of students, for schools, and test scores, for both. Numbers of students in schools differ because some schools do not have above or below cohorts for certain grade levels: for example, a grade K-5 school does not have above-cohort school peers for fifth graders, and a grade 6-8 school does not have below-cohort school peers for sixth graders.⁵ Other than test scores, the observable data suggests that the assumption of random assignment holds.

While test scores are consistently higher for students born in December than November, prior literature suggests this should not be a cause for concern on other unobservable background characteristics. Test scores are 0.1 standard deviation higher for December students in both school-cohorts and block-cohorts. This likely stems from relative age effects, a well-known effect in education in which the oldest students within a cohort perform better than the youngest, especially at younger ages (Elder, 2010; Black et al., 2011). Given the literature showing that this results directly from cohort assignment, we can reasonably conclude that this does not pose a problem for selection on other confounding unobservable background characteristics. As long as 1) students' innate ability is constant across the cutoff threshold, 2) the test score difference largely captures remaining differences in student ability due to age differences at test time and different schooling experiences as the oldest/youngest student in cohort, and 3) the linear-in-means assumption holds (i.e., higher levels of test scores at baseline aren't a cause for concern, since changes in test scores are the primary driver of results), this particular difference should not confound my estimates.

Before running my cutoff IV specification, I need to make several changes to the sample. First, I drop all students born within one month of the school-entry cutoff for their year. As described in Section E.1, this mostly means dropping students born in November or December, but includes November and October for the 2018-2019 school year and October and September for the 2019-

⁵As discussed further in Section 1.5, I include dummy variables for all cases where students have zero school-cohort, block-cohort, or school or block cutoff peers. Excluding these cases would severely restrict the sample to only a limited number of school-cohorts.

2020 school year. Second, I drop all students who do not have at least one cutoff student in any of the four peer cohort cutoff groups in either school or block: November of cohort above, December of own cohort, November of own cohort, and December of cohort below. Third, I drop all students who are missing test score data on any of the four peer cohort cutoff groups in either school or block. These three changes leave a final cutoff analysis sample of 1,914,686 observations for math and 1,906,253 for reading.

E.2 Cutoff Student IV Estimation

The estimation strategy for the cutoff student IV estimation is similar to a fuzzy border discontinuity. Figure A.4 helps illustrate the connection between the first stage equations above and the identification strategy and Figure A.3. The first two terms use the average ability of cutoff students born just before or after the cutoff for being the oldest peers in student i 's school-cohort. The second two terms do the same, but with the average abilities of cutoff students born just before or after the cutoff for being the *youngest* peers in student i 's school-cohort. The next four terms repeat the process, but now for block-cohort cutoff peers, rather than school-cohort. Adding in controls for counts of students and proportions of observable characteristics in each cutoff group, and the first stage equation looks like:

The second stage equation is the same as the main estimation strategy, but with the inclusion of the controls for counts of students and proportions of observable characteristics in each cutoff group instead of in each assigned cohort group:

$$Y_{i,t} = \beta_0 + \sum_{(j=-1)}^1 \beta_j \widehat{Y}_{-i,c+j,s,t-1} + \sum_{(j=-1)}^1 \alpha_j \widehat{Y}_{-i,c+j,b,t-1} \\ + X_{i,t} + \sum_{(m \in Nov, Dec)} \sum_{(k \in s, b)} \sum_{(j=-1)}^1 \bar{X}_{-i,m,c+j,k,t} + \Gamma_{m,c,g,s,b_g,t} + \epsilon_{i,t}$$

where β_j through α_j are again our exogenous peer effects for school and block peers.

CHAPTER 2

EQUALIZING INPUTS, ENDURING GAPS: EXAMINING CHANGES IN LEVELS AND CORRELATES OF GENDER GAPS IN NONCOGNITIVE SKILLS OVER TIME

2.1 Abstract

I examine how gender gaps in noncognitive skills change over time by comparing two nationally representative datasets of elementary school students. I determine that girls' advantages in four out of five noncognitive measures remain large and unchanged between the 1998-1999 and 2010-2011 national cohorts, ranging from 0.35 to 0.4 standard deviations, substantially larger than gender gaps in cognitive test scores. Focusing on family background and parental input measures examined in previous literature, I investigate the extent to which these measures continue to explain noncognitive gender gaps despite no change in the overall level of gender gaps. I find that the influence of these measures in predicted gender gaps has declined, likely due to an equalization of parent reports of educational activities and warmth between boys and girls. Single motherhood and teen motherhood remain predictors of gender gaps, though the correlation between kindergarten socioeconomic status and gender gaps has decreased.

JEL classification: I21, J12, J13, J16, J24

2.2 Introduction

Although the past few decades have seen a widespread increase in math and reading testing in schools, economists and education researchers have become increasingly aware of the importance of other skills not measured by traditional cognitive tests. Although often harder to measure, these noncognitive skills, such as self-control, interpersonal skills, impulsiveness, approaches to learning, internalizing problems and externalizing problems, have been increasingly associated with both difficulties in school and long-term labor market and educational outcomes (Lindqvist and Vestman, 2011; Deming, 2017; Heckman et al., 2006). Additionally, researchers (Jacob, 2002; Goldin et al., 2006; Becker et al., 2010) are increasingly finding evidence that gender gaps in long-term educational outcomes in particular may be mainly explained by gender gaps in these noncognitive skills. Further, economists have identified gender gaps in noncognitive skills as playing a role in gender gaps of more short term outcomes, such as the historical gender gap in grades received in school (Cornwell et al., 2013; Fortin et al., 2015). In other fields, researchers in sociology (DiPrete and Jennings, 2012) and psychology (Duncan et al., 2007; Raver et al., 2007) have conducted studies demonstrating that noncognitive skills affect the accumulation of cognitive skills in the short and medium term, thus influencing long term outcomes both directly and indirectly through the process of cognitive skill formation.

To investigate these gender gaps in noncognitive skills, researchers have taken several different approaches. One branch of research that looks at descriptive and correlational evidence from time diaries on how parents spend their time with their children finds that parents spend more time with children of their own gender, leaving boys without fathers in the home with less parental time investment overall (Baker and Milligan, 2016; Bibler, 2018). Another branch of gender gap research looks outside of specific measures of noncognitive skills and focuses on differential responses by gender to disadvantaged backgrounds. These papers find evidence that growing up in disadvantaged environments, such as impoverished neighborhoods or low socioeconomic status families, seem to have a larger negative effect on boys than girls, suggesting that boys may have a higher responsiveness to disadvantage and human capital inputs more generally (Autor et al., 2019,

2020; Chetty et al., 2016, e.g.).

This paper follows a third branch of literature, which examines how gender gaps in the development of noncognitive skills correlate with other measures in young students. A prominent example of this literature is Bertrand and Pan (2013), which provides evidence that students from single mother families, low SES families, and teen mother families experience larger gender gaps in eighth grade suspension rates and one type of noncognitive skill: externalizing behavior. The paper also provides suggestive evidence that, in terms of externalizing behavior, boys are more responsive than girls to more disadvantaged family backgrounds and lower levels of parental inputs. In particular, the paper emphasizes evidence on boys' higher degree of negative responsiveness to single mother households. I build on this literature by examining changing gender gaps in noncognitive skills between elementary school cohorts that entered school 12 years apart. Specifically, I examine how the influence of family background and parental input characteristics on these noncognitive gender gaps has changed over time, and how these changes vary by a broader array of noncognitive skills than externalizing behavior alone.

I find that gender gaps in four out of five noncognitive measures in two nationally representative datasets remain large and unchanged between the 1998-1999 and 2010-2011 national cohorts. For these four measures, I can rule out any changes in girls' advantages of 0.1 standard deviations or greater across all grades. I then combine all five noncognitive measures into a single latent noncognitive skill using factor analysis and analyze how the correlations of family background and parental input measures have changed in relation to this latent measure. An Oaxaca-Blinder decomposition of the noncognitive gender gap shows that the portion of the gender gap explained by these measures decreases by fifth grade for the 2010-2011 cohort compared to the 1998-1999 cohort, despite no change in the overall gender gap. This change is likely explained by an equalizing of parent reports of educational activities and feelings of warmth between boys and girls between the two cohorts: although parents continued to report engaging in more educational activities with girls, this advantage is smaller for the 2010-2011 cohort, and no longer report more parental warmth towards girls than boys. For family background measures, there are no statistically detectable

changes in gender gaps for either single mothers or teen mother families, and both continue to be substantially negative predictors. Socioeconomic status, on the other hand, appears to have a lesser role in enlarging gender gaps. Differences in gender gaps have compressed between lower and higher ends of the socioeconomic status distribution in the second half of elementary school when controlling for other family background and parental input measures, suggesting that its influence has waned.

This paper's structure is as follows: Section 2.3 describes the data and measures I use, section 2.4 goes over my results, and section 2.5 concludes.

2.3 Data

2.3.1 Data and Sample

I use two different versions of the Early Childhood Longitudinal Study, Kindergarten Cohort datasets for its analysis: the ECLS-K and the ECLS-K:2011. Both studies are nationally representative samples of children who entered kindergarten in the 1998-1999 and 2010-2011 school years, respectively. I refer to them as the 1998 and 2010 cohorts for the rest of this paper. Both studies contain data on over 7,000 children in their K-5 longitudinal panel samples. These children and their parents, teachers, and school administrators are interviewed repeatedly in several waves. The ECLS-K, collecting data on the 1998-1999 cohort, conducted interviews in fall of kindergarten, spring of kindergarten, fall of 1st grade, and spring of 1st, 3rd, 5th grades, and 8th grades. The ECLS-K:2011, collecting data on the 2010-2011 cohort, conducted interviews in spring and fall of kindergarten, 1st grade, and 2nd grade, as well as spring of 3rd, 4th, and 5th grades. In both of these studies, information was collected about children's cognitive, social, emotional, and physical development by interviewing children, parents, teachers, and administrators. Additional information was collected on the children's home environment (including parental educational activities), the environment at school, and school and teacher practices and qualifications.

To create the final analysis sample, I impose several sample restrictions. First, observations had to be respondents through all rounds of their surveys, indicated by having non-zero fifth grade panel weights. Second, respondents had to have non-missing responses on all control

variables. The base control variables to be used throughout this paper are dummy variables for race and school locale at kindergarten.¹ Third, respondents had to have non-missing responses on all family background and parental input measures. These measures are: mother's age at first birth, family socioeconomic status (derived by survey designers from household income, education, and occupations), family structure (i.e. two biological parents, single mother, or other family structure), parental educational activities (combined into a HOME index)², parental warmth (combined into a Warmth index)³, and parental disciplinary behavior (whether they spank their child). And fourth, respondents had to have non-missing data on the key outcomes of interest: externalizing behavior, self control, interpersonal skills, approaches to learning, and internalizing problems as reported by teachers in kindergarten in fifth grade. This leaves 6,630 observations for the ECLS-K dataset and 4,938 observations for the ECLS-K:2011 dataset. Further description of the sample constructions is available in Online Appendix Section C.1. Weighted descriptive statistics are reported for these two samples in Table B.1. Fifth grade panel weights included with the datasets are used for the remainder of the paper. These weights are designed to make the sample nationally representative in light of stratified sampling methodology and survey nonresponse.⁴

Table B.1 shows that, while the two cohorts are largely similar, the 2010 cohort is slightly more advantaged, primarily in terms of higher parental education and lower teen motherhood and single mother family rates. There is a 4 percentage point decrease in the number of children born to mothers who were teenagers at their first birth and a 6 percentage point increase in those born to mothers over 30 at first birth. There is also a 6 percentage point decrease in families with only a high school education and a 4 percentage point increase in families with at least one parent having a bachelor's degree or greater. Finally, there is a 2 percentage point decrease in children living in single mother households at kindergarten combined with a 4 percentage point increase in children with two biological parents at kindergarten.⁵ In sum, the 2010 cohort has larger proportions of

¹Results are robust to including a fuller set of controls more comparable to Bertrand and Pan (2013): race, age and age-squared at first assessment, birthweight, and number of older and younger brothers and sisters

²Described further in the next subsection.

³Described further in the next subsection.

⁴Results are robust to the use of inverse probability weights to account for item nonresponse in both surveys.

⁵Following Bertrand and Pan (2013) I define two *biological* parents as the base group for family structure, rather

educated parents and two parent households, and a smaller proportion of teenage mothers.

2.3.2 Key Measures

Following Bertrand and Pan (2013), I have created two indices of parental inputs: a HOME index, which standardizes the sum of eight measures of parental investment activities⁶ and a Warmth index, which standardizes the sum of eight measures on parental feelings towards their child.^{7,8} These measures are included as measured at kindergarten in order to limit potential endogeneity or reverse causality with child noncognitive abilities that may affect family structure or parental inputs. I use these two indices, in combination with an indicator for whether the parent reported spanking their child in the last week in kindergarten, as proxies for parental inputs for the duration of this paper.

In addition to data on parent-reported investment and child-rearing activities and attitudes, another important aspect of both ECLS-K datasets is their inclusion of measures of noncognitive skills, particularly teacher-reported noncognitive skills.⁹ Both datasets contain teacher-reported measures on externalizing behaviors, self-control, approaches to learning, interpersonal skills, and internalizing problems. These social skills scales were developed based on teachers' responses to questions taken from the Social Skills Rating System. The score on each scale is the mean rating of all questions included in the scale. Although the components of these measures are not available due to copyright reasons, the ECLS-K user's manual provides descriptions of each of the noncognitive measures (Tourangeau et al., 2001).

The scales are described as follows. Externalizing behaviors are constructed from “five items on this scale [that] rate the frequency with which a child argues, fights, gets angry, acts impulsively, and

than any two parent households. Results are robust to the use of two parent households as the base category instead.

⁶Measures are whether: read to child ≥ 3 times per week, child has ≥ 20 books, child reads ≥ 3 times per week outside school, have home computer child uses, has visited museum, concert, or library with child, and whether child participated in other outside school activities (dance, sports, music, etc.).

⁷Measures are Likert scales on how true parents felt following statements were: have warm or close times together, child likes me, always show child love, express affection, (reversed) being parent harder than expected, (reversed) child does things that bother me, (reversed) sacrifice to meet child's needs, and (reversed) often feel angry with child.

⁸Both were constructed following Bertrand and Pan as closely as possible, though several measures were dropped from the Warmth index due to lack of inclusion in the ECLS-K:2011 data.

⁹Parent-reported ratings of noncognitive skills are also available for early grades. However, I do not include them in this analysis out of concern that parents are less likely to be objective, unbiased assessors of their children's abilities.

disturbs ongoing activities.” Self-control is constructed from “four items that indicate the child’s ability to control behavior by respecting the property rights of others, controlling temper, accepting peer ideas for group activities, and responding appropriately to pressure from peers.” Approaches to learning is constructed from “six items that rate the child’s attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization.” Interpersonal skills are constructed from items that “rate the child’s skill in forming and maintaining friendships, getting along with people who are different, comforting or helping other children, expressing feelings, ideas and opinions in positive ways, and showing sensitivity to the feelings of others.” And finally, the internalizing problems scale is constructed from four items that ask about “the apparent presence of anxiety, loneliness, low self-esteem, and sadness.”

Two of these measures, externalizing behavior and internalizing behavior, have been reordered so that higher scores indicate the child exhibited a “better” score reflecting higher noncognitive skill in each respective category. This means more of each measure’s behaviors for positive scales (approaches to learning, self-control, and interpersonal skills) or less of each measure’s behaviors for negative scales (externalizing and internalizing problems). Additionally, in order to allow for comparability and reduce arbitrary scaling, all noncognitive measures are standardized within the estimated population of their respective surveys.

Both the Social Skills Rating System itself and these measures from the ECLS-K based on the SSRS are also used in numerous other studies involving the ECLS-K. This includes Neidell and Waldfogel (2010), who state “these scales have high construct validity as assessed by test-retest reliability, internal consistency, inter-rater reliability, and correlations with more advanced behavioral constructs (Elliott et al., 1988) and are considered the most comprehensive social skill assessment that can be widely administered in large surveys such as the ECLS-K (Demaray et al., 1995).” Taken together, these endorsements and descriptions provide evidence for the validity of the measures of noncognitive skills I will be using for the remainder of this paper.

2.4 Results

2.4.1 Gender Gaps Remain Wide

The first question to be addressed is whether gender gaps have changed between the two cohorts. To this end, I have recreated weighted measures of gender gaps in all five teacher-reported noncognitive skills in the two ECLS-K datasets and included them in Figure A.1.¹⁰ Each data point is the coefficient on a female dummy from a regression of each measure in each grade and each cohort on a female dummy variable. It is worth noting two additional points: (1) Fall-Kindergarten and Spring-Kindergarten survey waves are listed as KF and KS for the remainder of the paper, and (2) the 2010 cohort plots (in orange, dashed lines) have more data points due to conducting surveys at second and fourth grade, unlike for the 1998 cohort.

Figure A.1 shows that gender gaps in these noncognitive measures are largely unchanged for all but internalizing problems. While internalizing problems shows a narrowing of the gender gap to the point of no longer having a statistically detectable gender gap for the 2010 cohort, gender gaps in teacher reports of externalizing behavior, self control, interpersonal skills, and approaches to learning remain in the range of 0.35-0.6 standard deviations. For comparison, similar graphs for gender gaps in math and reading test scores are included in Appendix Figure C.1. As Bertrand and Pan (2013) note, the gender gaps in noncognitive skills remain much larger in relative magnitude to any gender gaps in test scores, which are in the range of 0.1 to 0.25 standard deviations. This difference holds across both ECLS-K cohorts.

Table B.2 shows and tests the differences in gender gaps between the two cohorts directly. As Table B.2 shows, when testing for differences between cohorts jointly across all grades, I can only reject the null only for internalizing behavior. In both Panels A and B of Table B.2, each cell in the first five columns is the coefficient on a female and 2010 cohort interaction term from regressions of each respective measure in each respective grade listed in the column title on a dummy for female, a dummy for the 2010 cohort, and a female times 2010 cohort interaction term. The last column shows the p-value of a joint F-test of the null that the gender gap in each measure across

¹⁰The sixth graph displays the same for the common factor, which will be described further in the next subsection.

all grades is unchanged between the two datasets. In panel A these regressions are run without any controls. In Panel B, the regressions are rerun with controls included for child race, school locale at kindergarten, family background at kindergarten (single motherhood, family socioeconomic quintile, and teen motherhood), and parental inputs at kindergarten (lower HOME index, lower Warmth index, and spanking at kindergarten). The results in Panel B largely confirm those of Panel A, with no statistically detectable changes for any of the measures other than internalizing problems.

Taken together, these findings suggest that gender gaps in noncognitive skills remain a substantial issue 12 years later. However, before fully concluding this, I must address one of the drawbacks in interpretation of measures based on subjective evaluations: I may not be able to separate changes in how teachers evaluate the same level of skill between genders from actual changes in the underlying skills themselves. To address this concern, I have included estimates of the gender gaps in teachers' subjective Academic Rating Scores¹¹ in the two cohorts in Appendix Figure C.2. If the endurance in subjective reports of gender gaps was solely due to teachers evaluating boys relatively more unfavorably in general, despite a real narrowing of objective skill gaps, we might expect this would result in a shift towards girls in Academic Rating Scores gender gaps. Instead, the change in the female-male gap in Academic Rating Scores between the two cohorts appears to have grown more *favorable* towards boys between the 1998 and 2010 cohorts.¹² In sum, the endurance of gender gaps in subjectively-rated noncognitive skills does not appear to be masking a real narrowing of objective skill, and are thus likely reflects a real persistence.

2.4.2 Factor Analysis

In addition to the issue of subjectivity bias, a second issue with interpreting the results in Table B.2 is that many of these measured noncognitive skills are highly correlated with each other. Table B.3 shows their correlation matrix across all grades and cohorts. The correlations Table B.3

¹¹These scores are produced by teachers evaluating the academic abilities of students in different subjects before observing the ECLS-K test scores of the students.

¹²While this comparison cannot rule out that teachers' relative evaluations of the skills of boys and girls may be driving the changes in *noncognitive skills specifically*, it does narrow the range of possible manifestations of confounding changes in subjective evaluations, since any changes that would affect subjective cognitive ratings are ruled out.

displays present a problem for interpreting the changes in gender gaps for any individual measure separately, as the skills being measured are also captured to some degree by the other noncognitive measures. Factor analysis presents a suitable solution to this issue by reducing dimensions and creating orthogonal latent factors based on the measures' correlation matrix. To illustrate, let X be a $5 \times n$ vector of the five noncognitive skills, let Θ be an $n \times f$ vector of latent factors, where f is the number of latent factors, let Λ be a $5 \times f$ factor loading matrix, and let U be a $5 \times n$ matrix of the idiosyncratic error terms, also known as the uniqueness matrix. I then have:

$$X = \Theta\Lambda' + U \quad (2.1)$$

The goal of factor analysis is to (1) separate out the communally explained variation in $\Theta\Lambda'$ from U and (2) to then create estimates of Θ and Λ , which are unobserved, using eigenvector decomposition of $\text{corr}(X) - U$.¹³

Table B.4 presents the results of the unrotated principal factor analysis on the five noncognitive measures across all grades and cohorts. This choice of the unrotated form produces one common factor, which I will call “Latent Noncognitive Skill” for the remainder of the paper. I choose this one factor rotation for simplicity of interpretation because, even though Information Criterion support the use of two factors, results run with these two factors are similar to results run with only one.¹⁴ See Online Appendix 8.4 for results run with the two orthogonal factors. The factor loadings column of Table B.4 shows estimates of Λ from equation 2.1 and the uniqueness column shows estimates of U from equation 2.1. The scores column is calculated from the factor loadings and the correlation matrix, and shows the weights used in creating the Latent Noncognitive Skill variable as a weighted linear combination of the five noncognitive skills.¹⁵

The estimates in Table B.4 show that the Latent Noncognitive Skill factor is comprised of two thirds self control and interpersonal skills, one third externalizing behavior and approaches to learning, and only a small remaining portion coming from internalizing problems. The uniqueness

¹³There is no unique solution to step 2, as various orthogonal rotations can produce equally valid solutions once the uniqueness matrix U has been estimated.

¹⁴This two factor orientation is produced by an orthogonal varimax rotation with a Kaiser correction.

¹⁵The score matrix S is defined as $S = (\text{corr}(X))^{-1} \times \Lambda$ in orthogonal factor analysis.

column, which shows how much variation each measure has that is not explained by the common factor, tells a similar story. This follows directly from the correlation matrix in Table B.3, which shows that internalizing problems is the measure least correlated with the other noncognitive skill and thus has the most unique variance not shared by the latent factor. This means that the estimates of the changes in the Latent Noncognitive Skill variable in the remainder of paper will mostly exclude the decreased gender gap in internalizing problems, reflecting instead the persistent gap of the other four measures.

However, one downside to factor analysis is that it relies on variation between the five subjective noncognitive measures, without any anchoring to more concrete and objective measures. To confirm that the common variation the factor analysis is capturing in the latent factor variable is meaningful, I have separately regressed four different objective eighth grade outcome measures on both latent noncognitive skill individually and jointly on its five components as measured in fifth grade and shown their respective adjusted R^2 s in Appendix Table C.1. These four measures are eighth grade suspensions, grade retention, math test scores, and reading test scores¹⁶. By examining the degree of common variation between the latent noncognitive skill and these four measures, I can see the degree to which the variation I am keeping through my factor analysis procedure is related to more concrete outcomes, rather than uninformative variation such as measurement error. Comparing the results for latent noncognitive skill in column 1 and its five components and column 2, Table C.1 provides evidence that much of the correlation the five noncognitive measures collectively has with these four grade 8 behavioral and cognitive measures remains when combined into the singular factor variable, particularly with respect to suspensions.¹⁷ These results suggest that factor analysis is indeed capturing meaningful common variation between the subjective noncognitive measures when regressed with the more objective measures available in the ECLS-K datasets.

Applying the latent noncognitive skill to the estimates obtained so far, Figure A.1 and Table B.2 both display analyses for Latent Noncognitive Skill, and the results show that, as the weights that

¹⁶Eighth grade measures are only available for the 1998 cohort

¹⁷Bertrand and Pan (2013) focus on eighth grade suspensions in particular due to its link to longer term outcomes suggested in other literature.

produce it imply, it combines the trends of the first four noncognitive skills. As can be seen, this latent noncognitive skill continues to display an unchanged gender gap across all grades, confirming the endurance of this issue. The question remains whether any of the underlying correlates of the gender gap has changed. In order to uncover what may be leading to this endurance of gender gaps, the remaining analyses will show how correlates of this gender gap have evolved between the 1998 and 2010 cohorts.

2.4.3 The Diminishing Influence of Predictors

While Figure A.1 and Table B.2 establish that gender gaps across most noncognitive skills have remained, Table B.5 investigates how the influence of these gender gaps as explained by the kindergarten family background and kindergarten parental inputs measures examined in Bertrand and Pan (2013) have changed. To this end, I regress fifth grade latent noncognitive skill on both kindergarten family background and kindergarten parental inputs measures with a full set of cohort and female interaction terms and controls for race and school locale, then I generate predicted values from this regression. I show these predicted gaps in the first column of Table B.5. This first column of predicted gender gaps closely mirrors the findings in Table B.2 of persistently large fifth grade gender gaps of 0.55 standard deviations across both cohorts. The next two columns, however, do display a change in the role of these predictors.

The second and third columns of Table B.5 display Oaxaca-Blinder decompositions for latent noncognitive skill in fifth grade. The Oaxaca-Blinder decomposition breaks down the fifth grade gender gaps into the portions of this gap that are either unexplained or explained by differences in levels of observables, where the covariates included for this analysis are the kindergarten family background, kindergarten parental inputs, racial demographics controls, and kindergarten school locale controls used throughout this paper. This decomposition is shown in equations 2.2 and 2.3 below. In equation 2.2, let y_i be fifth grade latent noncognitive skill for individual i , f_i an indicator for female, and X_i a vector of family background, parental inputs, racial demographics, and school locale controls. Equation 2.3 shows how I can decompose the gender gap in fifth grade latent noncognitive skill using equation 2.2. I have shown both possible Oaxaca-Blinder specifications in

panels A and B.

$$y_i = \beta_0 + \beta_f f_i + \beta_G(f_i \times X_i) + \beta_B((1 - f_i) \times X_i) + u_i \quad (2.2)$$

Let $\bar{X}_G = \overline{f_i \times X_i}$ and let $\bar{X}_B = \overline{(1 - f_i) \times X_i}$. Equation 2.3 shows the specification in panel A, which is derived by adding and subtracting $\beta_G \bar{X}_B$ and factoring out.¹⁸

$$E(y|j = G) - E(y|j = B) = \underbrace{\beta_f + [\beta_G - \beta_B]\bar{X}_B}_{\text{Unexplained}} + \underbrace{\beta_G[\bar{X}_G - \bar{X}_B]}_{\text{Role of X's}} \quad (2.3)$$

This Oaxaca-Blinder decomposition in Table B.5 suggests that the influence of levels of family background and parental inputs has slightly declined. Across both specifications, the difference in the portion of the gaps explained by levels is negative. Further, in panel B, there is a statistically significant decrease ($p < 0.05$) in the portion of the gender gap that is explained by the differing levels of family background and parental inputs between genders of 0.05 standard deviations. This suggests that the influence of these predictors in explaining the large gender gaps in noncognitive skill has decreased, even while the gaps themselves remain unchanged.

2.4.4 Changes in Levels of SES and Inputs

This slight decrease in the portion of the gap explained by family background and parental inputs leaves open the question of which of these predictors may have changed. To explore this further, I have broken down the levels of these predictors by gender and cohort, showing results for family background in Table B.6 and results for parental inputs in Table C.2.

Table B.6 looks at different levels of family background characteristics by gender and cohort. While gender¹⁹ is randomly assigned at conception, it is possible that in the presence of higher levels of difficult behavior from boys, family attributes, particularly family structure, may be negatively harmed by the psychic cost of raising a child with more behavioral problems. In Table B.6 I show

¹⁸Panel B is derived by adding and subtracting $\beta_B \bar{X}_G$, which gives us

$$E(y|j = G) - E(y|j = B) = \beta_f + [\beta_G - \beta_B]\bar{X}_G + \beta_B[\bar{X}_G - \bar{X}_B]$$

instead of Equation 2.3.

¹⁹More precisely, sex is assigned at conception, not gender. Practically speaking, the difference is likely minimal in the sample as a whole.

summary stats for all three family background characteristics by cohort and test whether there are detectable differences between genders along these measures. While the results for family structure and teen motherhood are consistent with random assignment by gender, socioeconomic status is not. F-tests shown in Table B.6 reject the null that the distribution of family socioeconomic status is equal across gender in either cohort. In the 1998 cohort, girls were in more socioeconomically advantaged households than boys, while in the 2010 cohort, the situation has reversed. As this change would likely decrease noncognitive gender gaps, because Bertrand and Pan (2013) show that lower family socioeconomic status at kindergarten is correlated with larger noncognitive gender gaps, it is probable that this reversal is part of the cause of the reduced influence of the predictors observed in Table B.5.

Table C.2 shows a breakdown by gender and cohort of each of the three kindergarten parental input measures and proxies: kindergarten HOME index, kindergarten Warmth index, and parent-reported spanking at kindergarten for both genders in both cohorts.²⁰ Additionally, Appendix Table C.2 reproduces Table C.2 for all components of the two indices. Like Bertrand and Pan (2013) and Baker and Milligan (2016), I find that parents spend more time on educational activities with girls than with boys in both cohorts, though less so for the 2010 cohort. For the 2010 and 1998 cohorts, girls have 0.07 and 0.16 standard deviations higher HOME indices, respectively, than boys. This change suggests that parents have moved relatively towards more equal levels of investment between the genders. The gender gap in parental warmth decreased from 0.11 to 0.00 standard deviations indicating a similar decline and a complete absence of any gender gap in reported parental Warmth for the 2010 cohort. Rates of spanking remain similar between the genders, though the levels of spanking for both genders have dropped. Overall, it appears that while girls still enjoy higher levels of parental investment, there was a substantial shift towards greater equality in reported parental inputs.

Together, the reversal of girls' SES advantage and the equalizing of parental inputs provide

²⁰Parental input measures are examined only at kindergarten to avoid issues of reverse causality. If we expect parents to respond endogenously either (1) to a higher psychic cost from parenting a child with externalizing behaviors or (2) with compensating or reinforcing behaviors in response to low observed levels of child noncognitive skill, then observed externalizing behaviors could be driving parental investment, rather than the other way around.

an explanation for the decreased influence of the two groups of measures on the noncognitive gender gap. Although the changes in the unexplained portion of the gender gap were collectively statistically undetectable, I now turn to examining whether any of the coefficients of the individual predictors changed between the cohorts.

2.4.5 Changes in Coefficients of SES and Warmth Index

To begin examining changes in coefficients, I display the regression results of the first column of Table B.5 in Appendix Table C.3, holding constant race and kindergarten school locale across gender and cohort. I estimate Appendix Table C.3 by regressing fifth grade latent noncognitive skill on parental inputs and family background characteristics at kindergarten, fully interacted with dummies for gender and cohort, and controls for race and school locale. The final column, which shows how gender differences in coefficients have changed between the two cohorts for each measure, shows that most gender differences in coefficients are unchanged, with two exceptions: kindergarten Warmth index and socioeconomic status. Both have statistically significant ($p < 0.05$) declines in the gender gaps in their coefficients between the two cohorts, with changes of over 0.13 standard deviations. For the remainder of this section, I explore further these two measures.

Starting with kindergarten socioeconomic status, I show the evolution of gender gaps across grades for each cohort and in each of the five quintiles in Figure A.2. Figure A.2 calculations are produced by regressing latent noncognitive skill in each grade on indicators for socioeconomic status quintile interacted with gender and cohort dummies, controlling for parental inputs at kindergarten, the remaining family background measures at kindergarten, race, and school locale at kindergarten. Table B.8 tests whether the gender gap for each quintile has changed significantly individually or jointly in fall-kindergarten, third grade, and fifth grade. As Figure A.2 and Table B.8 show, differences in gender gaps have narrowed for the middle of the distribution in later grades. Gender gaps are similar across quintiles at the start of kindergarten, but by third and fifth grade, gender gaps begin to decrease for lower quintiles and increase for higher quintiles: F-tests of whether differences in gender gaps between cohorts across all quintiles are jointly zero have p -values of 0.000 and 0.001 for third and fifth grade, respectively. This suggests that the influence of SES in

increasing gender gaps has declined between the two cohorts, despite continuing to play a role.

Next, for kindergarten Warmth index, I show the evolution of gender gaps across grades for each cohort in Figure A.3. Like Figure A.2, I produce Figure A.3 calculations by regressing latent noncognitive skill in each grade on indicators for kindergarten warmth index quintiles interacted with gender and cohort dummies, controlling for other parental inputs at kindergarten, family background measures at kindergarten, race, and school locale at kindergarten. Table B.9 tests whether the gender gap for each quintile of kindergarten Warmth index has changed significantly individually or jointly in fall-kindergarten, grade 3, and grade 5. The story here is similar as for socioeconomic status in Figure A.2 and Table B.8: compression of gender gaps across the distribution. The main difference here is that this compression is already visible in kindergarten, rather than occurring only in later grades.

In summary, it appears there may have been some compression in the influence of socioeconomic status and parental warmth on gender gaps across their distributions, particularly by fifth grade. Even when I control for race, school locale, other family background measures, and parental inputs, gender gaps for lower family socioeconomic status quintiles decreased between the two cohorts by fifth grade, while increasing for higher quintiles. A similarly robust phenomenon occurred for parental warmth at kindergarten quintiles, though these changes appear even beginning in kindergarten and only intensifying by fifth grade. These trends are both equalizing in isolation, in that boys and girls now show more similar responses to lower family socioeconomic status and parental warmth than before. However, even though these changes alone point in the direction of decreasing gender gaps, like the reversal of girls' socioeconomic status advantage and equalization of reported parental inputs shown in the previous section, this change is not enough to decrease the overall gap by fifth grade, and is instead offset by unexplained factors outside of the six parental inputs and family background measures.

2.5 Conclusion

Using two cohorts of the nationally representative ECLS-K datasets, I show that gender gaps in noncognitive skills remain substantial and substantially larger than gender gaps in test scores.

Combining the five noncognitive measures into one latent noncognitive measure using principal factor analysis, I then show that, by fifth grade, the influence of two groups of three parental input and family background measures at kindergarten on gender gaps has waned. I then show that this declining influence is likely due to two factors: (1) the distribution of socioeconomic status at kindergarten switching from favoring girls in the 1998 cohort to favoring boys in the 2010 cohort, and (2) a substantial equalizing of reported parental inputs between boys and girls. Additionally, looking at the coefficients of kindergarten socioeconomic status and Warmth index measures, I show that their influence, as well as their levels, has changed to be more favorable to boys. In both cases, being in the lower end of the distribution is less correlated with higher gender gaps and being in the higher end of the distribution is less correlated with lower gender gaps, particularly in later grades. However, despite these changes in levels and coefficients, I find that changes in other unexplained factors of the noncognitive gender gap are what are keeping these differences intact.

Although much of this paper has focused on what has changed, it is worth re-emphasizing what has not. Gender gaps in noncognitive skills remain substantially large, and family structure continues to play a prominent role. There may be a decrease in the role of low socioeconomic status in these gaps, but it does continue to increase gender gaps relative to higher socioeconomic status. As differing levels of parental input between genders fades away as potential cause, the remaining differential responses to adverse family background conditions becomes more important than ever to study. This study makes clear that gender gaps in noncognitive skills have not dissipated to any meaningful degree, and without a further understanding of how family background characteristics play such an important role, policy makers will have difficulty closing them for the student cohorts to come.

BIBLIOGRAPHY

- Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments. *Unpublished Manuscript, Arizona State University*.
- Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2020). It takes a village: the economics of parenting with neighborhood and peer effects. Technical report, National Bureau of Economic Research.
- Agostinelli, F. and G. Sorrenti (2018). Money vs. time: family income, maternal labor supply, and child development. *University of Zurich, Department of Economics, Working Paper* (273).
- Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11.
- 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.
- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics* 11(3), 338–81.
- Autor, D. H., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2020). Males at the tails: How socioeconomic status shapes the gender gap. *NBER Working Paper* (w27196).
- Baker, M. and K. Milligan (2016). Boy-girl differences in parental time investments: Evidence from three countries. *Journal of Human Capital* 10(4), 399–441.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences* 6(6), 248–254.
- Baron-Cohen, S. (2003). *The Essential Difference: Men, Women, and the Extreme Male Brain*. London: Allan Lane.
- Baum-Snow, N., D. A. Hartley, and K. O. Lee (2019). The long-run effects of neighborhood change on incumbent families.
- Becker, G. S., W. H. Hubbard, and K. M. Murphy (2010). The market for college graduates and the worldwide boom in higher education of women. *American Economic Review* 100(2), 229–33.
- 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.
- Bibler, A. (2018). Household composition and gender differences in parental time investments.

Available at SSRN 3192649.

- Black, S. E. and P. J. Devereux (2010). Recent developments in intergenerational mobility. *New This Week*, 2 – 90.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics* 93(2), 455–467.
- Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* 81(3), 855–882.
- Cascio, E. U. and D. W. Schanzenbach (2016). First in the class? age and the education production function. *Education Finance and Policy* 11(3), 225–250.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *The American Economic Review* 104(9), 2633 – 2679.
- Chetty, R., N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi (2016). Childhood environment and gender gaps in adulthood. *American Economic Review* 106(5), 282–88.
- Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review* 108(10), 3028–56.
- Cornwell, C., D. B. Mustard, and J. Van Parys (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources* 48(1), 236–264.
- 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.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources* 42(3), 528–554.
- Demaray, M. K., S. L. Ruffalo, J. Carlson, R. Busse, A. E. Olson, S. M. McManus, and A. Leventhal (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review* 24(4), 648–671.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics* 132(4), 1593–1640.
- Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School choice, school quality,

- and postsecondary attainment. *The American Economic Review* 104(3), 991 – 1013.
- DiPrete, T. A. and J. L. Jennings (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research* 41(1), 1–15.
- Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, et al. (2007). School readiness and later achievement. *Developmental psychology* 43(6), 1428.
- Elder, T. E. (2010). The importance of relative standards in adhd diagnoses: evidence based on exact birth dates. *Journal of Health Economics* 29(5), 641–656.
- Elder, T. E. and D. H. Lubotsky (2009). Kindergarten entrance age and children’s achievement impacts of state policies, family background, and peers. *Journal of Human Resources* 44(3), 641–683.
- Elliott, S. N., F. M. Gresham, T. Freeman, and G. McCloskey (1988). Teacher and observer ratings of children’s social skills: Validation of the social skills rating scales. *Journal of Psychoeducational Assessment* 6(2), 152–161.
- Fernández, A. B. (2021). Neighbors’ effects on university enrollment. *American Economic Journal: Applied Economics* (forthcoming).
- Fortin, N. M., P. Oreopoulos, and S. Phipps (2015). Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources* 50(3), 549–579.
- Gensowski, M., R. Landersø, B. Dorthe, P. Dale, A. Højen, and L. Justice (2020). Public and parental investments and children’s skill formation. *The ROCKWOOL Foundation Research Unit* (155).
- Goldin, C., L. F. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives* 20(4), 133–156.
- Heckman, J. J. and R. Landersø (2021). Lessons from denmark about inequality and social mobility. Technical report, National Bureau of Economic Research.
- Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics* 6(1), 689–733.
- 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.
- Imberman, S. A., A. D. Kugler, and B. I. Sacerdote (2012). Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review* 102(5), 2048–82.

- Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes : Evidence from school finance reforms. *The Quarterly Journal of Economics* 131(1), 157 – 218.
- Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review* 21(6), 589–598.
- Johann, A. (2020). The increasing fragility of boys: Examining changes in levels and correlates of gender gaps in noncognitive skills over time. *Available at <https://sites.google.com/site/alwjohann/working-papers/gender-gaps-in-noncognitive-skills>*.
- Kling, J. R., J. Ludwig, and L. F. Katz (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics* 120(1), 87–130.
- Knickmeyer, R., S. Baron-Cohen, P. Raggatt, and K. Taylor (2005). Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology and Psychiatry* 46(2), 198–210.
- Laliberté, J.-W. P. (2018). Long-term contextual effects in education: Schools and neighborhoods. *University of Calgary, unpublished manuscript*.
- Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics* 3(1), 101–28.
- List, J. A., F. Momeni, and Y. Zenou (2020). The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods. Technical report, National Bureau of Economic Research.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3), 531–542.
- Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics* 92(3), 562–576.
- Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The quarterly journal of economics* 118(4), 1533–1575.
- Raver, C., P. W. Garner, and R. Smith-Donald (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In *School readiness and the transition to kindergarten in the era of accountability*, pp. 121–147. Paul H Brookes Publishing.
- Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics* 6(1), 253–272.

- Sanbonmatsu, L., L. F. Katz, J. Ludwig, L. A. Gennetian, G. J. Duncan, R. C. Kessler, E. K. Adam, T. McDade, and S. T. Lindau (2011). Moving to opportunity for fair housing demonstration program: Final impacts evaluation.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518 – 529.
- Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* 113(485), F3–F33.
- Tourangeau, K., J. Burke, T. Le, S. Wan, M. Weant, E. Brown, N. Vaden-Kiernan, E. Rinker, R. Dulaney, K. Ellingsen, B. Barrett, I. Flores-Cervantes, N. Zill, J. Pollack, D. Rock, S. Atkins-Burnett, and S. Meisels (2001). *ECLS-K Base Year Public-Use Data Files and Electronic Codebook*. Washington, DC: National Center for Education Statistics: U.S. Department of Education. (NCES 2001-029).

APPENDIX A

FIGURES

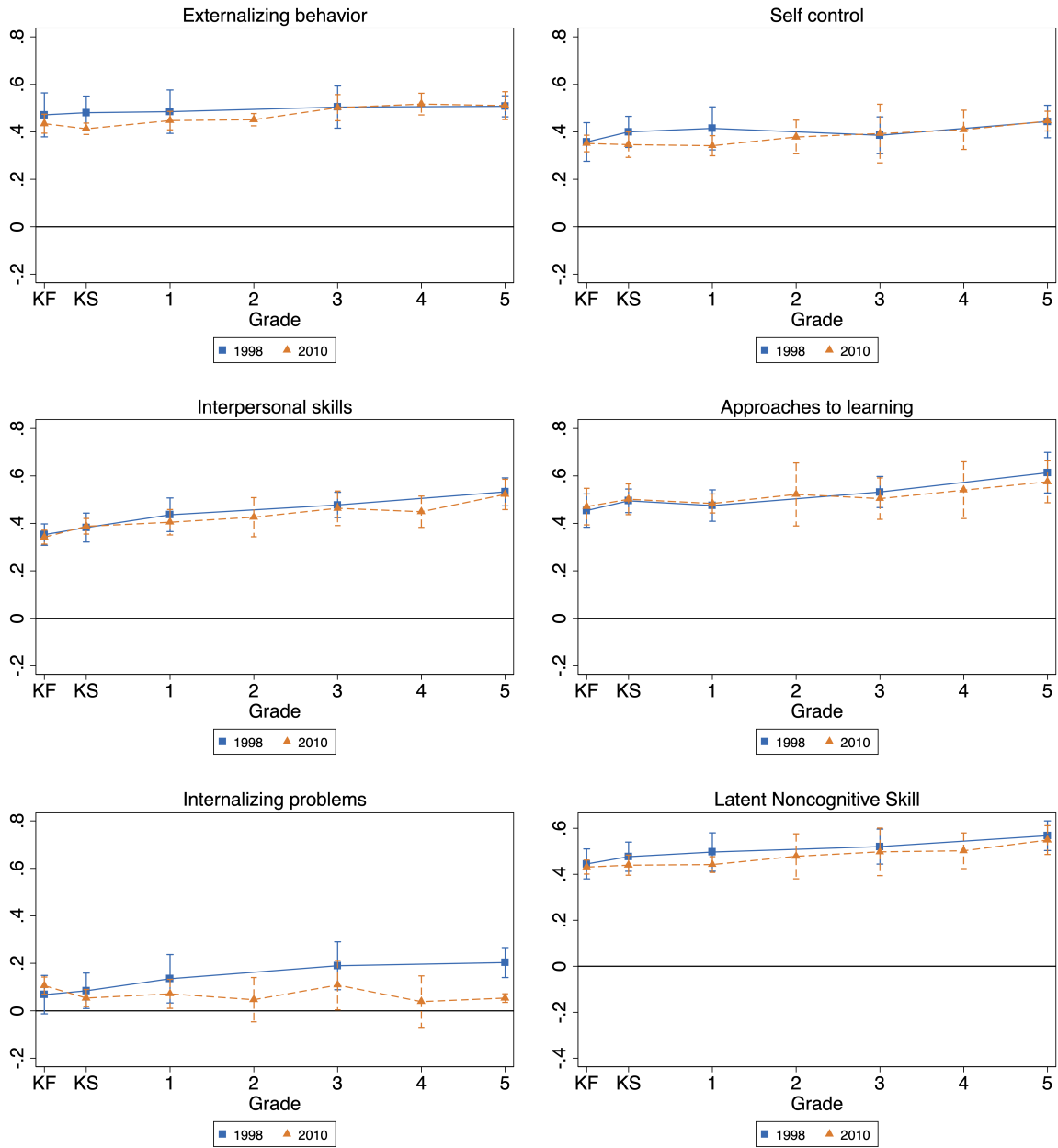


Figure 2.A.1 Female-Male Gaps in Teacher Ratings of Noncognitive Skills

Notes: Each graph shows the coefficient on a female dummy from a regression of each respective teacher-reported noncognitive skill in each respective grade on a female dummy variable. KF refers to the fall of kindergarten, KS refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

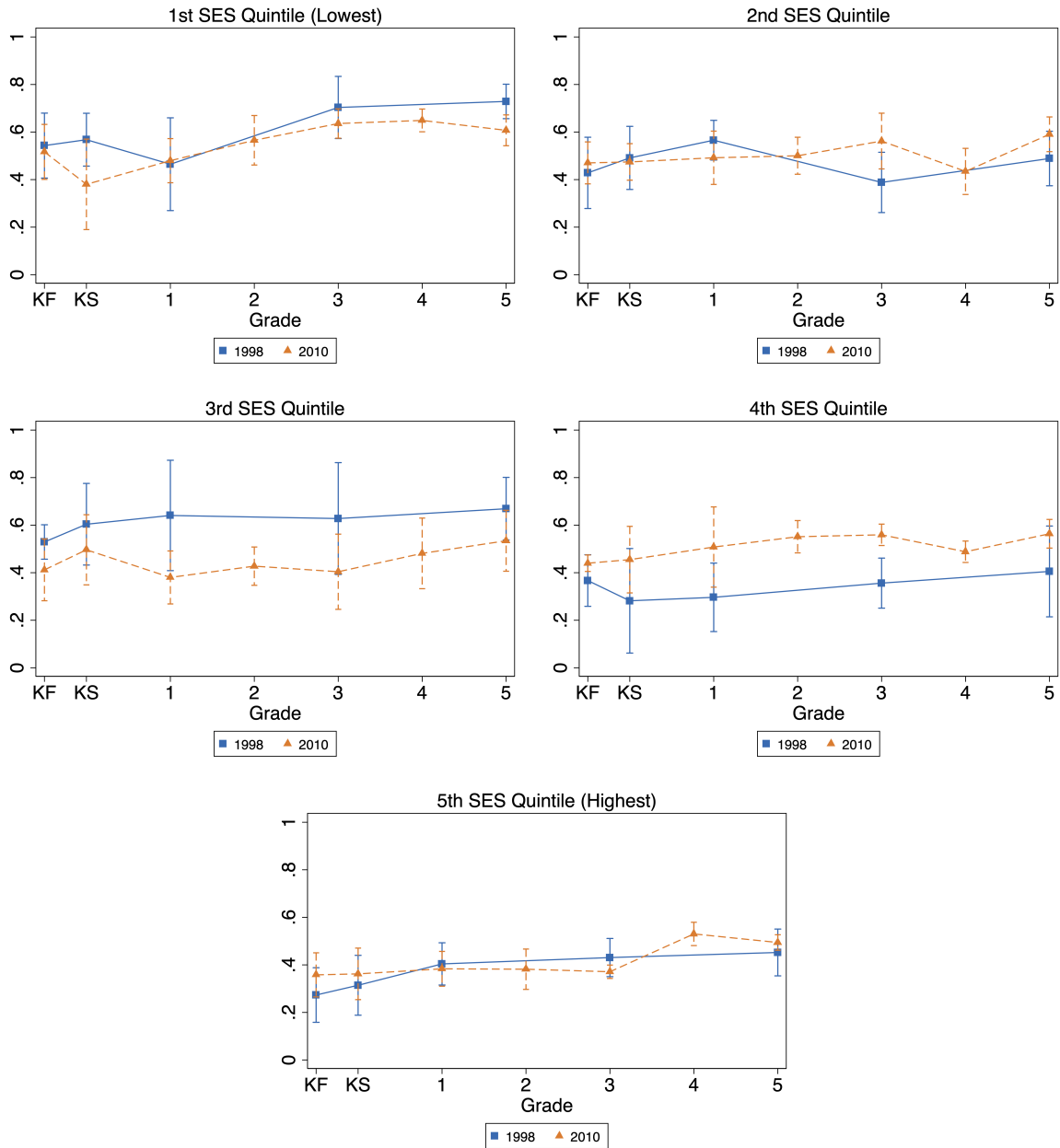


Figure 2.A.2 Female-Male Gaps in Latent Noncognitive Skill, By SES at Kindergarten

Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by SES quintile interaction term, and a female by 2010 cohort by SES quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female by SES quintile interaction term (for the 1998 estimates) from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

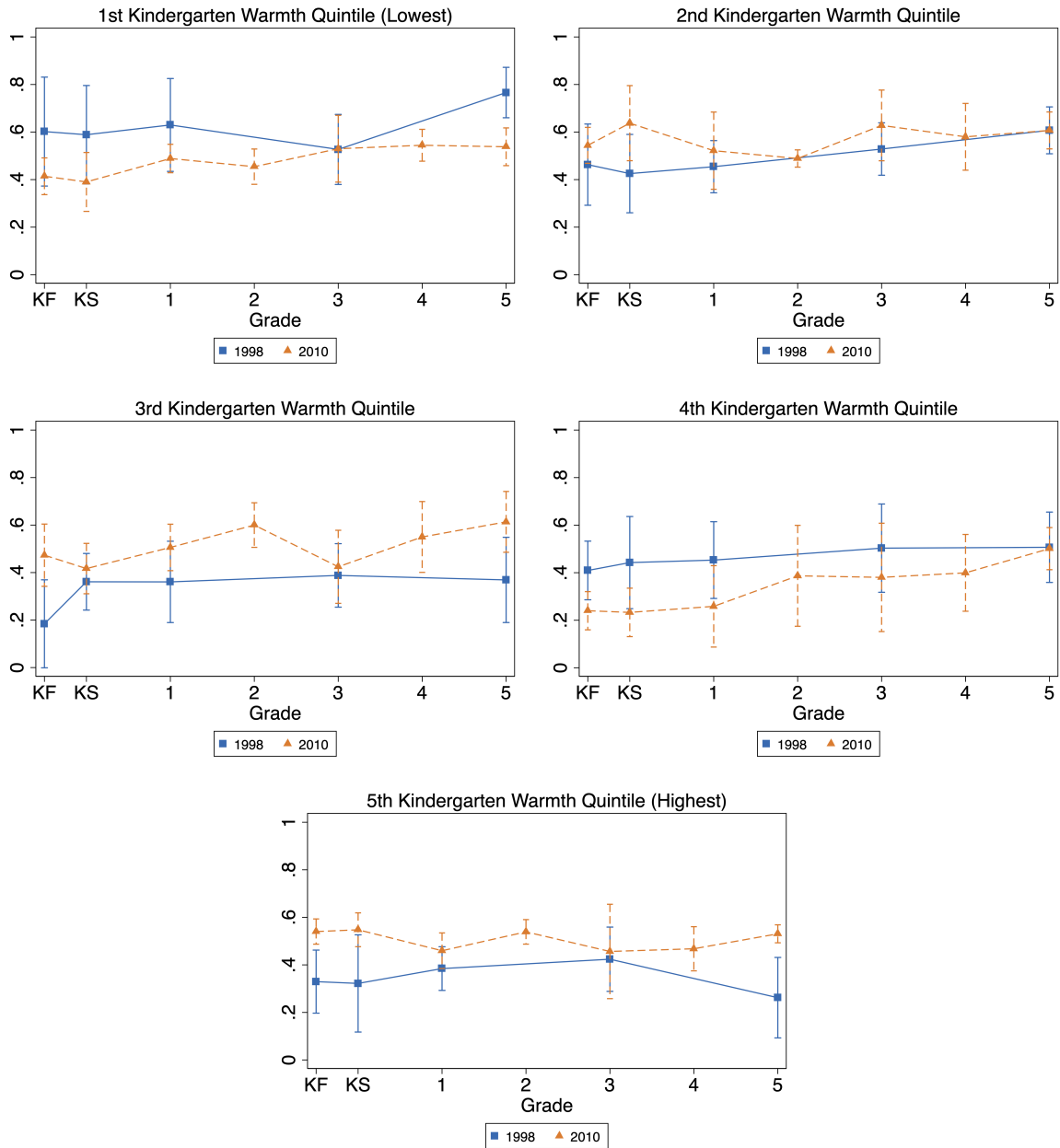


Figure 2.A.3 Female-Male Gaps in Latent Noncognitive Skill, By Kindergarten Warmth Index

Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by Warmth index quintile interaction term, and a female by 2010 cohort by Warmth index quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female by Warmth index quintile interaction term (for the 1998 estimates) from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of Warmth index in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, socioeconomic status at kindergarten, HOME index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

APPENDIX B TABLES

Table 2.B.1 Sample Summary Statistics

Variable	2010 Cohort Mean	1998 Cohort Mean
White	0.58	0.61
Black	0.12	0.14
Hispanic	0.22	0.18
Asian	0.03	0.03
Other race/ethnicity	0.06	0.05
Female	0.49	0.49
School locale: City	0.30	0.36
School locale: Suburbs	0.35	0.42
School locale: Town or Rural	0.35	0.22
1st SES quintile (lowest)	0.14	0.17
2nd SES quintile	0.20	0.21
3rd SES quintile	0.23	0.20
4th SES quintile	0.21	0.21
5th SES quintile (highest)	0.21	0.21
Parents' Highest Education: Less than HS	0.10	0.09
Parents' Highest Education: High School	0.21	0.27
Parents' Highest Education: Some college	0.34	0.33
Parents' Highest Education: College or greater	0.35	0.31
Age first birth < 20	0.21	0.25
Age first birth \geq 20 and < 30	0.56	0.59
Age first birth \geq 30	0.22	0.16
Single mom	0.18	0.20
Both biological parents	0.73	0.69
Other family structure	0.09	0.11

Notes: Each cell shows the weighted mean of each variable in each respective dataset. Column 2 shows the means for the 2010 cohort, in the ECLS-K:2011 data, and column 3 shows the means for the 1998 cohort, in the ECLS-K data dataset. Sample restrictions are imposed as described in text. Fifth grade parent panel weights are used for each calculation.

Table 2.B.2 Changes in Female-Male Gaps in Teacher Ratings of Noncognitive Skills

Variable	Fall-K	Spring-K	Grade 1	Grade 3	Grade 5	Joint test <i>p</i> -value
Panel A: Unadjusted						
Externalizing behavior	-0.037 [0.050]	-0.068+ [0.037]	-0.038 [0.049]	-0.002 [0.051]	0.003 [0.035]	0.181
Self control	-0.006 [0.044]	-0.054 [0.041]	-0.073 [0.049]	0.007 [0.069]	0.002 [0.039]	0.215
Interpersonal skills	-0.011 [0.026]	0.006 [0.034]	-0.031 [0.043]	-0.014 [0.043]	-0.011 [0.042]	0.937
Approaches to learning	0.017 [0.050]	0.007 [0.039]	0.009 [0.038]	-0.028 [0.052]	-0.038 [0.059]	0.734
Internalizing problems	0.038 [0.044]	-0.031 [0.041]	-0.064 [0.058]	-0.081 [0.070]	-0.150** [0.033]	0.000
Latent Noncognitive Skill	-0.014 [0.035]	-0.037 [0.038]	-0.054 [0.045]	-0.023 [0.061]	-0.019 [0.043]	0.531
Panel B: Adjusted						
Externalizing behavior	-0.022 [0.052]	-0.047 [0.039]	-0.023 [0.048]	0.025 [0.044]	0.030 [0.032]	0.400
Self control	0.008 [0.041]	-0.032 [0.045]	-0.053 [0.045]	0.036 [0.058]	0.031 [0.036]	0.345
Interpersonal skills	0.011 [0.024]	0.030 [0.030]	-0.006 [0.042]	0.013 [0.031]	0.018 [0.035]	0.926
Approaches to learning	0.039 [0.044]	0.035 [0.038]	0.037 [0.036]	0.002 [0.040]	-0.006 [0.057]	0.473
Internalizing problems	0.046 [0.043]	-0.016 [0.044]	-0.046 [0.058]	-0.065 [0.061]	-0.131** [0.030]	0.000
Latent Noncognitive Skill	0.010 [0.030]	-0.010 [0.039]	-0.030 [0.041]	0.007 [0.047]	0.014 [0.039]	0.856

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets

Notes: Each cell shows the coefficient on an interaction term for female and 2010 data with the row measure in each column grade as the left hand side variable. The last column displays the *p*-value from a joint F-test of the null that the differences across all grades for each measure are zero. Teacher ratings and test scores are standardized to have a mean of zero and standard deviation one in the population based on weighting and sampling methodology correction after imposing the sample restrictions, with additional correction for reference bias. Regressions in panel B include controls for race, school locale, family background, and parental inputs as reported at kindergarten. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.B.3 Noncognitive Skills Correlation Matrix

	Externalizing	Self control	Interpersonal skills	Approaches to learn	Internalizing
Externalizing behavior	1.000	0.726	0.621	0.587	0.295
Self control	0.726	1.000	0.803	0.687	0.305
Interpersonal skills	0.621	0.803	1.000	0.715	0.345
Approaches to learning	0.587	0.687	0.715	1.000	0.371
Internalizing problems	0.295	0.305	0.345	0.371	1.000

Notes: Results are shown from a weighted correlation matrix of all five standardized noncognitive skills across all grades and cohorts. Fifth grade parent panel weights are used for this calculation.

Table 2.B.4 Factor Loadings, Scores, and Uniqueness

	Eigenvalue 2.879	Proportion Explained 1.064	
Noncog Variables	Factor Loadings	Factor Scores	Uniqueness
Externalizing behavior	0.754	0.161	0.422
Self control	0.888	0.380	0.197
Interpersonal skills	0.863	0.296	0.255
Approaches to learning	0.787	0.199	0.368
Internalizing problems	0.397	0.058	0.811

Notes: Results are shown from an unrotated principal factor analysis of all five standardized noncognitive skills across all grades and cohorts. Fifth grade parent panel weights are used for this calculation. Results for further factors are not displayed due to low eigenvalues.

Table 2.B.5 Oaxaca-Blinder Decomposition of Fifth Grade Gender Gaps

Cohort	Predicted Gender Gap (girls – boys)	Unexplained	Due to Levels
Panel A: Boys' X's, Girls' Betas			
2010	0.549** [0.035]	0.553** [0.032]	-0.005 [0.010]
1998	0.558** [0.042]	0.542** [0.039]	0.015 [0.016]
Difference	-0.009 [0.055]	0.011 [0.050]	-0.020 [0.019]
Panel B: Girls' X's, Boys' Betas			
2010	0.549** [0.034]	0.555** [0.033]	-0.006 [0.015]
1998	0.558** [0.044]	0.512** [0.038]	0.045* [0.020]
Difference	-0.009 [0.055]	0.043 [0.051]	-0.052* [0.025]

** p<0.01, * p<0.05, + p<0.1. Bootstrapped standard errors in brackets

Notes: The Oaxaca-Blinder decompositions shown here are performed as described in text. Gender gaps as reported in the first column are the predicted gender gap from a regression of each measure on family background, parental input, racial demographics, and school locale measures as reported at kindergarten interacted separately by cohort and gender. Standard errors are bootstrapped with 100 replications, with each row's estimates produced jointly in each bootstrapping iteration. Sample restrictions are imposed as described in text. Fifth grade parent panel weights are used for these estimates.

Table 2.B.6 Kindergarten Family Background Characteristics, By Gender and Cohort

Variable	2010 Cohort			1998 Cohort			Diff-in-Diff
	Girls	Boys	Difference	Girls	Boys	Difference	
1st SES quintile (lowest)	0.144 [0.021]	0.141 [0.024]	0.002 [0.007]	0.151 [0.007]	0.172 [0.024]	-0.022 [0.021]	0.024 [0.022]
2nd SES quintile	0.213 [0.013]	0.193 [0.008]	0.020** [0.007]	0.203 [0.013]	0.219 [0.011]	-0.016 [0.011]	0.036** [0.013]
3rd SES quintile	0.223 [0.009]	0.237 [0.005]	-0.013+ [0.007]	0.199 [0.008]	0.209 [0.022]	-0.010 [0.027]	-0.003 [0.028]
4th SES quintile	0.204 [0.016]	0.218 [0.014]	-0.013** [0.005]	0.227 [0.009]	0.190 [0.008]	0.036** [0.013]	-0.050** [0.014]
5th SES quintile (highest)	0.215 [0.009]	0.212 [0.014]	0.004 [0.008]	0.221 [0.013]	0.210 [0.011]	0.011 [0.012]	-0.007 [0.014]
F-test jointly zero <i>p</i> -value			0.000			0.008	0.000
Age first birth < 20	0.219 [0.018]	0.210 [0.012]	0.009 [0.009]	0.239 [0.020]	0.253 [0.012]	-0.014 [0.017]	0.023 [0.019]
More than 20 years old	0.781 [0.018]	0.790 [0.012]	-0.009 [0.009]	0.761 [0.020]	0.747 [0.012]	0.014 [0.017]	-0.023 [0.019]
F-test jointly zero <i>p</i> -value			0.308			0.429	0.241
Single mom	0.181 [0.005]	0.187 [0.013]	-0.005 [0.009]	0.193 [0.011]	0.200 [0.009]	-0.007 [0.008]	0.001 [0.012]
Both biological parents	0.728 [0.005]	0.729 [0.016]	-0.001 [0.018]	0.697 [0.011]	0.689 [0.012]	0.007 [0.012]	-0.008 [0.021]
Other family structure	0.090 [0.007]	0.084 [0.006]	0.006 [0.010]	0.110 [0.007]	0.111 [0.011]	-0.001 [0.008]	0.007 [0.013]
F-test jointly zero <i>p</i> -value			0.201			0.704	0.854

Robust standard errors in brackets

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$

Notes: Columns 1-2 and 4-5 show the means of each row measure for each gender in the 2010 and 1998 cohort, respectively. Columns 3 and 6 show the difference between coefficients in columns 1-2 and 4-5, respectively. Column 7 shows the difference between columns 3 and 6. Significance stars are only included in columns 3, 6, and 7.

Estimates are calculated by regressing each grouping of measures simultaneously (using SUR) on a female dummy, a 2010 cohort dummy, and a female by 2010 cohort dummy. The final rows of each section shows the *p*-value from a joint F-test of the null that the coefficients from a regression on a female dummy in each cohort on all listed measures are jointly zero. Sample is restricted as reported in the text. Observations are weighted using fifth grade parent panel weights for the 1998 cohort and fifth grade panel weights for the 2010 cohort. Standard errors are heteroskedasticity robust and clustered at the primary sampling unit level.

Table 2.B.7 Kindergarten Parental Inputs, By Gender and Cohort

Variable	2010 Cohort			1998 Cohort			Diff-in-Diff
	Girls	Boys	Difference	Girls	Boys	Difference	
Kindergarten HOME index	0.034 [0.026]	-0.033 [0.025]	0.067+ [0.036]	0.082 [0.030]	-0.082 [0.031]	0.163** [0.043]	-0.096+ [0.056]
Kindergarten Warmth index	-0.002 [0.025]	0.002 [0.024]	-0.004 [0.035]	0.058 [0.027]	-0.058 [0.035]	0.116** [0.044]	-0.120* [0.056]
Spanked child last week, kindergarten	0.149 [0.009]	0.169 [0.009]	-0.020 [0.013]	0.260 [0.013]	0.276 [0.014]	-0.015 [0.019]	-0.004 [0.023]

Robust standard errors in brackets

** p<0.01, * p<0.05, + p<0.1

Notes: Columns 1-2 and 4-5 show the means of each row measure for each gender in the 2010 and 1998 cohort, respectively. Columns 3 and 6 show the difference between coefficients in columns 1-2 and 4-5, respectively. Column 7 shows the difference between columns 3 and 6. Significance stars are only included in columns 3, 6, and 7. Estimates are calculated by regressing each row measure on a female dummy, a 2010 cohort dummy, and a female by 2010 cohort dummy. The final rows of each section shows the p-value from a joint F-test of the null that the coefficients from a regression on a female dummy in each cohort on all listed measures are jointly zero. Sample is restricted as reported in the text. Observations are weighted using fifth grade parent panel weights for the 1998 cohort and fifth grade panel weights for the 2010 cohort. Standard errors are heteroskedasticity robust and clustered at the primary sampling unit level.

Table 2.B.8 Changes in Gender Gaps Between Cohorts, The Role of Socioeconomic Status at Kindergarten

Latent Noncognitive Skill In:	Fall-K	Grade 3	Grade 5
1st SES quintile (lowest)	-0.009 [0.086]	-0.060 [0.074]	-0.135** [0.049]
2nd SES quintile	0.050 [0.088]	0.169* [0.081]	0.122+ [0.073]
3rd SES quintile	-0.116 [0.079]	-0.232+ [0.139]	-0.144 [0.094]
4th SES quintile	0.084 [0.056]	0.204** [0.053]	0.166 [0.105]
5th SES quintile (highest)	0.099 [0.077]	-0.048 [0.046]	0.060 [0.054]
Joint F-test of no change p -value	0.253	0.000	0.001

Robust standard errors in brackets

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$

Notes: Each estimate shows the coefficients on a female by 2010 cohort by SES quintile interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.B.9 Changes in Gender Gaps Between Cohorts, The Role of Kindergarten Warmth Index

Latent Noncognitive Skill In:	Fall-K	Grade 3	Grade 5
1st Kindergarten Warmth quintile (lowest)	-0.182 [0.130]	-0.012 [0.101]	-0.232** [0.077]
2nd Kindergarten Warmth quintile	0.084 [0.105]	0.092 [0.095]	-0.001 [0.059]
3rd Kindergarten Warmth quintile	0.285* [0.113]	0.042 [0.106]	0.254* [0.119]
4th Kindergarten Warmth quintile	-0.158* [0.071]	-0.120 [0.150]	-0.000 [0.086]
5th Kindergarten Warmth quintile (highest)	0.178* [0.086]	0.003 [0.125]	0.265** [0.085]
Joint F-test of no change <i>p</i> -value	0.001	0.835	0.000

Robust standard errors in brackets

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$

Notes: Each estimate shows the coefficients on a female by 2010 cohort by kindergarten Warmth index quintile interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, SES at kindergarten, HOME index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

APPENDIX C ONLINE APPENDIX

C.1 Analysis Sample Creation

The full K-5 longitudinal panel sample described in Section 2.3.1 is defined as observations with non-missing and non-zero panel weights.¹ This full sample contains 8,370 observations for the 1998 cohort and 7,326 observations for the 2010 cohort. To get to the analysis sample, observations are dropped in three steps. First, I dropped observations if they have missing data on any of the five teacher-reported noncognitive measures in either (spring) kindergarten or fifth grade. This drops 1,410 observations from the 1998 cohort and 526 observations from the 2010 cohort. Second, I dropped observations if they were missing any basic demographic information, including gender, race/ethnicity, urbanicity, and parental education. This step drops less than 10 observations from the 1998 cohort and 140 observations from the 2010 cohort. Third, I dropped observations missing any family background or parental inputs variables, which include the Kindergarten HOME index, the Kindergarten Warmth index, spanked at kindergarten, family structure, mother's age at first birth, and kindergarten socioeconomic status. This last step drops 329 observations from the 1998 cohort and 1,722 observations from the 2010 cohort. The majority of the observations lost in the 2010 cohort are missing values for the parental input variables (Kindergarten HOME index, Kindergarten Warmth index, or spanked at kindergarten). Together, this process leaves 6,630 observations in the 1998 analysis sample and 4,938 observations in the 2010 analysis sample.

To investigate the possibility of nonrandom item nonresponse reflected by the first and third step, I reran my analysis using weights adjusted by the inverse probability of appearing in the final analysis sample among all longitudinal panel observations. These weights are calculated using a logit regression of an indicator for analysis sample membership on the demographic variables used in Step 2: gender, race/ethnicity, parental education, and urbanicity. Provided K-5 longitudinal panel weights were then multiplied by the inverse of these predicted probabilities to create inverse probability weights. All results were then rerun with these alternative weights, with no notable

¹The panel weight variables used are named C1_6FP0 for the 1998 and W9C19P_2T290 for the 2010 cohort in the corresponding ECLS-K manuals. These weight variables were generated by the ECLS-K survey administrators to correct for nonrandom attrition and other types of nonresponse bias between the kindergarten and grade 5 waves of the ECLS-K surveys.

differences. Tables and figures created using these weights are available upon request.

C.2 Appendix Figures

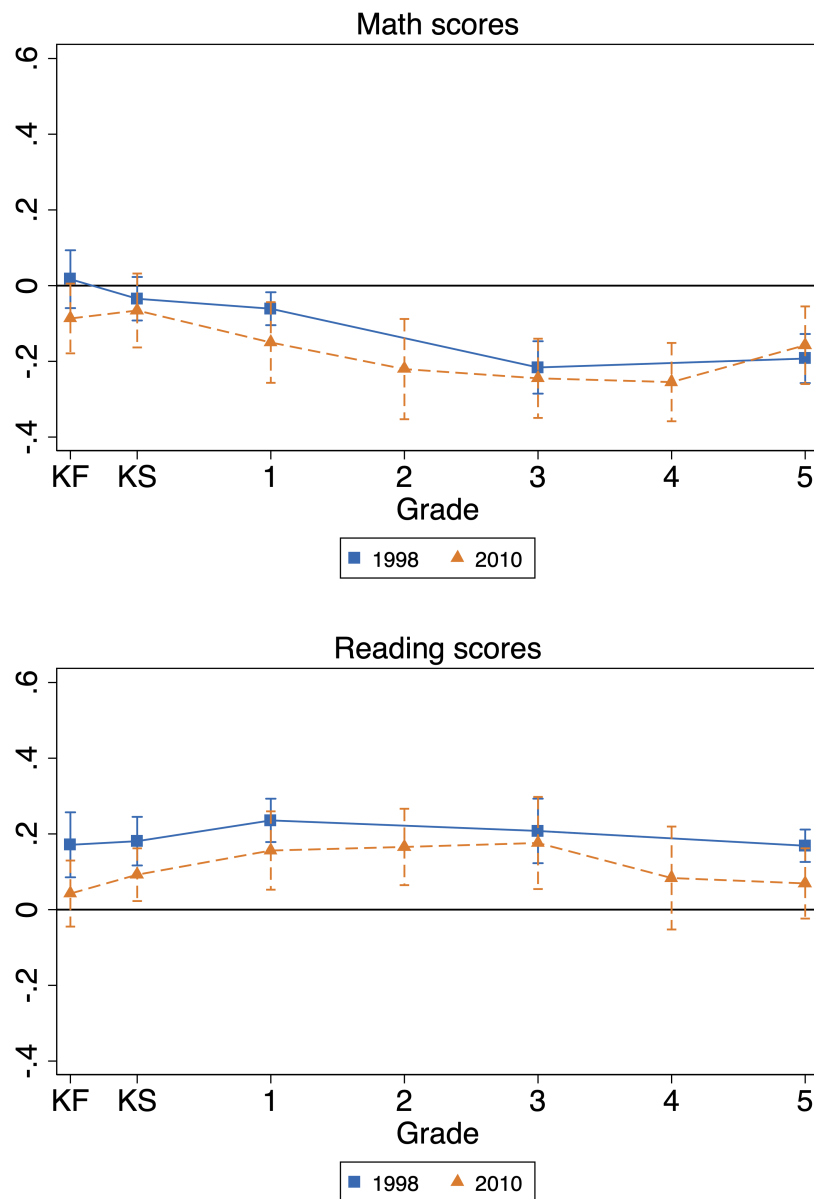


Figure 2.C.1 Female-Male Gaps in Test Scores

Notes: Each graph shows the coefficient on a female dummy from a regression of each test score in each respective grade on a female dummy variable. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

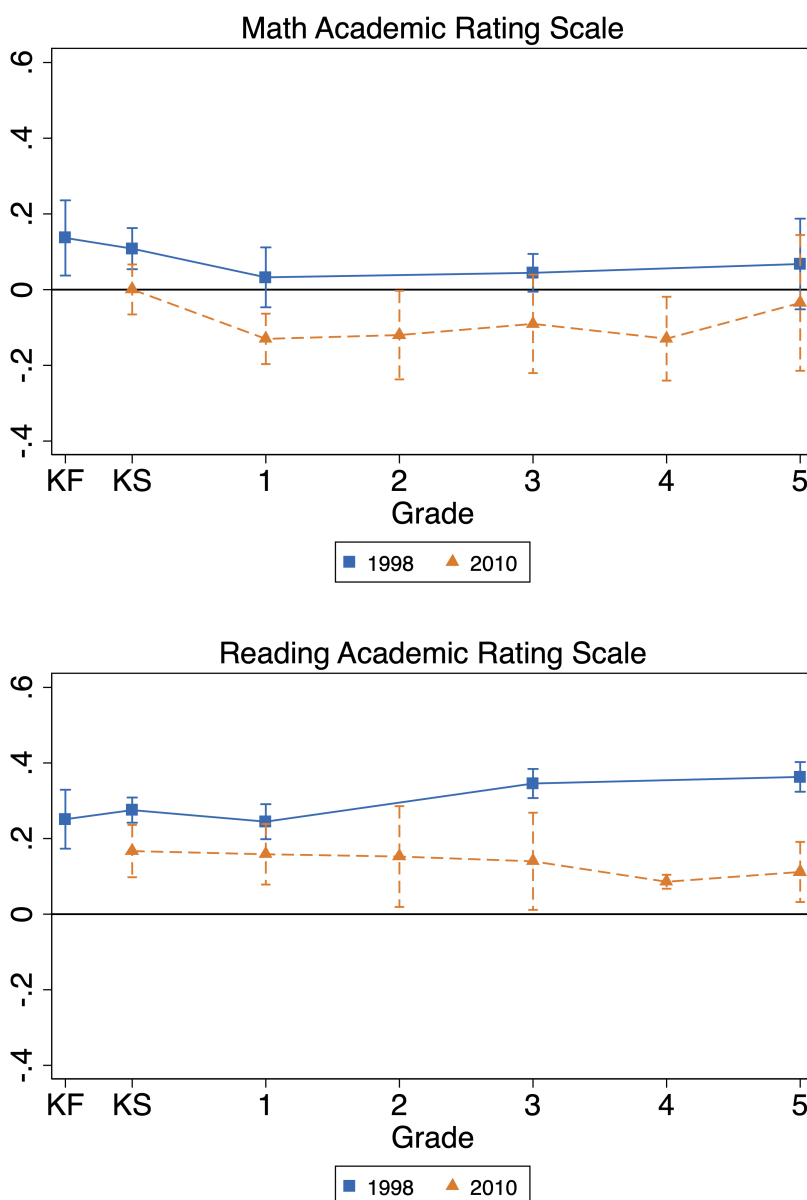


Figure 2.C.2 Female-Male Gaps in Teacher Cognitive Evaluations

Notes: Each graph shows the coefficient on a female dummy from a regression of each respective teacher-reported rating of cognitive ability in each respective grade on a female dummy variable. Academic Rating Scores were not reported for the 2010 Cohort Fall-Kindergarten survey wave. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

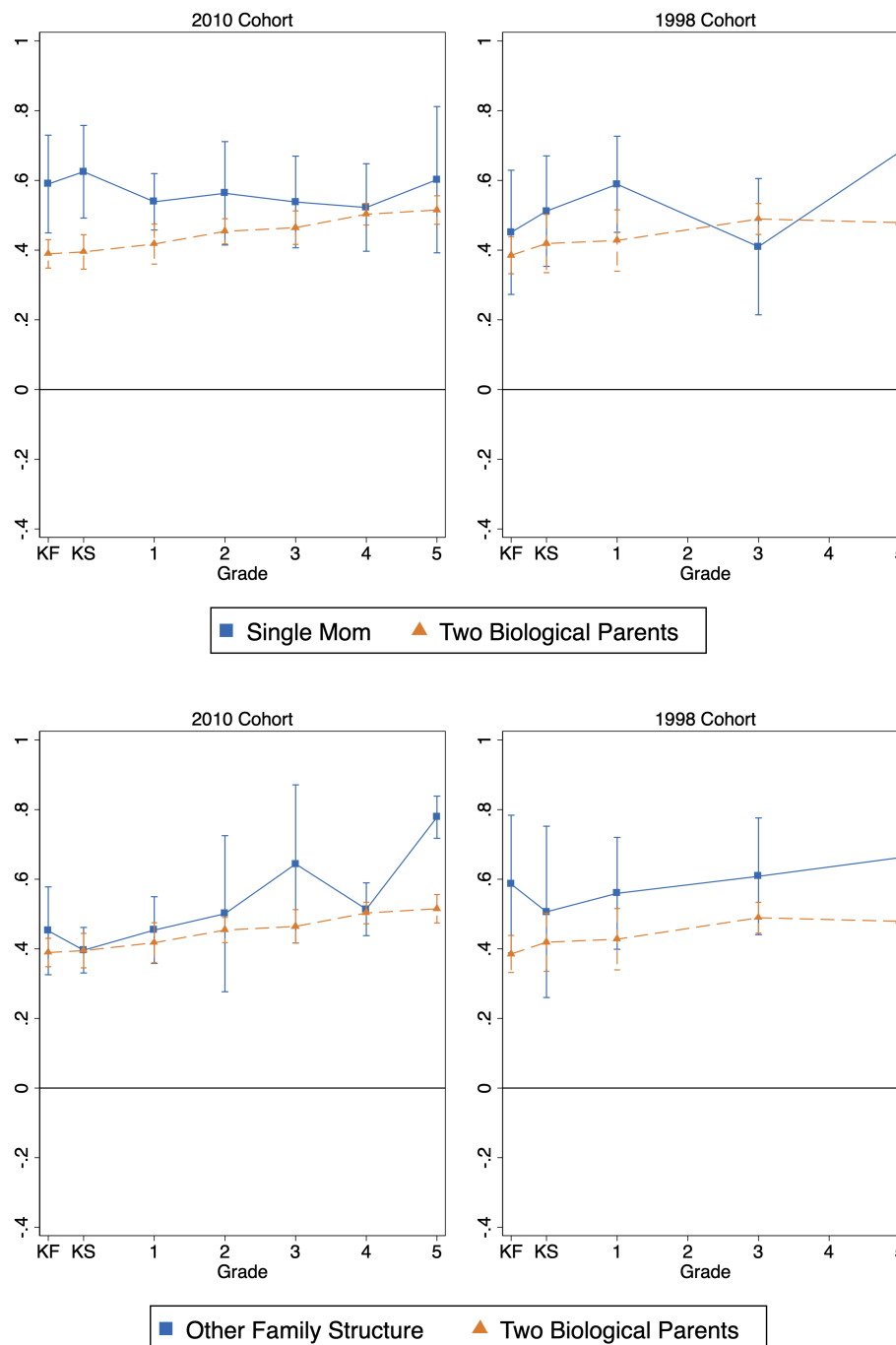


Figure 2.C.3 Female-Male Gaps in Latent Noncognitive Skill, By Kindergarten Family Structure

Notes: Estimates for each grade in all four graphs come from one regression of latent noncognitive skill in each grade with controls for teen motherhood, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. Separate estimates of gender gaps for each cohort and subgroup are produced using interaction terms for gender and cohort. Two biological parent estimates are the same in both rows. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions and reference bias corrections are imposed. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

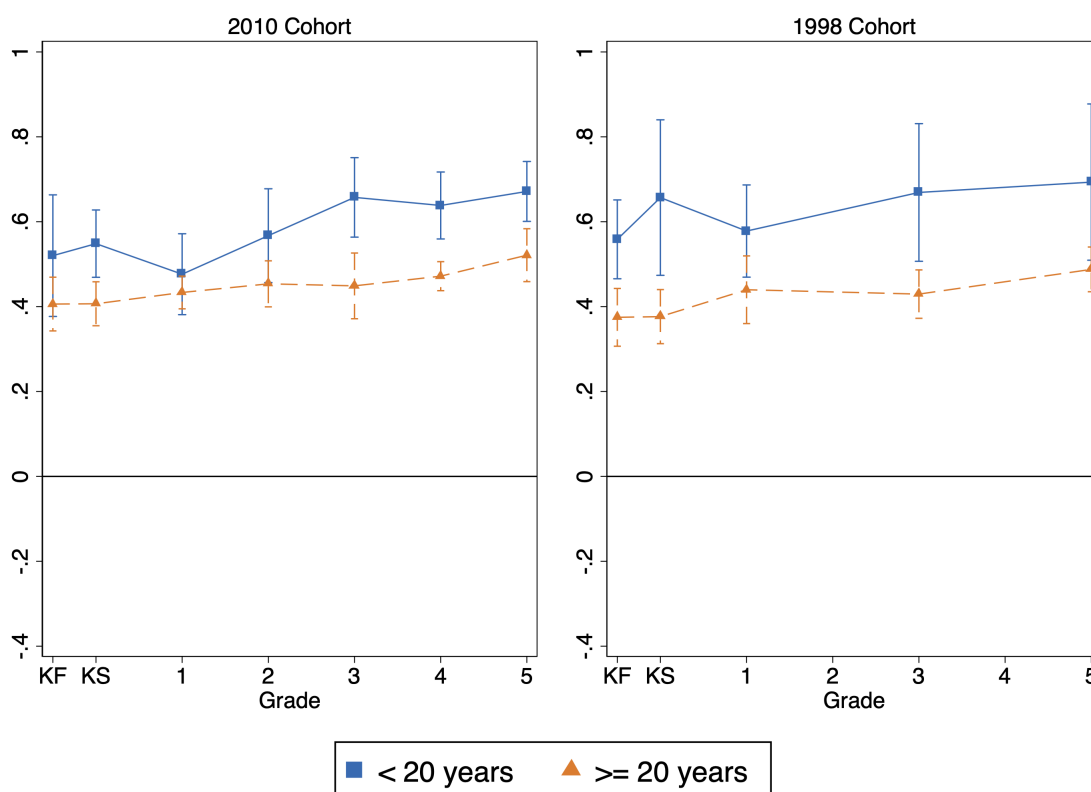


Figure 2.C.4 Female-Male Gaps in Latent Noncognitive Skill, By Mother's Age at First Birth

Notes: Estimates for each grade in both graphs come from one regression of latent noncognitive skill in each grade with controls for family structure at kindergarten, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. Separate estimates of gender gaps for each cohort and subgroup are produced using interaction terms for gender and cohort. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions and reference bias corrections are imposed. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

C.3 Appendix Tables

Table 2.C.1 Adjusted R^2 s from Regressions of Outcomes on Latent Noncognitive Skill or Components in 5th Grade

	Latent Noncognitive Skill	All Noncognitive Measures
Had any out-of-school suspensions by 8th Grade	0.176	0.201
Was held back by 8th Grade	0.042	0.064
Math scores in 8th Grade	0.086	0.142
Reading scores in 8th Grade	0.113	0.162

Notes: Each cell shows the adjusted R^2 from a regression of each row outcome on the noncognitive measure or measures listed in each column. The right-hand-side measure in column 1 is fifth grade latent noncognitive skill. The right-hand-side measures in column 2 are fifth grade externalizing behavior, self control, interpersonal skills, approaches to learning, and internalizing problems. Results only include observations from the 1998 cohort as there is no eighth grade data for the 2010 cohort, as well as no information on suspensions or disciplinary incidents. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.2 Kindergarten Parental Inputs, By Gender

Variable	2010 Cohort			1998 Cohort			
	Girls	Boys	Difference	Girls	Boys	Difference	
Kindergarten HOME index	0.034 [0.026]	-0.033 [0.025]	0.067+ [0.036]	0.082 [0.030]	-0.082 [0.031]	0.163** [0.043]	-0.096+ [0.056]
Read book to child 3+ times per week	0.883 [0.008]	0.868 [0.008]	0.015 [0.012]	0.837 [0.011]	0.794 [0.013]	0.043* [0.017]	-0.028 [0.021]
Child has ≥ 20 books around house	0.874 [0.009]	0.866 [0.008]	0.008 [0.012]	0.887 [0.009]	0.857 [0.011]	0.030* [0.015]	-0.023 [0.019]
Visited the library	0.618 [0.012]	0.571 [0.012]	0.047** [0.017]	0.572 [0.015]	0.550 [0.015]	0.021 [0.021]	0.026 [0.027]
Gone to a play/concert/show	0.426 [0.012]	0.410 [0.012]	0.016 [0.017]	0.422 [0.015]	0.346 [0.014]	0.076** [0.021]	-0.060* [0.027]
Visited art/museum/historical site	0.338 [0.012]	0.344 [0.011]	-0.006 [0.016]	0.308 [0.014]	0.297 [0.014]	0.011 [0.020]	-0.017 [0.025]
Child reads outside school 3+ times per week	0.249 [0.021]	0.009 [0.024]	0.241** [0.032]	0.139 [0.026]	-0.309 [0.034]	0.448** [0.042]	-0.207** [0.053]
Have home computer child uses	0.761 [0.011]	0.764 [0.011]	-0.003 [0.015]	0.602 [0.015]	0.563 [0.015]	0.038+ [0.021]	-0.041 [0.026]
Child engages in other outside school activity	0.753 [0.011]	0.740 [0.011]	0.012 [0.016]	0.683 [0.014]	0.629 [0.015]	0.054* [0.021]	-0.041 [0.026]
Kindergarten Warmth index	-0.002 [0.025]	0.002 [0.024]	-0.004 [0.035]	0.058 [0.027]	-0.058 [0.035]	0.116** [0.044]	-0.120* [0.056]
Warm, close times together	0.962 [0.005]	0.950 [0.005]	0.012+ [0.007]	0.956 [0.007]	0.947 [0.008]	0.009 [0.011]	0.002 [0.013]
Child likes me	0.981 [0.003]	0.970 [0.004]	0.011* [0.005]	0.979 [0.004]	0.968 [0.007]	0.012 [0.008]	-0.001 [0.009]
Always show child love	0.931 [0.006]	0.934 [0.006]	-0.004 [0.008]	0.870 [0.010]	0.862 [0.010]	0.008 [0.015]	-0.011 [0.017]
Express affection	0.992 [0.002]	0.985 [0.003]	0.007+ [0.004]	0.986 [0.002]	0.976 [0.004]	0.010* [0.005]	-0.004 [0.006]
Being parent harder than I thought (reverse)	0.413 [0.012]	0.420 [0.012]	-0.007 [0.017]	0.522 [0.015]	0.513 [0.015]	0.009 [0.021]	-0.015 [0.027]
Child does things that bother me (reverse)	0.916 [0.006]	0.902 [0.007]	0.014 [0.010]	0.912 [0.008]	0.882 [0.011]	0.030* [0.014]	-0.016 [0.017]
Sacrifice to meet child's needs (reverse)	0.715 [0.012]	0.751 [0.010]	-0.036* [0.016]	0.767 [0.013]	0.754 [0.013]	0.013 [0.018]	-0.050* [0.024]
Often feel angry with child (reverse)	0.988 [0.002]	0.985 [0.003]	0.003 [0.004]	0.984 [0.004]	0.985 [0.004]	-0.002 [0.005]	0.004 [0.007]
Spanked child last week, kindergarten	0.149 [0.009]	0.169 [0.009]	-0.020 [0.013]	0.260 [0.013]	0.276 [0.014]	-0.015 [0.019]	-0.004 [0.023]

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets

Notes: Columns 1-2 and 4-5 show the means of each row measure for each gender in the 2010 and 1998 cohort, respectively. Columns 3 and 6 show the difference between coefficients in columns 1-2 and 4-5, respectively. Column 7 shows the difference between columns 3 and 6. Significance stars are only included in columns 3, 6, and 7. Estimates are calculated by regressing each row measure on a female dummy, a 2010 cohort dummy, and a female by 2010 cohort dummy. The final rows of each section shows the p-value from a joint F-test of the null that the coefficients from a regression on a female dummy in each cohort on all listed measures are jointly zero. Sample is restricted as reported in the text. Observations are weighted using fifth grade parent panel weights for the 1998 cohort and fifth grade panel weights for the 2010 cohort. Standard errors are heteroskedasticity robust and clustered at the primary sampling unit level.

Table 2.C.3 Fifth Grade Joint Returns, by Gender and Cohort

Latent Noncognitive Skill in Fifth Grade	2010 Cohort			1998 Cohort			Diff-in-Diff
	Girls	Boys	Difference	Girls	Boys	Difference	
Lower kindergarten HOME index	0.039* [0.015]	0.016 [0.011]	0.023 [0.021]	0.038+ [0.023]	0.023* [0.011]	0.014 [0.029]	0.009 [0.036]
Lower kindergarten Warmth index	-0.040+ [0.022]	-0.048 [0.034]	0.008 [0.014]	-0.022* [0.011]	-0.166** [0.030]	0.145** [0.029]	-0.137** [0.032]
Spanked child last week, kindergarten	-0.178** [0.035]	-0.192** [0.011]	0.014 [0.039]	-0.083 [0.056]	-0.154** [0.046]	0.072 [0.051]	-0.057 [0.064]
Single mom	-0.279** [0.103]	-0.343** [0.030]	0.064 [0.112]	-0.156** [0.035]	-0.286** [0.068]	0.130 [0.081]	-0.066 [0.136]
Other family structure	-0.314** [0.057]	-0.547** [0.047]	0.233** [0.039]	-0.320** [0.062]	-0.428** [0.083]	0.108 [0.083]	0.125 [0.092]
Age first birth < 20	0.002 [0.039]	-0.110* [0.044]	0.112 [0.080]	-0.214* [0.095]	-0.353** [0.048]	0.138 [0.123]	-0.027 [0.147]
1st SES quintile (lowest)	-0.397** [0.065]	-0.402** [0.063]	0.005 [0.041]	-0.262* [0.127]	-0.400** [0.046]	0.138 [0.122]	-0.134 [0.127]
2nd SES quintile	-0.330** [0.055]	-0.333** [0.068]	0.003 [0.064]	-0.325** [0.094]	-0.269** [0.047]	-0.056 [0.096]	0.059 [0.113]
3rd SES quintile	-0.261** [0.086]	-0.255** [0.059]	-0.005 [0.059]	-0.090+ [0.048]	-0.278** [0.031]	0.188** [0.051]	-0.193* [0.079]
4th SES quintile	-0.099** [0.020]	-0.147* [0.056]	0.047 [0.043]	-0.129+ [0.066]	-0.074 [0.046]	-0.054 [0.092]	0.102 [0.102]
SES F-test of jointly zero, <i>p</i> -value			0.345			0.000	0.000

** p<0.01, * p<0.05, + p<0.1. Standard errors in brackets

Notes: Estimates are produced from one regression of latent noncognitive skill in fifth grade on both sets of three parental inputs and family background measures at kindergarten interacted fully with a set of dummy variables for female and 2010 cohort. Kindergarten HOME and Warmth indices used in this regression are multiplied by negative one to match the direction of the other measures in the table. The first two rows, the only continuous measures, report differing slopes between the subgroups. The remaining columns and rows report estimates as follows. 1998 boys: coefficient on row variable. 1998 girls: sum of coefficients on row variable and row variable by female interaction term. 1998 difference: coefficient on row variable by female interaction term. 2010 boys: sum of coefficients on row variable and row variable by 2010 cohort interaction term. 2010 girls: sum of coefficients on row variable, row variable by 2010 cohort interaction term, row variable by female interaction term, and row variable by female by 2010 cohort interaction term. 2010 difference: sum of coefficients on row variable by female interaction term, and row variable by female by 2010 cohort interaction term. Diff-in-diff: coefficient on row variable by female by 2010 cohort interaction term. Controls for child race and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.4 Changes in Gender Gaps Between Cohorts, The Role of Family Structure at Kindergarten

Category	Fall-K	Grade 3	Grade 5
Single mom	0.168 [0.115]	0.147 [0.127]	-0.065 [0.128]
Both biological parents	0.015 [0.043]	-0.024 [0.027]	0.040 [0.045]
Other family structure	-0.147 [0.118]	0.026 [0.138]	0.122 [0.087]
Joint F-test of no change <i>p</i> -value	0.265	0.366	0.441

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by family structure category interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for single motherhood and other family structure at kindergarten interacted with female and cohort dummies. Controls for socioeconomic status at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.5 Changes in Gender Gaps Between Cohorts, The Role of Mother's Age at First Birth

Category	Fall-K	Grade 3	Grade 5
Less than 20 years old	-0.017 [0.087]	-0.003 [0.093]	-0.016 [0.106]
More than 20 years old	0.039 [0.052]	0.019 [0.045]	0.040 [0.042]
Joint F-test of no change <i>p</i> -value	0.649	0.842	0.636

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by teen motherhood interaction term from a regression of latent noncognitive skill in each respective grade on an indicator for teen motherhood with female and cohort dummies. Controls for socioeconomic status at kindergarten, family structure at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

C.4 Results for Two Orthogonal Latent Factors

C.4.1 Introduction

As described in Section 2.4.2 on my factor analysis, Information Criterion support the use of two latent factors, rather than one. Because of this, I have included all results rerun with this two factor orientation, rather than the one factor "latent noncognitive skill" used in the main body of the paper. This two factor orientation is produced with a an orthogonal varimax rotation with a Kaiser correction. The factor analysis results using two factors and this rotation are shown below in Table B1, which is analogous to Table B.4 in the main analysis. For ease of understanding, I have named the first latent factor "social behavior" and the second latent factor "learning and socializing", (this naming will be explained further below), and they will be referred to as such for the remainder of Appendix 8.4.

The first panel of Table B1 and the uniqueness column of the second panel show us some of the overall differences obtained by using two factors, rather than one. As the first panel of Table B1 shows, whereas "latent noncognitive skill" captured all of the common variation, now the first latent factor, social behavior, captures about 60% and the second latent factor, learning and socializing, captures about 40%. Comparing the uniqueness column in the second panel of Table B1 to that of Table B.4, we can see that the use of two factors explains more of the variation in each individual factor, with lower uniqueness scores in each row. The drop is largest for both externalizing and internalizing behavior, suggesting that these two skills have more variation that is uncorrelated with each other but *is* correlated with the other three factors.

Next, the factor scores column in the second panel of Table B1 allow us to see the weights used in creating the weighted averages that become the two latent factors. This section shows us that social behavior is composed mainly of self control, with the remaining shares taken up evenly by externalizing behavior and interpersonal skills. It also shows us that the second latent factor, learning and socializing, is composed of about one third each interpersonal skills and approaches to learning, with the remainder taken up by internalizing problems and a little bit of externalizing behavior. Comparing to the factor scores in Table B.4, we can see that social behavior most

closely resembles the latent noncognitive skill measure used in the main analysis, with learning and socializing leaning more heavily on factors given less weight in the main analysis. Intuitively, we can understand the latent social behavior measure as capturing the degree to which students act out, act impulsively, and get along with others. Learning and socializing, on the other hand, has a much heavier emphasis on the more cognitively-related skills of approaches to learning and interpersonal skills, which are more reflective of a student's abilities to focus and participate in class in an effective and engaged manner. Looking ahead to Table B5 (analogous to Table C.1), which shows the R^2 s for each measure when regressing 8th grade outcomes on each latent factor, we can see support for this intuitive interpretation. Latent social behavior has greater explanatory power for 8th grade suspensions, whereas latent learning and socializing has greater explanatory power for 8th grade math and reading scores.

Table 2.C.6 Factor Loadings, Scores, and Uniqueness

	Eigenvalue	Proportion Explained			
Social Behavior	1.857	0.627			
Learning and Socializing	1.313	0.444			
Noncog Variables	Factor Loadings		Factor Scores		Uniqueness
	Social Behavior	Learning and Socializing	Social Behavior	Learning and Socializing	
Externalizing behavior	0.652	0.459	0.167	0.062	0.3642
Self control	0.765	0.493	0.576	-0.110	0.171
Interpersonal skills	0.671	0.567	0.132	0.285	0.228
Approaches to learning	0.565	0.596	-0.033	0.346	0.326
Internalizing problems	0.277	0.427	-0.059	0.172	0.741

Notes: Results are shown from an unrotated principal factor analysis of all five standardized noncognitive skills across all grades and cohorts. Fifth grade parent panel weights are used for this calculation. Results for further factors are not displayed due to low eigenvalues.

C.4.2 Figures

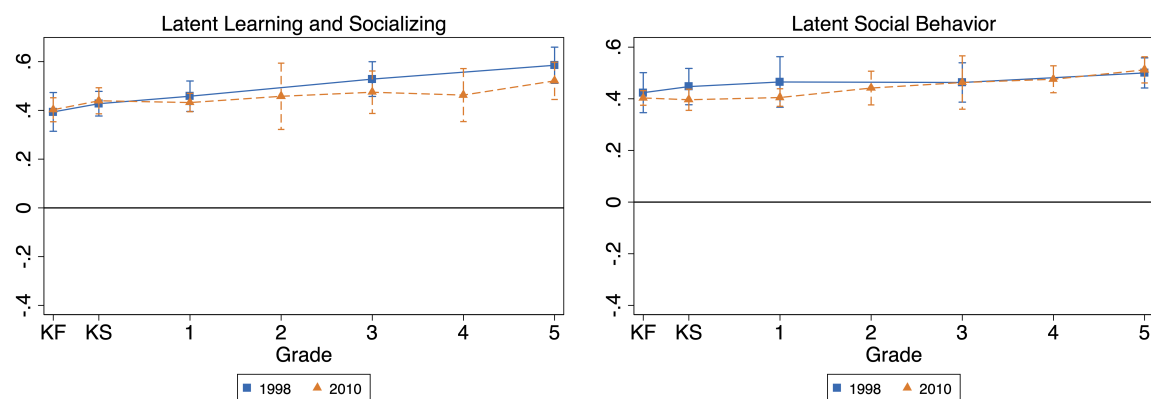


Figure 2.C.5 Female-Male Gaps in Teacher Ratings of Noncognitive Skills

Notes: Each graph shows the coefficient on a female dummy from a regression of each respective teacher-reported noncognitive skill in each respective grade on a female dummy variable. KF refers to the fall of kindergarten, KS refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

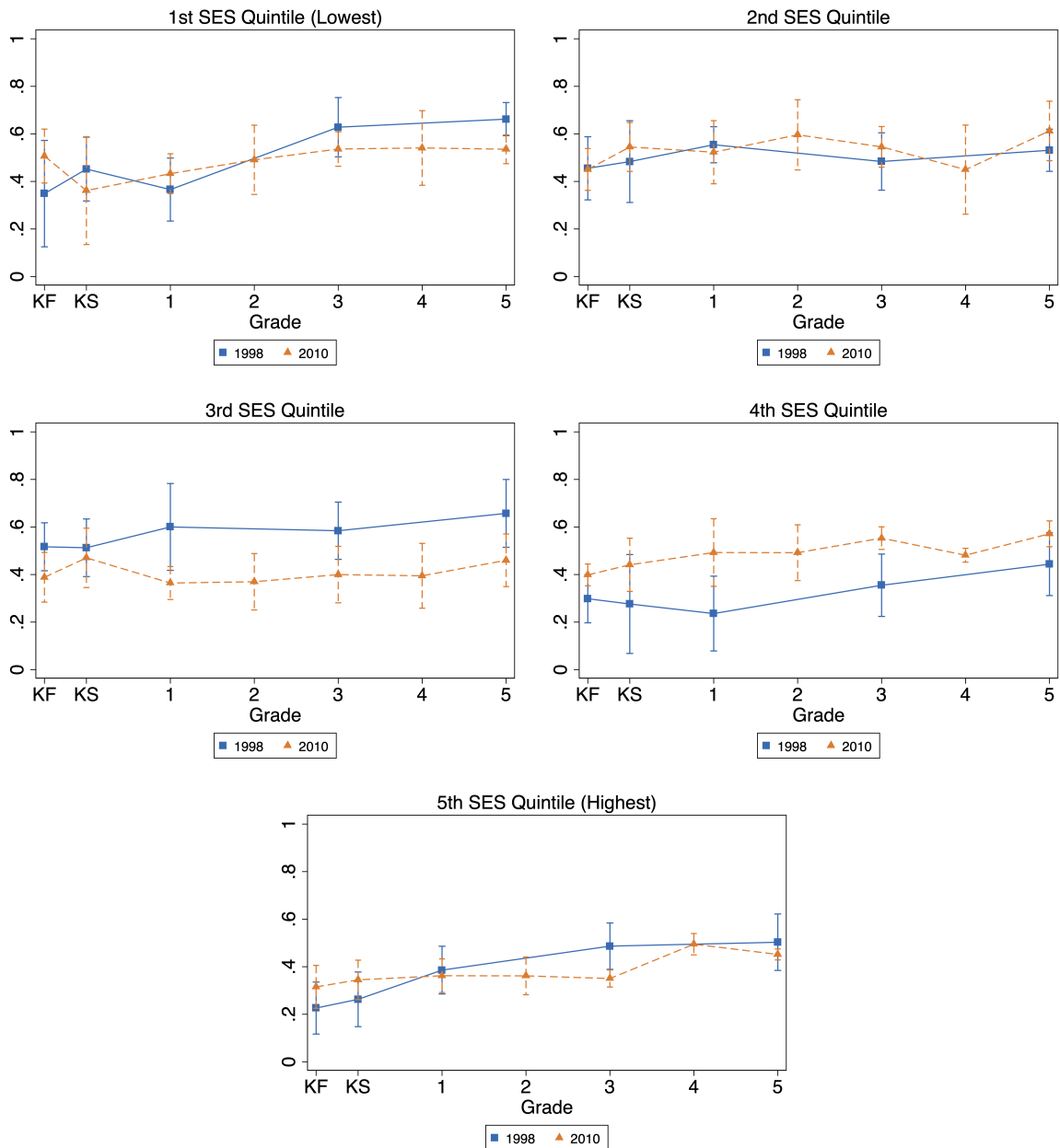


Figure 2.C.6 Female-Male Gaps in Latent Learning and Socializing, By SES at Kindergarten

Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by SES quintile interaction term, and a female by 2010 cohort by SES quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female by SES quintile interaction term (for the 1998 estimates) from a regression of latent learning and socializing in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

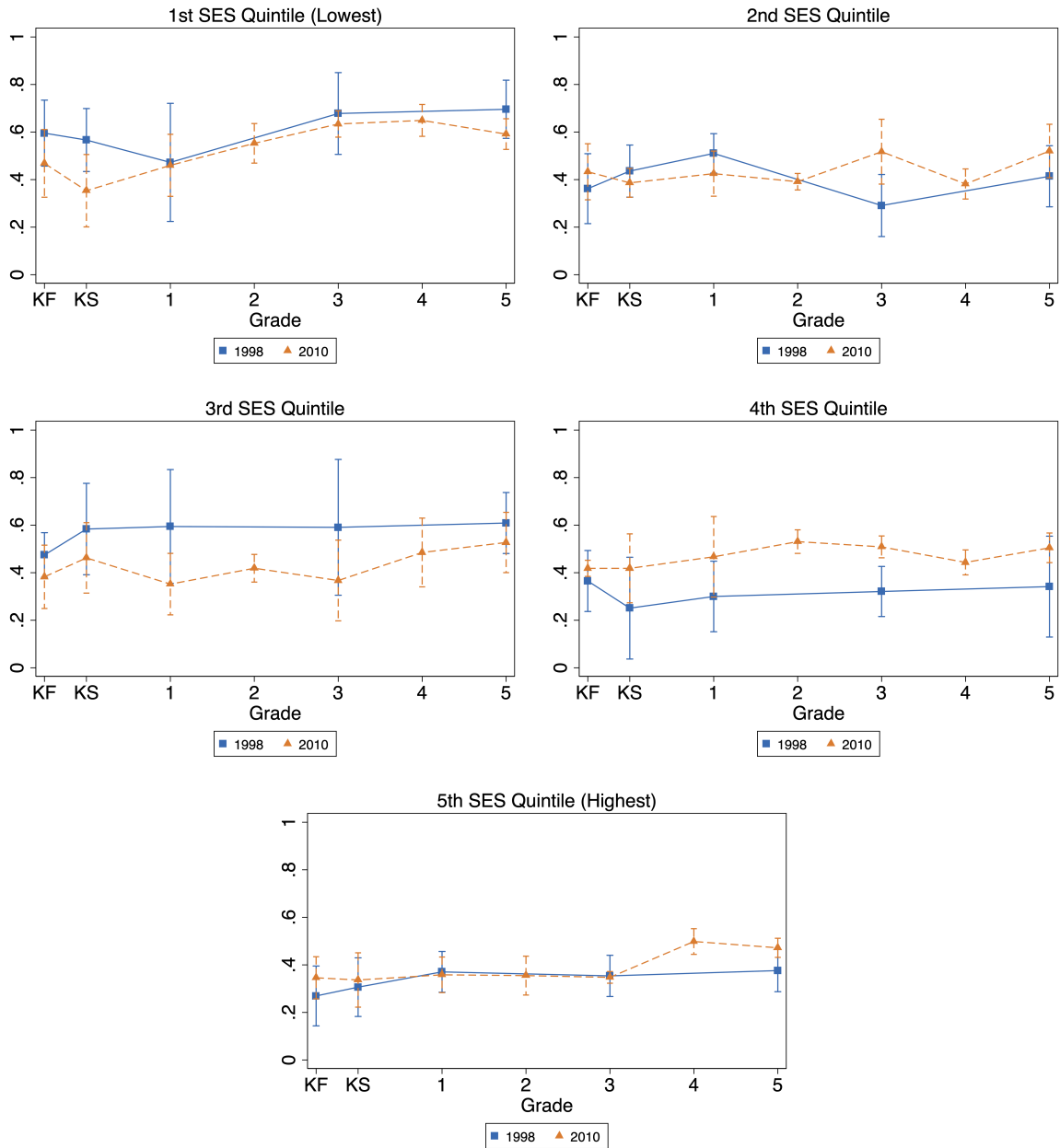


Figure 2.C.7 Female-Male Gaps in Latent Social Behavior, By SES at Kindergarten

Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by SES quintile interaction term, and a female by 2010 cohort by SES quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female by SES quintile interaction term (for the 1998 estimates) from a regression of Latent Social Behavior in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

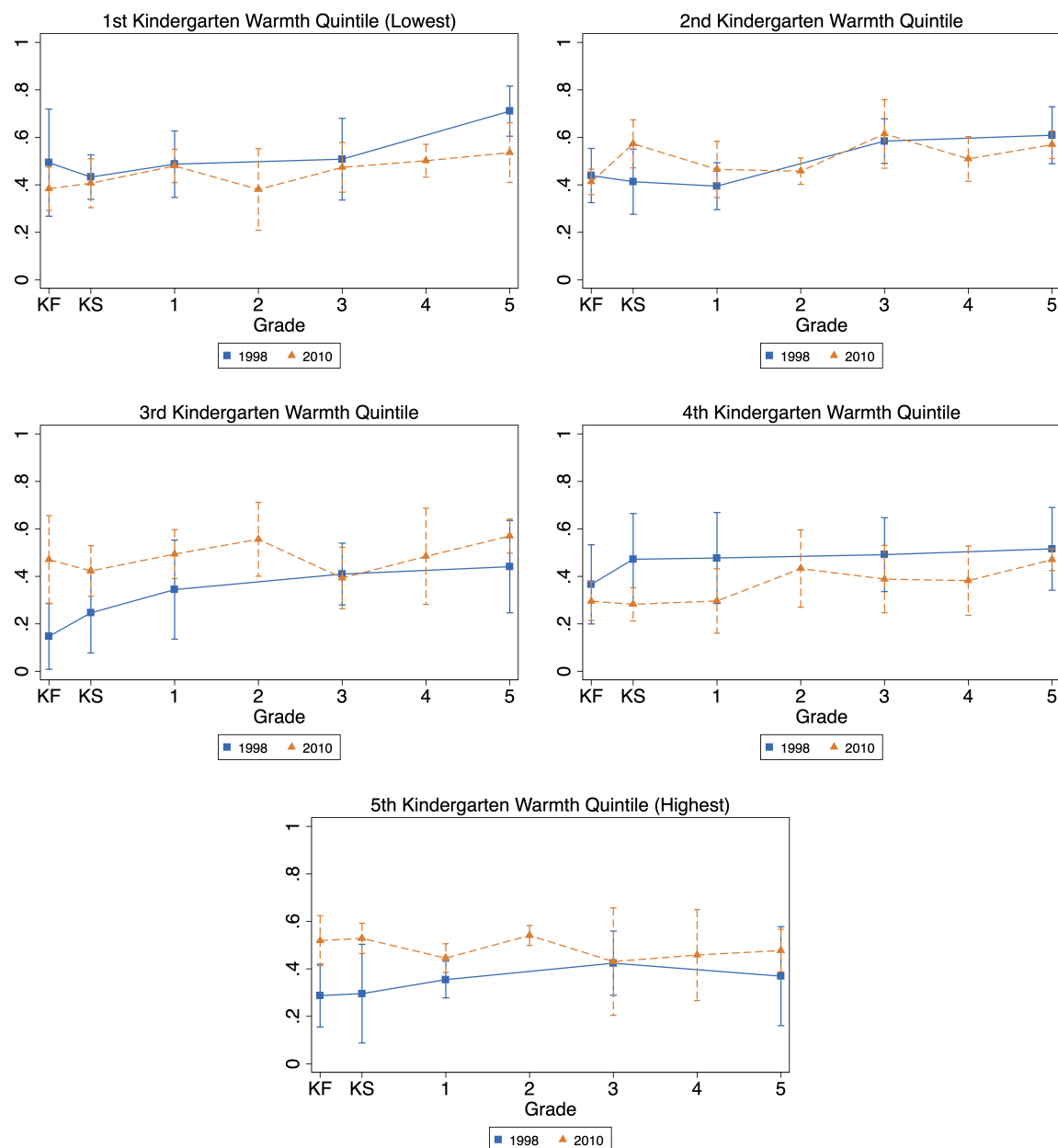


Figure 2.C.8 Female-Male Gaps in Latent Learning and Socializing, By Kindergarten Warmth Index

Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by Warmth index quintile interaction term, and a female by 2010 cohort by Warmth index quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female by Warmth index quintile interaction term (for the 1998 estimates) from a regression of latent learning and socializing in each respective grade on a set of indicators for four quintiles of Warmth index in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, socioeconomic status at kindergarten, HOME index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

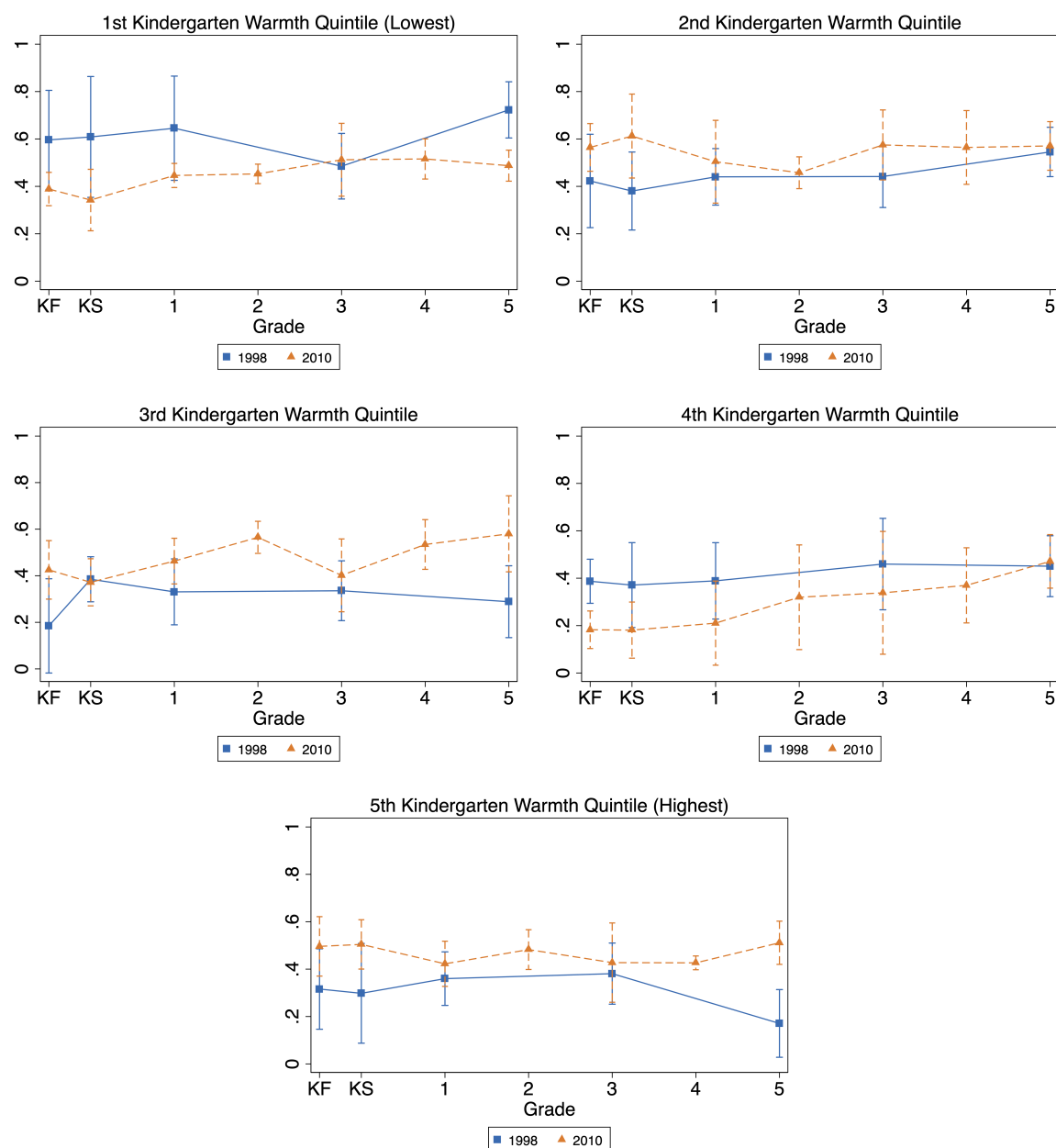


Figure 2.C.9 Female-Male Gaps in Latent Social Behavior, By Kindergarten Warmth Index

Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by Warmth index quintile interaction term, and a female by 2010 cohort by Warmth index quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female by Warmth index quintile interaction term (for the 1998 estimates) from a regression of Latent Social Behavior in each respective grade on a set of indicators for four quintiles of Warmth index in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, socioeconomic status at kindergarten, HOME index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

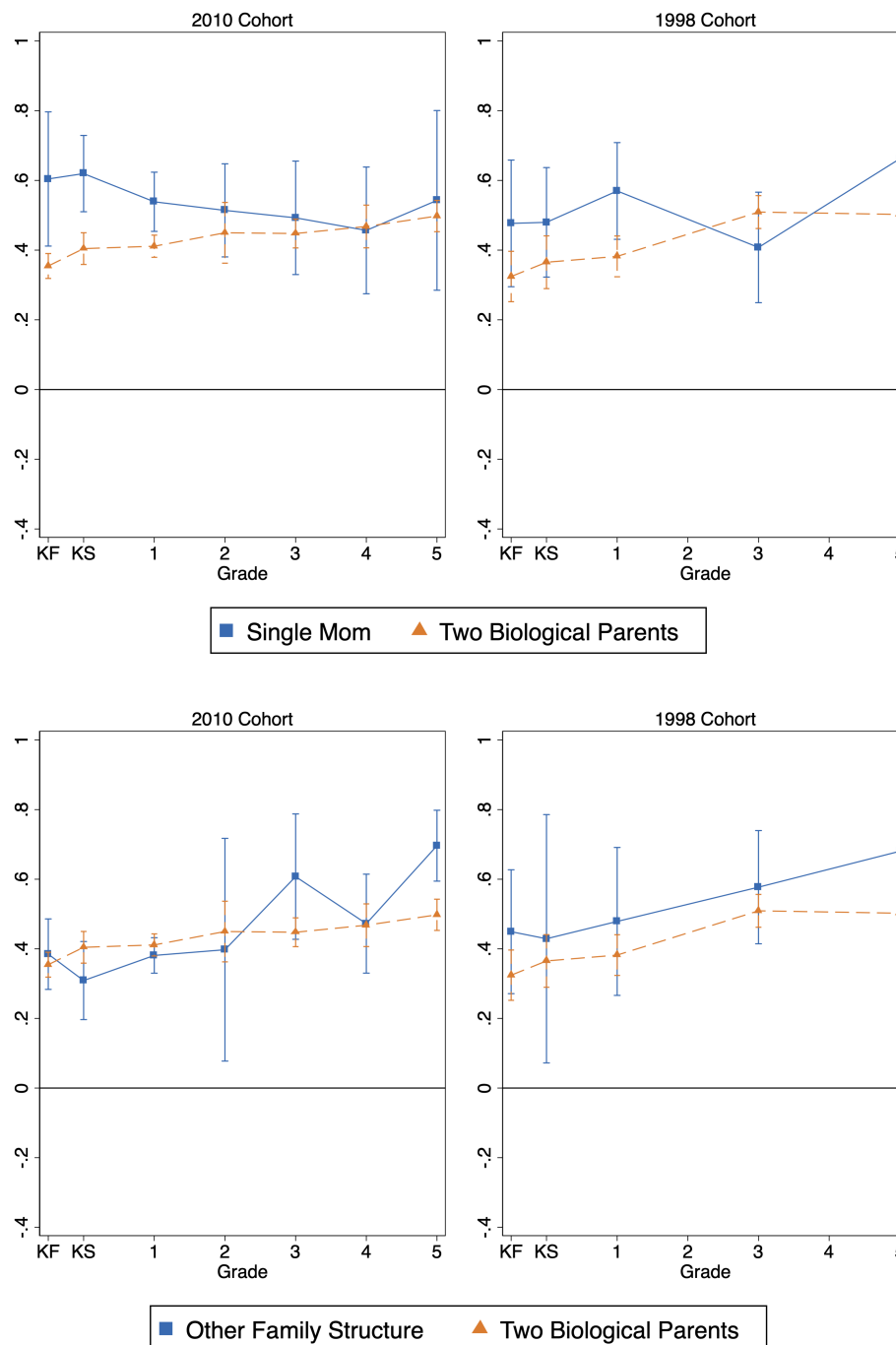


Figure 2.C.10 Female-Male Gaps in Latent Learning and Socializing, By Kindergarten Family Structure

Notes: Estimates for each grade in all four graphs come from one regression of latent learning and socializing in each grade with controls for teen motherhood, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. Separate estimates of gender gaps for each cohort and subgroup are produced using interaction terms for gender and cohort. Two biological parent estimates are the same in both rows. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions and reference bias corrections are imposed. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

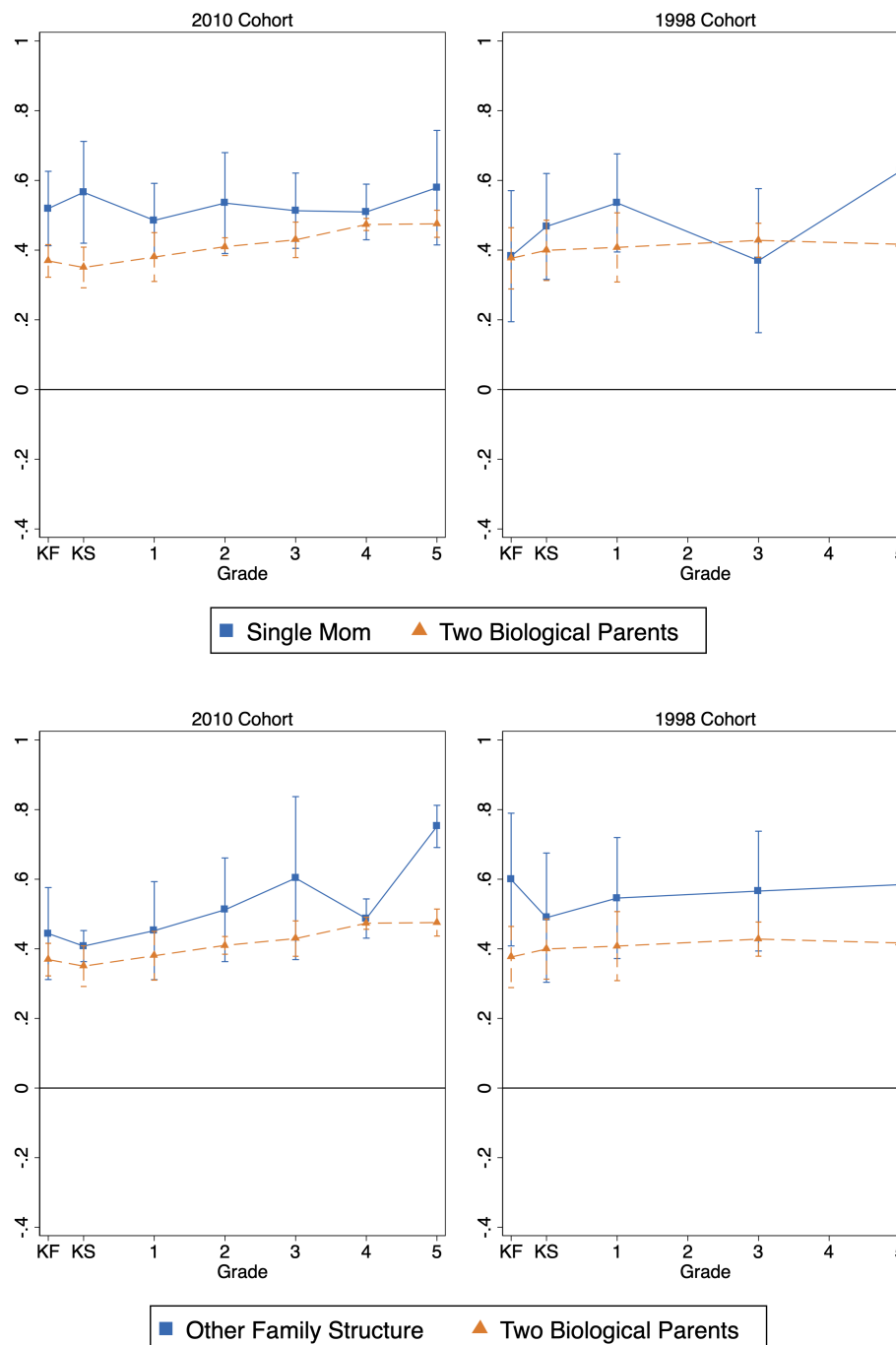


Figure 2.C.11 Female-Male Gaps in Latent Social Behavior, By Kindergarten Family Structure

Notes: Estimates for each grade in all four graphs come from one regression of Latent Social Behavior in each grade with controls for teen motherhood, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. Separate estimates of gender gaps for each cohort and subgroup are produced using interaction terms for gender and cohort. Two biological parent estimates are the same in both rows. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions and reference bias corrections are imposed. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

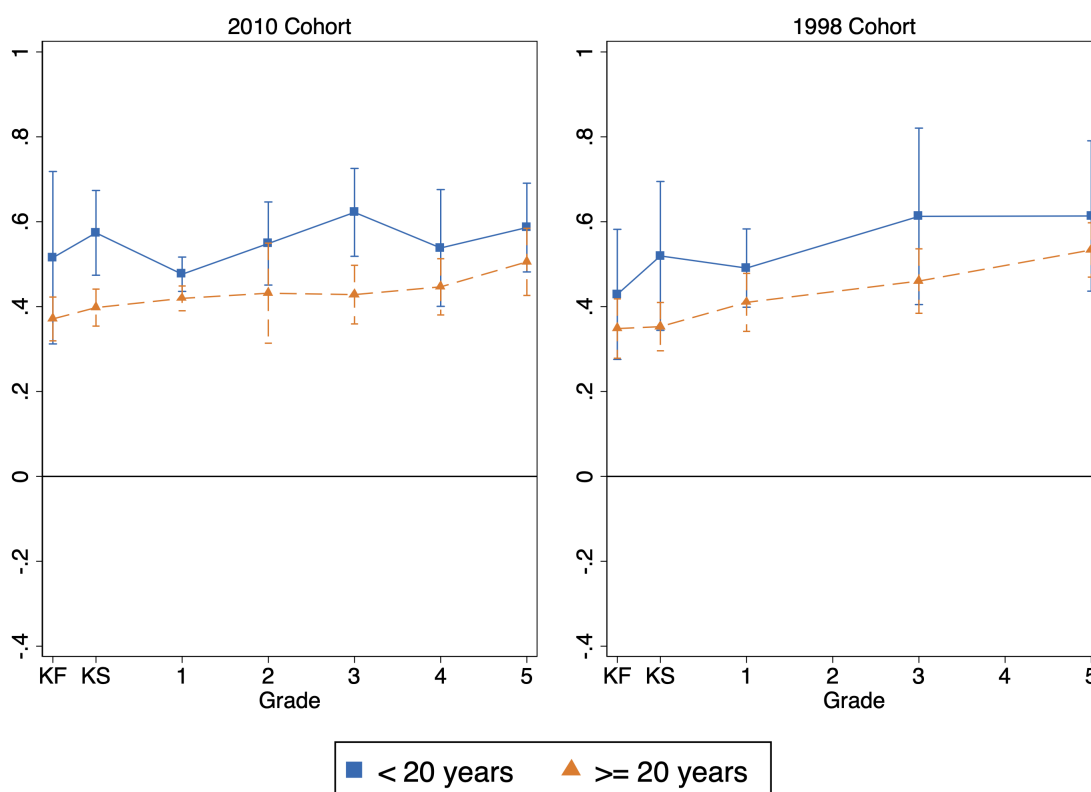


Figure 2.C.12 Female-Male Gaps in Latent Learning and Socializing, By Mother's Age at First Birth

Notes: Estimates for each grade in both graphs come from one regression of Latent Learning and Socializing in each grade with controls for family structure at kindergarten, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. Separate estimates of gender gaps for each cohort and subgroup are produced using interaction terms for gender and cohort. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions and reference bias corrections are imposed. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

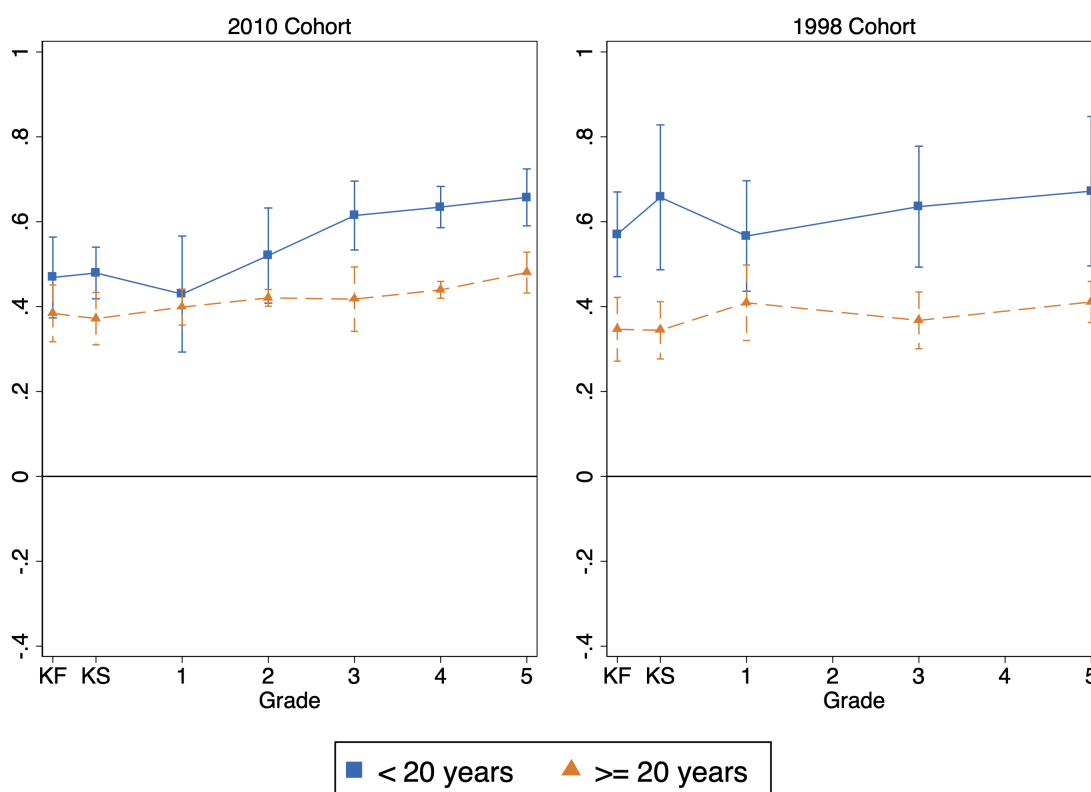


Figure 2.C.13 Female-Male Gaps in Latent Social Behavior, By Mother's Age at First Birth

Notes: Estimates for each grade in both graphs come from one regression of Latent Social Behavior in each grade with controls for family structure at kindergarten, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. Separate estimates of gender gaps for each cohort and subgroup are produced using interaction terms for gender and cohort. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions and reference bias corrections are imposed. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

C.4.3 Tables

Table 2.C.7 Changes in Female-Male Gaps in Teacher Ratings of Noncognitive Skills

Variable	Fall-K	Spring-K	Grade 1	Grade 3	Grade 5	Joint test <i>p</i> -value
Panel A: Unadjusted						
Latent Social Behavior	-0.021 [0.041]	-0.051 [0.040]	-0.060 [0.051]	-0.000 [0.061]	0.012 [0.037]	0.566
Latent Learning and Socializing	0.009 [0.046]	0.012 [0.035]	-0.026 [0.036]	-0.054 [0.054]	-0.064 [0.051]	0.717
Panel B: Adjusted						
Latent Social Behavior	-0.001 [0.038]	-0.028 [0.042]	-0.041 [0.047]	0.029 [0.049]	0.041 [0.033]	0.545
Latent Learning and Socializing	0.033 [0.043]	0.041 [0.035]	0.002 [0.036]	-0.027 [0.040]	-0.031 [0.047]	0.350

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets

Notes: Each cell shows the coefficient on an interaction term for female and 2010 data with the row measure in each column grade as the left hand side variable. The last column displays the *p*-value from a joint F-test of the null that the differences across all grades for each measure are zero. Teacher ratings and test scores are standardized to have a mean of zero and standard deviation one in the population based on weighting and sampling methodology correction after imposing the sample restrictions, with additional correction for reference bias. Regressions in panel B include controls for race, school locale, family background, and parental inputs as reported at kindergarten. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.8 Oaxaca-Blinder Decomposition of Fifth Grade Gender Gaps: Latent Learning and Socializing

Cohort	Predicted Gender Gap (girls – boys)	Unexplained	Due to Levels
Panel A: Boys' X's, Girls' Betas			
2010	0.521** [0.034]	0.523** [0.034]	-0.002 [0.011]
1998	0.585** [0.038]	0.560** [0.036]	0.025+ [0.014]
Difference	-0.064 [0.051]	-0.038 [0.049]	-0.026 [0.017]
Panel B: Girls' X's, Boys' Betas			
2010	0.521** [0.036]	0.526** [0.033]	-0.004 [0.012]
1998	0.585** [0.041]	0.538** [0.039]	0.048** [0.017]
Difference	-0.064 [0.055]	-0.012 [0.051]	-0.052* [0.021]

** p<0.01, * p<0.05, + p<0.1. Bootstrapped standard errors in brackets

Notes: The Oaxaca-Blinder decompositions shown here are performed as described in text. Gender gaps as reported in the first column are the predicted gender gap from a regression of each measure on family background, parental input, racial demographics, and school locale measures as reported at kindergarten interacted separately by cohort and gender. Standard errors are bootstrapped with 100 replications, with each row's estimates produced jointly in each bootstrapping iteration. Sample restrictions are imposed as described in text. Fifth grade parent panel weights are used for these estimates.

Table 2.C.9 Oaxaca-Blinder Decomposition of Fifth Grade Gender Gaps: Latent Social Behavior

Cohort	Predicted Gender Gap (girls – boys)	Unexplained	Due to Levels
Panel A: Boys' X's, Girls' Betas			
2010	0.512** [0.033]	0.518** [0.033]	-0.006 [0.010]
1998	0.500** [0.046]	0.489** [0.042]	0.011 [0.014]
Difference	0.012 [0.057]	0.029 [0.054]	-0.017 [0.017]
Panel B: Girls' X's, Boys' Betas			
2010	0.512** [0.030]	0.519** [0.029]	-0.007 [0.015]
1998	0.500** [0.044]	0.456** [0.041]	0.044+ [0.023]
Difference	0.012 [0.053]	0.063 [0.050]	-0.051+ [0.027]

** p<0.01, * p<0.05, + p<0.1. Bootstrapped standard errors in brackets

Notes: The Oaxaca-Blinder decompositions shown here are performed as described in text. Gender gaps as reported in the first column are the predicted gender gap from a regression of each measure on family background, parental input, racial demographics, and school locale measures as reported at kindergarten interacted separately by cohort and gender. Standard errors are bootstrapped with 100 replications, with each row's estimates produced jointly in each bootstrapping iteration. Sample restrictions are imposed as described in text. Fifth grade parent panel weights are used for these estimates.

Table 2.C.10 Changes in Gender Gaps Between Cohorts, The Role of Socioeconomic Status at Kindergarten

Latent Learning and Socializing In:	Fall-K	Grade 3	Grade 5
1st SES quintile (lowest)	0.176 [0.115]	-0.090 [0.071]	-0.142** [0.049]
2nd SES quintile	0.002 [0.081]	0.056 [0.064]	0.095 [0.082]
3rd SES quintile	-0.129 [0.079]	-0.189* [0.082]	-0.205* [0.092]
4th SES quintile	0.111* [0.053]	0.206** [0.069]	0.135+ [0.072]
5th SES quintile (highest)	0.102 [0.072]	-0.124* [0.056]	-0.034 [0.057]
Joint F-test of equality <i>p</i> -value	0.001	0.000	0.000

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by SES quintile interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.11 Changes in Gender Gaps Between Cohorts, The Role of Socioeconomic Status at Kindergarten

Latent Social Behavior In:	Fall-K	Grade 3	Grade 5
1st SES quintile (lowest)	-0.112 [0.111]	-0.033 [0.090]	-0.114 [0.071]
2nd SES quintile	0.079 [0.097]	0.222* [0.092]	0.129 [0.088]
3rd SES quintile	-0.090 [0.083]	-0.233 [0.164]	-0.090 [0.093]
4th SES quintile	0.063 [0.069]	0.184** [0.057]	0.171 [0.117]
5th SES quintile (highest)	0.090 [0.081]	0.005 [0.042]	0.113* [0.055]
Joint F-test of equality <i>p</i> -value	0.268	0.000	0.023

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by SES quintile interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.12 Changes in Gender Gaps Between Cohorts, The Role of Kindergarten Warmth Index

Latent Learning and Socializing In:	Fall-K	Grade 3	Grade 5
1st Kindergarten Warmth quintile (lowest)	-0.110 [0.126]	-0.035 [0.102]	-0.175* [0.080]
2nd Kindergarten Warmth quintile	-0.027 [0.064]	0.030 [0.086]	-0.040 [0.068]
3rd Kindergarten Warmth quintile	0.323** [0.120]	-0.017 [0.096]	0.129 [0.105]
4th Kindergarten Warmth quintile	-0.071 [0.094]	-0.103 [0.108]	-0.046 [0.092]
5th Kindergarten Warmth quintile (highest)	0.232* [0.089]	0.007 [0.131]	0.108 [0.113]
Joint F-test of no change <i>p</i> -value	0.084	0.850	0.024

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by kindergarten Warmth index quintile interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, SES at kindergarten, HOME index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.13 Changes in Gender Gaps Between Cohorts, The Role of Kindergarten Warmth Index

Latent Social Behavior In:	Fall-K	Grade 3	Grade 5
1st Kindergarten Warmth quintile (lowest)	-0.207+ [0.113]	0.027 [0.106]	-0.235** [0.070]
2nd Kindergarten Warmth quintile	0.142 [0.114]	0.133 [0.101]	0.025 [0.074]
3rd Kindergarten Warmth quintile	0.240+ [0.126]	0.066 [0.106]	0.292* [0.115]
4th Kindergarten Warmth quintile	-0.205** [0.061]	-0.121 [0.165]	0.020 [0.088]
5th Kindergarten Warmth quintile (highest)	0.180+ [0.106]	0.046 [0.104]	0.340** [0.090]
Joint F-test of equality <i>p</i> -value	0.038	0.651	0.001

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by kindergarten Warmth index quintile interaction term from a regression of latent noncognitive skill in each respective grade on a set of indicators for four quintiles of socioeconomic status in kindergarten interacted with female and cohort dummies. Controls for family structure at kindergarten, teen motherhood, SES at kindergarten, HOME index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.14 Adjusted R^2 s from Regressions of Outcomes on Latent Noncognitive Skill or Components

5th Grade Measure(s)	Latent Learning and Socializing	Latent Social Behavior	Latent Noncognitive Skill	All Noncognitive Measures
Had any out-of-school suspensions by 8th Grade	0.132	0.171	0.176	0.201
Was held back by 8th Grade	0.054	0.028	0.042	0.064
Math scores in 8th Grade	0.120	0.053	0.086	0.142
Reading scores in 8th Grade	0.143	0.076	0.113	0.162

Notes: Each cell shows the adjusted R^2 from a regression of each row outcome on the noncognitive measure or measures listed in each column. The right-hand-side measure in column 1 is fifth grade latent noncognitive skill. The right-hand-side measures in column 2 are fifth grade externalizing behavior, self control, interpersonal skills, approaches to learning, and internalizing problems. Results only include observations from the 1998 cohort as there is no eighth grade data for the 2010 cohort, as well as no information on suspensions or disciplinary incidents. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.15 Fifth Grade Learning and Socializing Joint Returns, by Gender and Cohort

	2010 Cohort			1998 Cohort			Diff-in-Diff
	Girls	Boys	Difference	Girls	Boys	Difference	
Lower kindergarten HOME index	0.013 [0.014]	0.005 [0.016]	0.008 [0.026]	0.008 [0.015]	-0.029 [0.021]	0.036 [0.022]	-0.028 [0.035]
Lower kindergarten Warmth index	-0.025 [0.022]	-0.043 [0.031]	0.018 [0.017]	-0.013 [0.018]	-0.119** [0.022]	0.106** [0.033]	-0.087* [0.037]
Spanked child last week, kindergarten	-0.213** [0.036]	-0.185** [0.013]	-0.028 [0.030]	-0.107 [0.066]	-0.112** [0.031]	0.005 [0.059]	-0.033 [0.066]
Single mom	-0.281** [0.095]	-0.319** [0.043]	0.038 [0.128]	-0.178** [0.039]	-0.291** [0.077]	0.113 [0.092]	-0.075 [0.156]
Other family structure	-0.335** [0.056]	-0.512** [0.034]	0.178** [0.060]	-0.235** [0.062]	-0.377** [0.056]	0.142 [0.087]	0.036 [0.104]
Age first birth < 20	-0.012 [0.050]	-0.055 [0.058]	0.043 [0.106]	-0.195** [0.068]	-0.215** [0.041]	0.020 [0.095]	0.022 [0.144]
1st SES quintile (lowest)	-0.442** [0.037]	-0.478** [0.054]	0.036 [0.064]	-0.354** [0.094]	-0.417** [0.064]	0.063 [0.072]	-0.027 [0.097]
2nd SES quintile	-0.345** [0.071]	-0.463** [0.056]	0.118 [0.090]	-0.332** [0.084]	-0.295** [0.046]	-0.037 [0.075]	0.154 [0.117]
3rd SES quintile	-0.293** [0.035]	-0.282** [0.053]	-0.011 [0.058]	-0.150** [0.043]	-0.287** [0.034]	0.137** [0.036]	-0.147* [0.070]
4th SES quintile	-0.093** [0.011]	-0.205** [0.040]	0.111* [0.046]	-0.139* [0.067]	-0.075 [0.047]	-0.064 [0.090]	0.175+ [0.101]

** p<0.01, * p<0.05, + p<0.1. Standard errors in brackets

Notes: Estimates are produced from one regression of latent learning and socializing in fifth grade on both sets of three parental inputs and family background measures at kindergarten interacted fully with a set of dummy variables for female and 2010 cohort. Kindergarten HOME and Warmth indices used in this regression are multiplied by negative one to match the direction of the other measures in the table. The first two rows, the only continuous measures, report differing slopes between the subgroups. The remaining columns and rows report estimates as follows. 1998 boys: coefficient on row variable. 1998 girls: sum of coefficients on row variable and row variable by female interaction term. 1998 difference: coefficient on row variable by female interaction term. 2010 boys: sum of coefficients on row variable and row variable by 2010 cohort interaction term. 2010 girls: sum of coefficients on row variable, row variable by 2010 cohort interaction term, row variable by female interaction term, and row variable by female by 2010 cohort interaction term. 2010 difference: sum of coefficients on row variable by female interaction term, and row variable by female by 2010 cohort interaction term. Diff-in-diff: coefficient on row variable by female by 2010 cohort interaction term. Controls for child race and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.16 Fifth Grade Learning and Socializing Joint Returns, by Gender and Cohort

	2010 Cohort			1998 Cohort			Diff-in-Diff
	Girls	Boys	Difference	Girls	Boys	Difference	
Lower kindergarten HOME index	0.051** [0.019]	0.021+ [0.012]	0.030 [0.027]	0.052+ [0.028]	0.052* [0.022]	-0.001 [0.047]	0.031 [0.054]
Lower kindergarten Warmth index	-0.045* [0.021]	-0.046 [0.032]	0.001 [0.012]	-0.025 [0.016]	-0.178** [0.033]	0.153** [0.026]	-0.153** [0.028]
Spanked child last week, kindergarten	-0.140** [0.049]	-0.178** [0.014]	0.038 [0.057]	-0.060 [0.050]	-0.164** [0.056]	0.105+ [0.058]	-0.066 [0.080]
Single mom	-0.251* [0.100]	-0.324** [0.029]	0.073 [0.093]	-0.127** [0.039]	-0.254** [0.072]	0.128 [0.082]	-0.054 [0.124]
Other family structure	-0.271** [0.052]	-0.514** [0.057]	0.244** [0.031]	-0.339** [0.063]	-0.415** [0.101]	0.077 [0.098]	0.167 [0.104]
Age first birth < 20	0.011 [0.034]	-0.132** [0.032]	0.142* [0.059]	-0.204+ [0.105]	-0.399** [0.051]	0.196 [0.130]	-0.053 [0.143]
1st SES quintile (lowest)	-0.332** [0.085]	-0.317** [0.067]	-0.015 [0.042]	-0.180 [0.136]	-0.350** [0.052]	0.170 [0.154]	-0.184 [0.158]
2nd SES quintile	-0.288** [0.042]	-0.223** [0.075]	-0.066 [0.058]	-0.287** [0.094]	-0.225** [0.050]	-0.062 [0.105]	-0.004 [0.117]
3rd SES quintile	-0.216* [0.108]	-0.214** [0.065]	-0.002 [0.060]	-0.044 [0.052]	-0.244** [0.037]	0.200** [0.068]	-0.201* [0.091]
4th SES quintile	-0.093** [0.033]	-0.098 [0.065]	0.005 [0.044]	-0.109+ [0.063]	-0.066 [0.056]	-0.044 [0.095]	0.048 [0.104]

** p<0.01, * p<0.05, + p<0.1. Standard errors in brackets

Notes: Estimates are produced from one regression of latent learning and socializing in fifth grade on both sets of three parental inputs and family background measures at kindergarten interacted fully with a set of dummy variables for female and 2010 cohort. Kindergarten HOME and Warmth indices used in this regression are multiplied by negative one to match the direction of the other measures in the table. The first two rows, the only continuous measures, report differing slopes between the subgroups. The remaining columns and rows report estimates as follows. 1998 boys: coefficient on row variable. 1998 girls: sum of coefficients on row variable and row variable by female interaction term. 1998 difference: coefficient on row variable by female interaction term. 2010 boys: sum of coefficients on row variable and row variable by 2010 cohort interaction term. 2010 girls: sum of coefficients on row variable, row variable by 2010 cohort interaction term, row variable by female interaction term, and row variable by female by 2010 cohort interaction term. 2010 difference: sum of coefficients on row variable by female interaction term, and row variable by female by 2010 cohort interaction term. Diff-in-diff: coefficient on row variable by female by 2010 cohort interaction term. Controls for child race and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.17 Changes in Latent Learning and Socializing Gender Gaps Between Cohorts, The Role of Family Structure at Kindergarten

Category	Fall-K	Grade 3	Grade 5
Single mom	0.152 [0.135]	0.098 [0.119]	-0.104 [0.158]
Both biological parents	0.038 [0.043]	-0.059* [0.029]	-0.004 [0.044]
Other family structure	-0.076 [0.104]	0.030 [0.122]	0.022 [0.105]
Joint F-test of equality <i>p</i> -value	0.440	0.435	0.723
* $p < 0.01$, * $p < 0.05$, + $p < 0.1$ Standard errors in brackets			

Notes: Each estimate shows the coefficients on a female by 2010 cohort by family structure category interaction term from a regression of latent learning and socializing in each respective grade on a set of indicators for single motherhood and other family structure at kindergarten interacted with female and cohort dummies. Controls for socioeconomic status at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.18 Changes in Latent Social Behavior Gender Gaps Between Cohorts, The Role of Family Structure at Kindergarten

Category	Fall-K	Grade 3	Grade 5
Single mom	0.165 [0.111]	0.163 [0.127]	-0.033 [0.109]
Both biological parents	0.004 [0.060]	0.002 [0.030]	0.064 [0.048]
Other family structure	-0.168 [0.116]	0.025 [0.140]	0.172+ [0.089]
Joint F-test of equality <i>p</i> -value	0.205	0.384	0.323
** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Standard errors in brackets			

Notes: Each estimate shows the coefficients on a female by 2010 cohort by family structure category interaction term from a regression of latent social behavior in each respective grade on a set of indicators for single motherhood and other family structure at kindergarten interacted with female and cohort dummies. Controls for socioeconomic status at kindergarten, teen motherhood, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.19 Changes in Latent Learning and Socializing Gender Gaps Between Cohorts, The Role of Mother's Age at First Birth

Category	Fall-K	Grade 3	Grade 5
Less than 20 years old	0.104 [0.130]	0.014 [0.110]	-0.028 [0.111]
More than 20 years old	0.030 [0.047]	-0.029 [0.050]	-0.022 [0.051]
Joint F-test of equality <i>p</i> -value	0.598	0.766	0.961

* $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by teen motherhood interaction term from a regression of latent learning and socializing in each respective grade on an indicator for teen motherhood with female and cohort dummies. Controls for socioeconomic status at kindergarten, family structure at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

Table 2.C.20 Changes in Latent Social Behavior Gender Gaps Between Cohorts, The Role of Mother's Age at First Birth

Category	Fall-K	Grade 3	Grade 5
Less than 20 years old	-0.080 [0.076]	-0.009 [0.083]	-0.005 [0.101]
More than 20 years old	0.045 [0.056]	0.048 [0.045]	0.075* [0.036]
Joint F-test of equality <i>p</i> -value	0.240	0.538	0.468

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Standard errors in brackets

Notes: Each estimate shows the coefficients on a female by 2010 cohort by teen motherhood interaction term from a regression of latent social behavior in each respective grade on an indicator for teen motherhood with female and cohort dummies. Controls for socioeconomic status at kindergarten, family structure at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten are included. Teacher ratings are standardized to have a mean of zero and standard deviation one in the population based on weighting after imposing the sample restrictions. Please refer to the text for sample restrictions. Observations are weighted using fifth grade parent panel weights, with robust standard errors and clustering at the primary sampling unit level.

CHAPTER 3

RAISING BOYS, RAISING GIRLS: MODELING GENDER DIFFERENCES IN THE PROCESS OF EARLY CHILDHOOD SKILL FORMATION

3.1 Abstract

Using two nationally representative datasets with detailed information on students, parents, and schools, I create a model of early childhood human capital production and test for differences in parameters for boys versus girls. Building from a basic model of contemporaneous correlation of inputs up to a dynamic technology of skill formation model as proposed by Cunha and Heckman (2008), I test whether the model parameters that best fit the data are detectably different between boys and girls. Results are inconclusive, as correctly measuring and determining meaningful parental inputs in the investment process is tricky. I test the robustness of the Cunha and Heckman (2008) model to modeling assumptions and measurement of parental inputs, and find 1) the value-added model sufficiently captures the process of skill formation, relative to the cumulative model of Todd and Wolpin (2003), and 2) parental investment as captured by the measures available in the ECLS-K do not have a statistically detectable impact on the formation of noncognitive skills, regardless of the specification used.

3.2 Introduction

A significant amount of research has established the existence of educational gender gaps in both short-term and long-term outcomes. For short-term outcomes, researchers have noted boys' advantages in math test scores and girls' advantages in reading scores (Dee, 2007, e.g.), girls' advantages in grades in school (Cornwell et al., 2013; Fortin et al., 2015), and girls' advantages across several dimensions of noncognitive skills (Bertrand and Pan, 2013; Johann, 2020). For long term outcomes, girls have notably higher college entrance and persistence rates, which several studies have connected back to gender gaps in noncognitive skills (Jacob, 2002; Goldin et al., 2006; Becker et al., 2010). Like racial gaps observed in similar outcomes, the existence of these gender gaps raises important questions of interest for policymakers seeking to design interventions to equalize outcomes across groups. Primarily, these gaps raise the question: what may be causing boys and girls to have divergences in their short and long term outcomes?

There are two particular characteristics of the nature of gender gaps that differentiates research on them from that of other commonly researched gaps, such as race. First, gender is randomly assigned at conception; thus, boys and girls are no more or less likely to be born into disadvantaged families. Second, boys and girls have some biologically determined developmental differences. Research has shown that higher exposure in utero to sex hormones such as testosterone is associated with structural brain differences and lower levels of empathy and inhibitory control (Baron-Cohen, 2002, 2003; Knickmeyer et al., 2005). Developmental differences such as these open the possibility that some gender gaps in human capital development may be due to differing responses to similar inputs.¹ Putting these two characteristics together, we have that (1) background and family characteristics should be more evenly distributed between genders than between other groups, and (2) that there is a possible scientific basis for more immutable differences in development.

However, economists have documented both gender differences in inputs and differential

¹It is important to note here that whether any of these differing responses are due to unmeasured cultural practices in raising different gendered children or some fixed, immutable biological characteristics is very difficult to determine. Any findings in this paper about gender differences in development should *not* be interpreted as making claims about whether these gender differences are permanent, only that they exist alongside the cultural contexts in which the children were raised at the times of these surveys.

gender responses to similar levels of input. Several papers have looked at how parents in similar circumstances will report spending different amounts of time in various activities with boys versus girls (Baker and Milligan, 2016; Bibler, 2018), thus opening a pathway for post-birth differences to cause differing development. In addition, other research has examined the differing responses of boys versus girls to disadvantaged backgrounds such as impoverished neighborhoods or low family socioeconomic status and noted that these conditions of deprivation seem to have larger negative effects on boys and their human capital accumulation (Autor et al., 2019, 2020; Chetty et al., 2016; Kling et al., 2005). The existence of both strands of research findings suggests that whether gender gaps exist due to gender differences in inputs or different responses to the same level of inputs is an open question. That is, whether boys and girls truly have differing technologies of skill formation is an unanswered question (Cunha and Heckman, 2007).

To address this question, I propose using detailed information on parental inputs, school inputs, cognitive test scores, and noncognitive skill evaluations to model separate dynamic processes of skill formation for boys and girls. Specifically, I will be basing the formation of my model off Cunha and Heckman (2008), who propose a method of modeling and estimating the technology of skill formation using a dynamic process where next-period human capital formation is a function of current-period human capital, both cognitive and noncognitive, and current period parental investments.² Variations of this model have seen extensive use over the past few years in research on early childhood parenting and interventions, from understanding the mechanisms behind an early childhood education program in Colombia (Attanasio et al., 2020), to understanding how parents modify their investment behavior in response to expanded public preschool in Denmark (Gensowski et al., 2020), to estimating the tradeoffs of income and substitution effects on cognitive and behavioral development of children in the US in the context of maternal employment decisions (Agostinelli and Sorrenti, 2018). See Heckman and Mosso (2014) for a more thorough review of the pre-2014 state of this field. In short, this model has been established as a strong framework for evaluating dynamics in human capital investment and accumulation, as first proposed by Cunha

²See Section 3 for a more in-depth description of this model.

and Heckman (2007).³

Heckman et al. (2006), in related work also apply their model to estimating gender differences in the process of skill formation. For both males and females, they estimate separate models of schooling, employment, work experience, occupational choice, and wages based on latent cognitive and noncognitive skills using the NLSY. This model does not attempt to look at differing impacts of human capital investment, but instead sees latent abilities as factors influencing differing choices between the genders. This model finds that noncognitive skills play a slightly larger role for females in the wages that result from these choices for than for males. However, as this is a model of decision-making in adolescence through adulthood, rather than a dynamic process of skill formation beginning in early childhood, much about the gender differences in skill formation remains unmodeled. As Johann (2020) shows, large gender gaps in noncognitive skills are present even at the start of kindergarten, and only grow over the elementary school years. Employing the more recently developed and demonstrated dynamic skill formation model and examining earlier in childhood than Heckman et al. (2006) will likely provide better insight into how these skill gaps emerge.

My novel contribution to this field will be (1) the application of this dynamic model of skill formation to the field of educational gender gaps in elementary school and (2) the use of the ECLS-K datasets for the calibration, which include enough within-school variation to adequately control for time-invariant school quality contributions to the production function. The ECLS-K datasets repeatedly collect data on the same panel of students throughout elementary school students from kindergarten to fifth grade, collecting information on child test scores and noncognitive abilities, and parental educational and investment activities at each wave. This survey structure of repeatedly gathering these categories of information for the same individuals at different time periods is particularly well-suited to the dynamic nature of the technology of skill formation model described by Cunha and Heckman (2008), allowing me to see how each stage's abilities and investment decisions affect the next, and how the impacts of these decisions differ between boys and girls. In

³In Section 2, I propose a system of cross-model validation tests of whether this particular model does indeed fit the ECLS-K data.

summary, I believe the application of this model to the ECLS-K will allow me to estimate whether and how boys and girls may develop differently in response to measured inputs, helping answer the question of whether gender gaps are due to differences in inputs, or differences in technologies.

The remainder of this chapter of my dissertation proposal is as follows. Section 2 discusses how I validate that the model of Cunha and Heckman (2008) is best suited to the data, following Todd and Wolpin (2003). Section 3 describes the Cunha and Heckman (2008) model and its assumptions in greater detail. Section 4 describes the data and measures to which I apply this method. Section 5 shows the results of this model applied to the two ECLS-K datasets. Section 6 tests the robustness of the Cunha and Heckman (2008) model's findings to alternative assumptions. Section 7 concludes.

3.3 Cross-Model Validation of Human Capital Production Function

3.3.1 Basic Model

The most basic human capital production function sets ability in each period as the product of schooling, family inputs in the previous period, as well as permanent endowments. Let A_t be a measure of ability (e.g. math test scores) in period t , S_t be a measure of schooling inputs, F_t be a measure of family inputs, and μ be a vector of permanent endowments, such as genetic predisposition towards intelligence. Our production function is then:

$$A_t = g(S_{t-1}, F_{t-1}, \mu) \quad \forall t = 2, \dots, T$$

In this case, we can estimate a linear and additively separable version of this model with panel data on each period's ability is regressed on the previous period's schooling and family inputs along with some measures of endowments. Of course, we only have imperfect measures of these three inputs, but as long as our omitted measures are orthogonal to our included ones (a strong assumption), we can estimate this model with panel OLS.

3.3.2 Directly Including Ability

The next possible wrinkle to add is modeling the direct role of previous period ability in the production of current period ability. Cunha and Heckman (2007) argue that ability itself likely

modifies how future ability is produced. That is, a child with more ability may not only receive differing inputs, but that ability itself may also directly reinforce (if skill begets skill) the child's production of future skill. This step brings the model close to the value-added model popular in teacher evaluation.

Before getting into the model itself, I want to note that I view directly including skill in the production function as an alternative to including cumulative school and family inputs, as is suggested by Todd and Wolpin (2003). Because the previous period's ability is now a function of both schooling and family inputs from two periods ago as well as ability from two periods ago, this addition effectively allows previous periods of school and family inputs to influence current period ability production through the production of previous period ability.

Returning to the model itself, the model will be modified for periods $t > 2$ to be:

$$A_t = g(A_{t-1}, S_{t-1}, F_{t-1}, F_0, \mu) \quad \forall t = 2, \dots, T$$

Testing the coefficient on A_{t-1} should provide evidence for or against this particular extension. While it may be difficult to determine whether the direct inclusion of ability is preferable to the inclusion of cumulative schooling and family inputs, cumulative schooling and family inputs could be jointly tested against no inclusion at all as well.

3.3.3 Dynamic Function

In addition to the possibility of endogenous inputs, the human capital production function may not be constant over time. Indeed, there is evidence that is not (Cunha and Heckman, 2007; Todd and Wolpin, 2003). If our function is not constant, then we must subscript $g()$ to be period-specific. Our time-varying function is then:

$$A_1 = g_0(F_0, \mu)$$

$$A_t = g_t(A_{t-1}, S_{t-1}, F_{t-1}, F_0, \mu) \quad \forall t = 2, \dots, T$$

Notice that allowing our human capital production function to vary as the child ages also allows us to measure initial period production based on measures of previous family inputs before entering school, as well as endowments.

This addition can again be tested using seemingly unrelated regressions. Using SUR estimation to simultaneously model each period's production function, we can then test whether the production parameters in each period differ from each other.

3.3.4 Separating Out Cognitive and Noncognitive Production Functions

The final step that bridges the basic human capital production function from that of Cunha and Heckman (2008) is the expansion and bifurcation of the definition of ability. So far, ability has been defined as an abstract uni-dimensional; measure. However, economists are increasingly coming to see the benefits of multi-varied measures of ability, particularly of cognitive versus noncognitive ability (Cunha and Heckman, 2007). Besides describing different contributions to how they modify worker productivity in adulthood, these different forms of ability may also have different production functions. Taking this a step further, Cunha and Heckman (2007) present evidence that not only does skill beget skill, but that different types of skill have cross productivity in the production of other types of skill, such as emotional regulation aiding studying behavior. In the model, this means that both cognitive and noncognitive skills from last period appear directly in the production functions. Thus, we can replace A_t with the vector (C_t, N_t) for cognitive and noncognitive skills, respectively. Then, we split up our model into two differing functions, $g()$ and $f()$:

$$C_1 = g_0(F_0, \mu)$$

$$N_1 = f_0(F_0, \mu)$$

$$\begin{aligned} C_t &= g_t(C_{t-1}, N_{t-1}, S_{t-1}, F_{t-1}, F_0, \mu) \\ N_t &= f_t(N_{t-1}, C_{t-1}, S_{t-1}, F_{t-1}, F_0, \mu) \end{aligned} \quad \forall t = 2, \dots, T$$

This allows the production of cognitive and noncognitive human capital to be separate, but related processes. What increases cognitive capacity may have a more limited effect on noncognitive capacity, and vice versa.

Similar to the previous section, this can be tested with SUR. By simultaneously estimating the cognitive and noncognitive functions for each period, we can test for differences between within-period parameters across the two differing ability types.

3.4 Model Description

Cunha and Heckman (2008) outline a procedure for estimating a linear parametric version of this model. For the remainder of this section, I will summarize their proposed model and method. In this model, skills θ_t are divided into two categories: cognitive, θ_t^C , and noncognitive, θ_t^N . Each skill affects the production of its own type in the next period, as well as cross-productivity of the other skill type. Parental investment, θ_t^I , differentially effects each process, and for each process, there is an error component of omitted factors θ_t . The goal of this method is to estimate the resulting law of motion:

$$\begin{pmatrix} \theta_{t+1}^N \\ \theta_{t+1}^C \end{pmatrix} = \begin{pmatrix} \gamma_1^N & \gamma_2^N \\ \gamma_1^C & \gamma_2^C \end{pmatrix} \begin{pmatrix} \theta_t^N \\ \theta_t^C \end{pmatrix} + \begin{pmatrix} \gamma_3^N \\ \gamma_3^C \end{pmatrix} \theta_t^I + \begin{pmatrix} \eta_t^N \\ \eta_t^C \end{pmatrix} \quad (3.1)$$

This can be estimated both generally and per period. However, the researcher does not observe the true values of θ_t . Instead, the researcher observes m_t^k noisy measures, $t \in \{1, 2, \dots, N\}$, $k \in \{C, N, I\}$. Let $Y_{j,t}^k$, $j \in \{1, 2, \dots, m_t^k\}$ be the observed measure of type k , and let $\varepsilon_{j,t}^k$ be the error term. Then, in each period we have:

$$Y_{j,t}^C = \mu_{j,t}^C + \alpha_{j,t}^C \theta_t^C + \varepsilon_{j,t}^C \text{ for } j \in \{1, 2, \dots, m_t^C\} \quad (3.2)$$

$$Y_{j,t}^N = \mu_{j,t}^N + \alpha_{j,t}^N \theta_t^N + \varepsilon_{j,t}^N \text{ for } j \in \{1, \dots, m_t^N\} \quad (3.3)$$

$$Y_{j,t}^I = \mu_{j,t}^I + \alpha_{j,t}^I \theta_t^I + \varepsilon_{j,t}^I \text{ for } j \in \{1, \dots, m_t^I\} \quad (3.4)$$

For cognitive skills, noncognitive skills, and parental investment, respectively. The researcher thus seeks to identify the true underlying components, θ_j^k , from the intercepts $\mu_{j,t}^k$, the factor loadings $\alpha_{j,t}^k$, and the error terms $\varepsilon_{j,t}^k$. In order to do so, two normalizations must be made: $\alpha_{1,t}^k = 1$

and $E(\theta_t^k) = 0$. That is, factor loadings are identified relative to $\alpha_{1,t}^k = 1$, and, for $\theta \sim N(0, \Sigma)$, we seek to identify Σ .

3.4.1 First Assumptions

Following Cunha and Heckman (2008), I implement a method of semiparametric identification through covariance restrictions. This method rests on several assumptions. The first is Classical Measurement Error in $\varepsilon_{j,t}^k$, which is described by Cunha and Heckman for the case of two measurements per latent factor, $m_t^C = m_t^N = m_t^I = 2$, as follows:

“Assumption 1: $\varepsilon_{j,t}^k$ is mean zero and independent across agents and over time for $t \in \{1, \dots, T\}$; $j \in \{1, 2\}$; and $k \in \{C, N, I\}$;

Assumption 2: $\varepsilon_{j,t}^k$ is mean zero and independent of $(\theta_\tau^C, \theta_\tau^N, \theta_\tau^I)$ for all $t, \tau \in \{1, \dots, T\}$; $j \in \{1, 2\}$; and $k \in \{C, N, I\}$;

Assumption 3: $\varepsilon_{j,t}^k$ is mean zero and independent from $\varepsilon_{i,t}^l$ for $i, j \in \{1, 2\}$ and $i \neq j$ for $k = l$; otherwise $\varepsilon_{j,t}^k$ is mean zero and independent from $\varepsilon_{i,t}^l$ for $i, j \in \{1, 2\}$; $k \neq l$, $k, l \in \{C, N, I\}$ and $t \in \{1, \dots, T\}$.”

For the case of more than two measurements, we extend the assumption to $j \in \{1, \dots, m_t^k\}$.

3.4.2 Identifying factor loadings

Under these assumptions, the first step is to fix m^k for $t \in \{1, \dots, T\}$ and identify the factor loadings. Using cognitive skills as an example, this can be done using covariances of observed measures as follows :

$$\text{Cov}\left(Y_{1,t}^C, Y_{1,t+1}^C\right) = \text{Cov}\left(\theta_t^C, \theta_{t+1}^C\right), \quad (3.5)$$

$$\text{Cov}\left(Y_{j,t}^C, Y_{1,t+1}^C\right) = \alpha_{j,t}^C \text{Cov}\left(\theta_t^C, \theta_{t+1}^C\right), \text{ and} \quad (3.6)$$

$$\text{Cov}\left(Y_{1,t}^C, Y_{j,t+1}^C\right) = \alpha_{j,t+1}^C \text{Cov}\left(\theta_t^C, \theta_{t+1}^C\right) \quad (3.7)$$

Where $j \neq 1$. Equation 6 follows because we have normalized $\alpha_{1,t}^C = 1$. We can then identify $\alpha_{j,t}^C$ and $\alpha_{j,t+1}^C$ by taking the ratios of equations 7 and 6 and 8 and 6, respectively. By repeating this process for $t \in \{1, 2, \dots, N\}$, $j \in \{1, \dots, m^k\}$, and $k \in \{C, N, I\}$ we can identify all $\alpha_{j,t}^k$.

3.4.3 Identifying joint distribution of θ

Now that we have the factor loadings, we identify the joint distribution of $\{(\theta_t^C, \theta_t^N, \theta_t^I)\}_{t=1}^T$. First rewrite equations 3, 4, and 5 as

$$\frac{Y_{j,t}^k}{\alpha_{j,t}^k} = \frac{\mu_{j,t}^k}{\alpha_{j,t}^k} + \theta_t^k + \frac{\varepsilon_{j,t}^k}{\alpha_{j,t}^k}, j \in \{1, \dots, m^k\} \text{ for } \alpha_{j,t}^k \neq 0, k \in \{C, N, I\}; t \in \{1, \dots, T\}$$

Next, we can redefine Y_j , μ_j , and ε_j as

$$Y_j = \left\{ \left(\frac{Y_{j,t}^C}{\alpha_{j,t}^C}, \frac{Y_{j,t}^N}{\alpha_{j,t}^N}, \frac{Y_{j,t}^I}{\alpha_{j,t}^I} \right) \right\}_{t=1}^T \text{ for } j = \{1, \dots, m^k\}.$$

$$\mu_j = \left\{ \left(\frac{\mu_{j,t}^C}{\alpha_{j,t}^C}, \frac{\mu_{j,t}^N}{\alpha_{j,t}^N}, \frac{\mu_{j,t}^I}{\alpha_{j,t}^I} \right) \right\}_{t=1}^T \text{ for } j = \{1, \dots, m^k\}.$$

$$\varepsilon_j = \left\{ \left(\frac{\varepsilon_{j,t}^C}{\alpha_{j,t}^C}, \frac{\varepsilon_{j,t}^N}{\alpha_{j,t}^N}, \frac{\varepsilon_{j,t}^I}{\alpha_{j,t}^I} \right) \right\}_{t=1}^T \text{ for } j = \{1, \dots, m^k\}.$$

Finally, let θ denote the latent vector of skills and investment in all time periods

$$\theta = \left\{ (\theta_t^C, \theta_t^N, \theta_t^I) \right\}_{t=1}^T$$

Using these redefinitions, we can rewrite the measurement equations 3, 4, 5 as

$$Y_j = \mu_j + \theta + \varepsilon_j \text{ for } j \in \{1, \dots, m^k\}.$$

From here, we can identify the joint distribution of θ as well as the distributions of ε_1 and ε_2 .

3.4.4 Second Assumptions

The second assumption I am making through the use of this model is the independence of η^k from $(\theta_t^C, \theta_t^N, \theta_t^I)$ and serial independence of η_t^k over time. To reiterate, η_t^k is the residual term in equation 2. This assumption means that any omitted factors in the technology of skill formation are uncorrelated with the latent skill factors θ^C and θ^N and the latent investment factor θ^I . This last part, the independence with parental investment behavior, is likely the strongest assumption, and I somewhat ameliorate this with the inclusion of dummy variables for parents' highest education level in the final regression. I also rerun this model with individual fixed effects and find that the results are robust to unobserved individual heterogeneity.

3.4.5 Identifying Technology Parameters

Using the law of motion for noncognitive skills, I will outline the final step of this process under the assumptions outlined above. The law of motion for noncognitive skills is

$$\theta_{t+1}^N = \gamma_0^N + \gamma_1^N \theta_t^N + \gamma_2^N \theta_t^C + \gamma_3^N \theta_t^I + \eta_t^N \text{ for } t \in \{1, \dots, T\} \quad (3.8)$$

Define

$$\tilde{Y}_{1,t+1}^N = Y_{1,t+1}^N - \mu_{1,t+1}^N$$

$$\tilde{Y}_{1,t}^N = Y_{1,t}^N - \mu_{1,t}^N$$

$$\tilde{Y}_{1,t}^C = Y_{1,t}^C - \mu_{1,t}^C$$

$$\tilde{Y}_{1,t}^I = Y_{1,t}^I - \mu_{1,t}^I$$

Substitute in $\tilde{Y}_{1,t+1}^N, \tilde{Y}_{1,t}^N, \tilde{Y}_{1,t}^C, \tilde{Y}_{1,t}^I$ for $\theta_{t+1}^N, \theta_t^N, \theta_t^C, \theta_t^I$, respectively

$$\tilde{Y}_{1,t+1}^N = \gamma_0^N + \gamma_1^N \tilde{Y}_{1,t}^N + \gamma_2^N \tilde{Y}_{1,t}^C + \gamma_3^N \tilde{Y}_{1,t}^I + \left(\varepsilon_{1,t+1}^N - \gamma_{1,t}^N \varepsilon_{1,t}^N - \gamma_{2,t}^N \varepsilon_{1,t}^C - \gamma_{3,t}^N \varepsilon_{1,t}^I + \eta_t^N \right) \quad (3.9)$$

However, we cannot obtain consistent estimator of the technology parameters if we estimate equation 10 by OLS because $\tilde{Y}_{1,t}^N, \tilde{Y}_{1,t}^C, \tilde{Y}_{1,t}^I$ are correlated with ω_{t+1} , where

$$\omega_{t+1} = \varepsilon_{1,t+1}^N - \gamma_{1,t}^N \varepsilon_{1,t}^N - \gamma_{2,t}^N \varepsilon_{1,t}^C - \gamma_{3,t}^N \varepsilon_{1,t}^I + \eta_t^N$$

Instead, we can obtain consistent estimates by instrumenting for $\tilde{Y}_{1,t}^N, \tilde{Y}_{1,t}^C, \tilde{Y}_{1,t}^I$ with $\{Y_{j,t}^N, Y_{j,t}^C, Y_{j,t}^I\}_{j=2}^{m^k}$ with two-stage least squares. This follows because the Classical Measurement Error assumption and the η_t^N assumption give us our exclusion restrictions. Thus, once we have transformed our measures with the identified factor loadings and intercepts, we can use two-stage least squares to instrument for our latent factors and estimate the technology parameters in our law of motion.

3.4.6 Anchoring factors

The final piece of this process involves anchoring the scale of our factors in a long-term outcome variable of interest. Anchoring is necessary in order to avoid a common problem with test scores: lack of uniqueness to affine transformations. To solve this, education researchers often standardize their test scores. However, as the purpose of this paper is to identify how boys and girls come to have differing behavioral outcomes, I believe following Cunha and Heckman's process of anchoring can be beneficial with the right choice of outcome to anchor to: avoiding eighth grade suspensions. Let Y be a binary that indicators whether the student had an out-of-school suspension in eighth grade. I estimate the following equation using a linear probability model

$$Y = \mu_T + \delta_N \theta_T^N + \delta_C \theta_T^C + \varepsilon \quad (3.10)$$

where ε is not correlated with either θ or $\varepsilon_{j,t}^k$. For any affine transformation of θ_T^k , following a change in levels of scores, δ_k will adjust. Thus, while neither θ_T^k nor δ_k are uniquely determined, $\delta_k \theta_T^k$ is. Let

$$D = \begin{pmatrix} \delta_N & 0 \\ 0 & \delta_C \end{pmatrix}$$

By working with $\delta_k \theta_t^k$ and $\delta_k \theta_{t+1}^k$ in our law of motion estimation instead of θ_t^k and θ_{t+1}^k , I am effectively estimating the following transformation of equation 2 instead:

$$D\theta_{t+1} = \left(DAD^{-1}\right)(D\theta_t) + (DB)\theta_t^I + (D\eta_t) \quad (3.11)$$

The result is that while the self-productivity terms will be unchanged, the cross-productivity and investment terms are scaled.

3.5 Data

I use two different versions of the Early Childhood Longitudinal Study, Kindergarten Cohort datasets for this analysis: the ECLS-K and the ECLS-K:2011. Both studies are nationally representative samples of children who entered kindergarten in the 1998-1999 and 2010-2011 school years, respectively, referred to for the rest of this paper as the 1998 and 2010 cohorts. Additionally, both studies contain data on about 8,000⁴ children, parents, and teachers interviewed repeatedly in several waves. The ECLS-K conducted interviews in fall of kindergarten, spring of kindergarten, fall of 1st grade, and spring of 1st, 3rd, and 5th grades. The ECLS-K:2011 conducted interviews in spring and fall of kindergarten, 1st grade, and 2nd grade, as well as spring of 3rd, 4th, and 5th grades. In both of these studies, information was collected about children's cognitive, social, emotional, and physical development by interviewing children, parents, teachers, and administrators. Additional information was collected on the children's home environment, including parental educational activities, the environment at school, and school and teacher practices and qualifications.

For my analysis sample, I keep only observations with non-missing data on cognitive, non-cognitive, and parental investment measures in each of the four time periods I consider: kindergarten, first grade, third grade, and fifth grade. Of the 15,696 observations in the fifth-grade panel samples between the two datasets, I drop 3,359 observations for missing noncognitive measures in one of the four periods, 2,508 for missing parental investment measures, and 635 for missing other controls.⁵ This leaves a final analysis sample of 9,045 students between the two datasets. I use inverse probability weighting to adjust the final analysis sample to match the

⁴8,370 and 7,326 for the fifth grade panel samples of the 1999 and 2011 cohorts, respectively.

⁵Gender, race, urbanicity, parental education in kindergarten, and parental socioeconomic status in kindergarten.

full panel samples in Gender, race, urbanicity, parental education in kindergarten, and parental socioeconomic status in kindergarten. For the main analysis, data are reshaped to the student-year level, which creates a final panel sample size of 36,180 student-year observations.

3.5.1 Key Measures

Following Bertrand and Pan (2013), I have created an index of parental inputs: a HOME index, which standardizes the average of eight measures of parental investment activities⁶. I also create an index of all remaining parental investment measures in each wave of each survey, which I call the Other Parental Investment index. Like the HOME index, this index standardizes the average of all remaining measures, representing the degree to which each student's parents engaged in other investment-related activities, relative to other parents in the survey.

In addition to data on parent-reported investment and child-rearing activities and attitudes, another important aspect of both ECLS-K datasets is their measures of non-cognitive skills, particularly teacher-reported noncognitive skills⁷. Both datasets contain teacher-reported measures on externalizing behaviors, self-control, approaches to learning, interpersonal skills, and internalizing problems. These social skills scales were developed based on teachers' responses to items taken from the Social Skills Rating System. The score on each scale is the mean rating of all items included in the scale. Although the components of these measures are not available due to copyright reasons, the ECLS-K user's manual provides descriptions of each of the noncognitive measures (Tourangeau et al., 2001). Approaches to learning is constructed from "six items that rate the child's attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization." Self-control is constructed from "four items that indicate the child's ability to control behavior by respecting the property rights of others, controlling temper, accepting peer ideas for group activities, and responding appropriately to pressure from peers." Interpersonal skills are constructed from items that "rate the child's skill in forming and maintaining friendships,

⁶Measures include whether: read to child ≥ 3 times per week, child has ≥ 20 books, child reads ≥ 3 times per week outside school, have home computer child uses, has visited museum, concert, or library with child, and whether child participated in other outside school activities (dance, sports, music, etc.).

⁷Parent-reported ratings of noncognitive skills are also available for early grades. However, I do not include them in this analysis, as I believe that parents are less likely to be objective, unbiased assessors of their children's abilities.

getting along with people who are different, comforting or helping other children, expressing feelings, ideas and opinions in positive ways, and showing sensitivity to the feelings of others.” Externalizing problem behaviors are constructed from “Five items on this scale rate the frequency with which a child argues, fights, gets angry, acts impulsively, and disturbs ongoing activities.” And finally, internalizing problem behaviors is constructed from four items that ask about “the apparent presence of anxiety, loneliness, low self-esteem, and sadness.”

Two of these measures, externalizing behavior and internalizing behavior, have been reordered so that higher scores indicate the child exhibited a “better” score reflecting higher noncognitive skill in each respective category, which means more of each measure’s behaviors for positive scales (approaches to learning, self-control, and interpersonal skills) or less of each measure’s behaviors for negative scales (externalizing and internalizing problems). Additionally, in order to allow for comparability and reduce arbitrary scaling, all noncognitive measures, including externalizing behavior, are standardized within the estimated population of their respective surveys.

Both the Social Skills Rating System itself and these measures from the ECLS-K based on the SSRS are also used in numerous other studies involving the ECLS-K. This includes Neidell and Waldfogel (2010) (2010), who state “these scales have high construct validity as assessed by test-retest reliability, internal consistency, inter-rater reliability, and correlations with more advanced behavioral constructs (Elliott et al., 1988) and are considered the most comprehensive social skill assessment that can be widely administered in large surveys such as the ECLS-K (Demaray et al., 1995).” Taken together, these endorsements and descriptions provide evidence for the validity of the measures of noncognitive skills I will be using for the remainder of this paper.

As described in Section 3.4, the Cunha and Heckman (2008) model requires choosing two measures for each category of cognitive skill, non-cognitive skill, and parental investment in order to address measurement error. The measures of the latent factors used here are as follows: math and reading scores for cognitive measures, externalizing behavior and self-control for noncognitive measures, and the two indices described above, the HOME and Other Investment indices, for investment measures. The first measure listed in each of these categories was treated as the primary

measure for which the loading factor was normalized.

3.5.2 Descriptive Statistics

Table B.1 shows means for demographic characteristics for three groups: the analysis sample, with and without inverse probability weights, and the full fifth-grade panel sample. As the column with inverse probability weights (IPWs) shows, the inverse probability weights effectively correct imbalances in the analysis sample induced by item nonresponse to noncognitive and parental investment questions in the survey. For example, column 3 shows that without inverse probability weighting, the analysis sample would overrepresent white, higher-income, and higher-parentally-educated students relative to the nationally representative samples that the full sample is supposed to represent. Means between columns 2 and 4 do not match perfectly due to capping inverse probability weights at an overweighting of 10, to prevent an over-reliance on a small number of observations.

In Table B.2 and Figures A.1 and A.2, I compare metrics of parental investment for both the HOME and Other Investment Indices that will be used in the model. Figures A.1 and A.2 show the evolution over time in mean indices of parent-reported activities from kindergarten to fifth grade. They show that parents report a substantially larger investment in the development of girls compared to boys. Except for the Other Investment Index at Kindergarten, both figures show gender gaps in reported investment activities of 0.1-0.2 standard deviations across all grades. Table B.2 reports the distribution of these gender gaps by showing the percentage of boys and girls (in columns 3 and 4) who fall into each quintile of investment for the HOME index (the top panel) and the Other Investment Index (the bottom panel) within their respective grades. Column 2 shows the difference between the percentage of girls versus boys in each of the indices and tests for joint significance of these differences in the second-to-last row. The results in Table B.2 show that the gender gaps in the means appear largely in the tails of the distribution, with boys more likely to appear in the bottom two quintiles and girls more likely to appear in the top quintile. Overall, Table B.2 and Figures A.1 and A.2 confirm that parents are reporting more investment activities with girls than boys both across grades and across the distributions of these indices, supporting the case for nurture playing

a causal role in the existence of gender gaps.

3.6 Results

Although Table B.2 and Figures A.1 and A.2 show that parents report engaging in substantially more investment activities with girls rather than boys, it remains an open question whether the gender gap in noncognitive skills would be eliminated in the counterfactual where parental investment was equalized. To this end, I apply the Cunha and Heckman (2008) model to my data using their suggested strategies for both modeling the process of skill formation and for dealing with issues of measurement error and tractability.

3.6.1 Estimating Equations

In applying the model laid out in Section 3.4 to my data, I use instrumental variables to estimate the following reduced-form equations:

$$\begin{aligned} Y_{i,t+1}^N &= \beta_0 + \beta_1 Y_{i,t}^N + \beta_2 Y_{i,t}^C + \beta_3 I_{i,t} + \beta_4 female_i + \beta_5 female_i \times I_{i,t} + \delta X_i + \alpha_{s,t} + \epsilon_{i,t} \\ Y_{i,t+1}^C &= \beta_0 + \beta_1 Y_{i,t}^C + \beta_2 Y_{i,t}^N + \beta_3 I_{i,t} + \beta_4 female_i + \beta_5 female_i \times I_{i,t} + \delta X_i + \alpha_{s,t} + \epsilon_{i,t} \end{aligned} \quad (3.12)$$

Where $Y_{i,t}^N$ and $Y_{i,t}^C$ are standardized measures of non-cognitive and cognitive skills, respectively, for student i in period (grade) t , and $Y_{i,t+1}^N$ and $Y_{i,t+1}^C$ are the same measures for the following period (grade). I use externalizing behavior and math scores as the two primary measures of non-cognitive and cognitive skills. $I_{i,t}$ is a measure of parental investment in these skills, for which I use the HOME index from Bertrand and Pan (2013) as my primary measure. $female_i$ is a binary indicator for whether the student is female. $female_i \times I_{i,t}$ is the interaction of parental investment and $female_i$, which tests whether parental investment has a differing marginal effect on boys versus girls $\alpha_{s,t}$ represents a school fixed-effect for the school attended by student i in period (grade) t . And lastly, X_i is a vector of time-invariant controls that includes socioeconomic status, parental education, and urbanicity as measured in kindergarten. Then, β_5 is the coefficient of interest, testing (under modeling assumptions) whether boys and girls have differing responsiveness to parental investment.

However, as Cunha and Heckman (2008) note, if we estimate the value-added equation 3.12 directly, we may have biased estimates due to measurement error. To resolve this, I apply instrumental variables using rescaled and demeaned secondary measures of non-cognitive, cognitive, and investment as instruments. These three instruments are self-control, reading scores, and the other investments index.⁸

3.6.2 Findings

Tables B.3 and B.4 show the coefficients on parental investment and the parental investment-female interactions terms for non-cognitive and cognitive skills, respectively. Columns 5 and 6 in both Tables show the results of the Cunha and Heckman (2008) model, without and with individual fixed effects, respectively.⁹ Columns 1-4 build up the Cunha and Heckman (2008) model from the simplest model, with skills in $t + 1$ regressed on parental investment in t in Column 1, to a more complete model that includes controls for parental education, socioeconomic status, and urbanicity, school fixed effects, and a value-added specification that recognizes the ways in which skills beget skills (Cunha and Heckman, 2007), with column 4 replicating the reduced form equation 3.12. For columns 1-4, the cognitive measure used is math scores, the non-cognitive variable used is externalizing behavior, and the parental investment measure is the HOME index. Together, these tables show not just the results of my preferred model in Column 5, but also the role that each aspect of the model plays in producing the results in Column 5. In the next section, I will also explore some alternative methods of addressing the concerns that the Cunha and Heckman (2008) model attempts to fix.

As the results in column 5 of Tables B.3 and B.4 show, not only is there no evidence for a gender difference in the effects of parental investment, but there is also little evidence that the parental

⁸For their primary measures, Cunha and Heckman (2008) use PIAT Mathematics scores, the Antisocial Score of the Behavioral Problem Index (BPI), and family income for cognitive, non-cognitive, and parental investment latent factors, respectively. For instruments, they use PIAT Reading recognition scores, Anxiety, Headstrong, Hyperactivity, and Peer Conflict BPI Scores, and the Home Observation Measurement of the Environment - Short Form measures in the NLSY are used as instruments for cognitive, non-cognitive, and parental investment latent factors. The specific measures they use from the HOME-SF are the number of books, number of musical instruments, newspaper subscriptions, special lessons, trips to museums, and trips to the theater.

⁹Cunha and Heckman (2008) recommend using first differencing instead of individual fixed effects for concerns about unobserved individual heterogeneity, but both methods produce the same results with some minor differences in assumptions necessary for consistency.

investment measures present in the ECLS-K contribute to non-cognitive and cognitive development at all. Columns 1-4 show us that while the inclusion of controls plays a large role in addressing identification concerns about the correlation between family background, resources, and education may play in driving both investment and outcomes, it is not sufficient to address the issue. Both school fixed effects and controlling for lagged ability are necessary to identify the role of parental investment in producing skills. Notably, these factors are mostly sufficient for explaining away the role of parental investment, as shown in column 4. With the presence of controls, school fixed effects, and lagged ability, the coefficients on parental investment and its gender interaction become insignificant at the 0.1 level for non-cognitive skills, as well as gender difference in the effects of parental investment on cognitive skills. Measurement error correction, on the other hand, has no effect on the role of parental investment in non-cognitive skills, while playing some role in severing the link between parental investment and cognitive skills in Table B.4. Ultimately, these results suggest that not only is properly capturing the role of non-school parental investment challenging, but seeing the acquisition of skills as a dynamic process correlated with background characteristics is necessary for developing a more accurate model of childhood human capital development.

3.7 Robustness

Although columns 1-4 of Tables B.3 and B.4 show the influence of some of the modeling assumptions in producing the null results of the Cunha and Heckman (2008) model in columns 5 and 6, there are several additional modeling assumptions that could be affecting the results. In this section, I test the robustness of my findings by examining alternatives to the two primary attributes of the Cunha and Heckman (2008) model: 1) their approach to addressing issues of measurement error and the “curse of dimensionality”, and 2) their use of a value-added model that excludes inputs before time period t .

3.7.1 Testing Dimension Reduction and Measurement Error Correction

Two major related issues in the pursuit of modeling childhood skill formation can play a major role in affecting estimates: the curse of dimensionality and measurement error. Essentially, the researcher observes a number of different measures, such as the five different measures of non-

cognitive skill and large number of different parental investment questions in the ECLS-K datasets, and must decide what to do with them. Putting all variables individually into the model is the most flexible and comprehensive approach, but, in addition to creating a large degree of complexity which hampers tractability, also potentially introduces a lot of noise if these measures are correlated and are only weakly related to the underlying concept the researcher is trying to measure.¹⁰ The researcher then has a choice between picking and choosing individual measures to prioritize, such as Bertrand and Pan (2013)'s choice of externalizing behavior and the HOME index, which is potentially vulnerable to researcher¹¹ and measurement error, or combining measures into a smaller number of measures. The latter approach, combining measures, has two variations: 1) indexing, which creates a simple average of a group of measures, and 2) factor or principal component analysis, which takes a weighted average of a group of measures based on their correlation with each other. Bertrand and Pan (2013) took this first approach in creating the HOME index, and I've extended it in creating the Other Parental Investment index. Cunha and Heckman (2008) propose taking the second approach, with their version of factor analysis. Although it introduces complication, factor analysis has the advantage of reducing the potential for measurement error, assuming that each individual measure is a noisy measure of an underlying latent factor, non-cognitive skill or parental investment, that is commonly correlated between the available measures. See Johann (2020) for a more in-depth description of factor analysis and its benefits and drawbacks.

For estimating parental investment in my preferred specification, I combine these two approaches, indexing and factor analysis, by creating two indices of parental investment and, following Cunha and Heckman (2008), using the common covariance between the two indices over time as the true latent measure of parental investment. However, in addition to relying on Cunha and Heckman (2008)'s preferred method of factor analysis, this approach is still vulnerable to researcher error by requiring the researcher to divide the set of parental investment measures into two different groups before using the common variation. As a robustness check, I use factor analysis within

¹⁰In the case of the parental investment measures, these questions changed between waves and between datasets, making estimation even more challenging.

¹¹Whether the researcher truly picked the correct individual measures.

each wave (grade) and dataset that uses the correlation between the set of all parental investment measures within each grade-cohort group to determine a weighted average of the measures.¹²

The process of dimension reduction for non-cognitive and cognitive skills is more straightforward since there are fewer measures to work with. For generating a single measure of latent non-cognitive skill, I use a single latent factor between all five non-cognitive measures, following Johann (2020). For generating a single measure of cognitive skill I take a simple average of math and reading scores and standardize the result.

Column 2 of Table B.5 reruns a value-added model following reduced-form equation 3.12 and, like Tables B.3 and B.4, shows the coefficients on parental investment and the parental investment-female interaction terms. Columns 3 and 4 take an alternative approach instead of dimension reduction: including all components. All five non-cognitive skills are included on the right-hand side, along with both math and reading scores, and, rather than serving as instruments for each other, both parental investment indices are included as well. The estimates presented in columns 3 and 4 are the *sum* of the coefficients on the HOME and Other Parental Investment indices (in the first and fourth rows) as well as the sum of the coefficients of their interaction terms with the female indicator. Column 3 shows results for estimates with externalizing behavior (in the first panel) and math scores (in the second panel) on the left-hand side of equation 3.12 and Column 4 shows results for estimates with self-control (in the first panel) and reading scores (in the second panel) on the left-hand side.

Comparing the preferred Cunha and Heckman (2008) estimates in column 1 to column 2, we can see that taking an alternative approach to dimension reduction does not seem to affect the null effects of parental investment on the production of non-cognitive or cognitive skills. This suggests that the main estimates, in columns 5 and 6 in Tables B.3 and B.4, are not driven by the particular approach to measurement error correction and dimension reduction suggested by Cunha and Heckman (2008). Columns 3 and 4, which attempt to limit dimension reduction to the extent

¹²Since the goal here is tractability I use a single latent factor selected by its largest eigenvalue among potential latent factors, though information criterion suggests that 2-3 factors may be more appropriate for capturing the full spectrum of common variation between measures.

feasible, find similar null results for externalizing behavior and reading scores, but interestingly, find significant effects for self-control and reading scores. This presents the possibility that attempting to use only common variation between the left-hand side variables in columns 3 and 4, such as through factor analysis, may be overly restrictive, although the possibility of spurious results due to the multiple comparisons problem suggests we should view these findings with caution.

3.7.2 Testing the Value-Added Model

A second assumption of the Cunha and Heckman (2008) model that could be driving my null results is the use of a value-added model. A value-added model assumes that all inputs in period t are sufficient for capturing effects on outcomes in period $t + 1$, following equation 3.12. However, if effects persist over time, then this will leave out important parts of the skill production function. Todd and Wolpin (2003) propose instead using a cumulative model, in which *all* previous inputs are included on the right-hand side. Rewriting equation 3.12 to reflect the Todd and Wolpin (2003) model gives us:

$$\begin{aligned} Y_{i,t+1}^N &= \beta_0 + \sum_{j=1}^t (\beta_j Y_{i,j}^N + \gamma_j Y_{i,j}^C + \kappa_j I_{i,j} + \zeta_j female_i \times I_{i,j} + \alpha_{s,j}) + \beta_4 female_i + \delta X_i + \epsilon_{i,t} \\ Y_{i,t+1}^C &= \beta_0 + \sum_{j=1}^t (\beta_j Y_{i,j}^C + \gamma_j Y_{i,j}^N + \kappa_j I_{i,j} + \zeta_j female_i \times I_{i,j} + \alpha_{s,j}) + \beta_4 female_i + \delta X_i + \epsilon_{i,t} \end{aligned} \quad (3.13)$$

Now, all terms from $t = 1$ up to the current period are included on the right-hand side of the equation. Intuitively, this reflects the possibility that inputs could take several periods to "sink in", rather than only immediately affecting the next period. Fortunately, this addition is straightforward to test with a joint test of terms preceding period t .

Tables B.6 and B.7 apply this alternative specification in columns 2-4, using the dimension reduction methods from the previous section. Because the number of additional terms varies by grade, results are reported separately by grade t (as in, the effects of inputs in grade t and before on skills in grade $t + 1$). Similar to Table B.5, the sum of relevant coefficients are shown for parental investment and the parental investment-female interaction terms. However, now the sum shown

includes coefficients before t , rather than only in period t . Additionally, for grades 1 and 3, an additional row has been added showing the p-value of a joint test that all coefficients on parental investment and parental investment-female interaction terms before period t are jointly zero.

Tables B.6 and B.7 confirm that the findings of the main specifications are robust to the cumulative specification of Todd and Wolpin (2003). Like Table B.5, almost all parental investment and parental investment-female interaction terms are not statistically significant at the 0.1 level, indicating that the null results on parental investment are not driven by the exclusion of the terms before period t . Interestingly, the joint tests of time periods before t are all statistically significant at the 0.01 level. This is explained by the fact that the individual terms have differing signs, even though they are not individually significant.

3.8 Conclusion

While plenty of research has documented gender differences in cognitive (Dee, 2007; Cornwell et al., 2013; Fortin et al., 2015), and non-cognitive (Bertrand and Pan, 2013; Johann, 2020) skills, the question of what may be causing these gaps remains unanswered. Although we can observe differences in the amount of time and energy parents invest in girls versus boys (Baker and Milligan, 2016; Bibler, 2018) and observe boys responding more poorly to disadvantaged environments (Autor et al., 2019, 2020; Chetty et al., 2016; Kling et al., 2005), it remains an open question whether these gender gaps in skill arise from nature versus nurture. From an optimal policy perspective, this leaves us guessing whether the best solution to resolving gender gaps is to equalize inputs, or whether to adopt differing strategies for children of each gender.

In this paper, I apply a popular model on the development of skills over time proposed by Cunha and Heckman (2008) to the ECLS-K:1998 and ECLS-K:2011 datasets. In doing so, I am able to model skill formation as a multi-dimensional and dynamic process vulnerable to measurement error. While I am able to show significant gender gaps in parental investment throughout elementary school in these nationally representative datasets, my modeling results do not show a significant impact of parental investment, as captured by the measures available in the ECLS-K datasets, on either cognitive or non-cognitive skill production. I further test whether these findings are driven

by either the Cunha and Heckman (2008) model's use of measurement error correction or by its reliance on a value-added, rather than cumulative, model, and find that the null results are robust to these alternative specifications. Not only are there no gender differences in responsiveness to parental investment, but parental investment also does not appear to aid skill formation at all once the previous period's skills are taken into account.

The findings of this paper emphasize the challenges of answering questions about early childhood skill formation. Not only does parental investment often suffer from a lack of a clean cardinal measure for regression analysis, there are also a large number of potential activities that could be considered parental investment, and deciding which ones matter and how to include them in estimation in a tractable manner remains an ongoing challenge. Bibler (2018) attempts to overcome this by focusing on time spent, but even there, it's not clear whether time itself is the most important factor, and time estimates can be subject to significant measurement error themselves. Given these issues, researchers attempting to work in this area should attempt and show results from a wide number of alternative parental investment measures before settling on a specific preferred strategy to avoid the significant potential for researcher and measurement error.

BIBLIOGRAPHY

- Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments. *Unpublished Manuscript, Arizona State University*.
- Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2020). It takes a village: the economics of parenting with neighborhood and peer effects. Technical report, National Bureau of Economic Research.
- Agostinelli, F. and G. Sorrenti (2018). Money vs. time: family income, maternal labor supply, and child development. *University of Zurich, Department of Economics, Working Paper* (273).
- Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11.
- 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.
- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics* 11(3), 338–81.
- Autor, D. H., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2020). Males at the tails: How socioeconomic status shapes the gender gap. *NBER Working Paper* (w27196).
- Baker, M. and K. Milligan (2016). Boy-girl differences in parental time investments: Evidence from three countries. *Journal of Human Capital* 10(4), 399–441.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences* 6(6), 248–254.
- Baron-Cohen, S. (2003). *The Essential Difference: Men, Women, and the Extreme Male Brain*. London: Allan Lane.
- Baum-Snow, N., D. A. Hartley, and K. O. Lee (2019). The long-run effects of neighborhood change on incumbent families.
- Becker, G. S., W. H. Hubbard, and K. M. Murphy (2010). The market for college graduates and the worldwide boom in higher education of women. *American Economic Review* 100(2), 229–33.
- 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.
- Bibler, A. (2018). Household composition and gender differences in parental time investments.

Available at SSRN 3192649.

- Black, S. E. and P. J. Devereux (2010). Recent developments in intergenerational mobility. *New This Week*, 2 – 90.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics* 93(2), 455–467.
- Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* 81(3), 855–882.
- Cascio, E. U. and D. W. Schanzenbach (2016). First in the class? age and the education production function. *Education Finance and Policy* 11(3), 225–250.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *The American Economic Review* 104(9), 2633 – 2679.
- Chetty, R., N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi (2016). Childhood environment and gender gaps in adulthood. *American Economic Review* 106(5), 282–88.
- Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review* 108(10), 3028–56.
- Cornwell, C., D. B. Mustard, and J. Van Parys (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources* 48(1), 236–264.
- 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.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources* 42(3), 528–554.
- Demaray, M. K., S. L. Ruffalo, J. Carlson, R. Busse, A. E. Olson, S. M. McManus, and A. Leventhal (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review* 24(4), 648–671.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics* 132(4), 1593–1640.
- Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School choice, school quality,

- and postsecondary attainment. *The American Economic Review* 104(3), 991 – 1013.
- DiPrete, T. A. and J. L. Jennings (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research* 41(1), 1–15.
- Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, et al. (2007). School readiness and later achievement. *Developmental psychology* 43(6), 1428.
- Elder, T. E. (2010). The importance of relative standards in adhd diagnoses: evidence based on exact birth dates. *Journal of Health Economics* 29(5), 641–656.
- Elder, T. E. and D. H. Lubotsky (2009). Kindergarten entrance age and children’s achievement impacts of state policies, family background, and peers. *Journal of Human Resources* 44(3), 641–683.
- Elliott, S. N., F. M. Gresham, T. Freeman, and G. McCloskey (1988). Teacher and observer ratings of children’s social skills: Validation of the social skills rating scales. *Journal of Psychoeducational Assessment* 6(2), 152–161.
- Fernández, A. B. (2021). Neighbors’ effects on university enrollment. *American Economic Journal: Applied Economics* (forthcoming).
- Fortin, N. M., P. Oreopoulos, and S. Phipps (2015). Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources* 50(3), 549–579.
- Gensowski, M., R. Landersø, B. Dorte, P. Dale, A. Højen, and L. Justice (2020). Public and parental investments and children’s skill formation. *The ROCKWOOL Foundation Research Unit* (155).
- Goldin, C., L. F. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives* 20(4), 133–156.
- Heckman, J. J. and R. Landersø (2021). Lessons from denmark about inequality and social mobility. Technical report, National Bureau of Economic Research.
- Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics* 6(1), 689–733.
- 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.
- Imberman, S. A., A. D. Kugler, and B. I. Sacerdote (2012). Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review* 102(5), 2048–82.

- Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes : Evidence from school finance reforms. *The Quarterly Journal of Economics* 131(1), 157 – 218.
- Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review* 21(6), 589–598.
- Johann, A. (2020). The increasing fragility of boys: Examining changes in levels and correlates of gender gaps in noncognitive skills over time. Available at <https://sites.google.com/site/alwjohann/working-papers/gender-gaps-in-noncognitive-skills>.
- Kling, J. R., J. Ludwig, and L. F. Katz (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics* 120(1), 87–130.
- Knickmeyer, R., S. Baron-Cohen, P. Raggatt, and K. Taylor (2005). Foetal testosterone, social relationships, and restricted interests in children. *Journal of Child Psychology and Psychiatry* 46(2), 198–210.
- Laliberté, J.-W. P. (2018). Long-term contextual effects in education: Schools and neighborhoods. *University of Calgary, unpublished manuscript*.
- Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics* 3(1), 101–28.
- List, J. A., F. Momeni, and Y. Zenou (2020). The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods. Technical report, National Bureau of Economic Research.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3), 531–542.
- Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics* 92(3), 562–576.
- Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The quarterly journal of economics* 118(4), 1533–1575.
- Raver, C., P. W. Garner, and R. Smith-Donald (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In *School readiness and the transition to kindergarten in the era of accountability*, pp. 121–147. Paul H Brookes Publishing.
- Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics* 6(1), 253–272.

- Sanbonmatsu, L., L. F. Katz, J. Ludwig, L. A. Gennetian, G. J. Duncan, R. C. Kessler, E. K. Adam, T. McDade, and S. T. Lindau (2011). Moving to opportunity for fair housing demonstration program: Final impacts evaluation.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518 – 529.
- Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* 113(485), F3–F33.
- Tourangeau, K., J. Burke, T. Le, S. Wan, M. Weant, E. Brown, N. Vaden-Kiernan, E. Rinker, R. Dulaney, K. Ellingsen, B. Barrett, I. Flores-Cervantes, N. Zill, J. Pollack, D. Rock, S. Atkins-Burnett, and S. Meisels (2001). *ECLS-K Base Year Public-Use Data Files and Electronic Codebook*. Washington, DC: National Center for Education Statistics: U.S. Department of Education. (NCES 2001-029).

APPENDIX A FIGURES

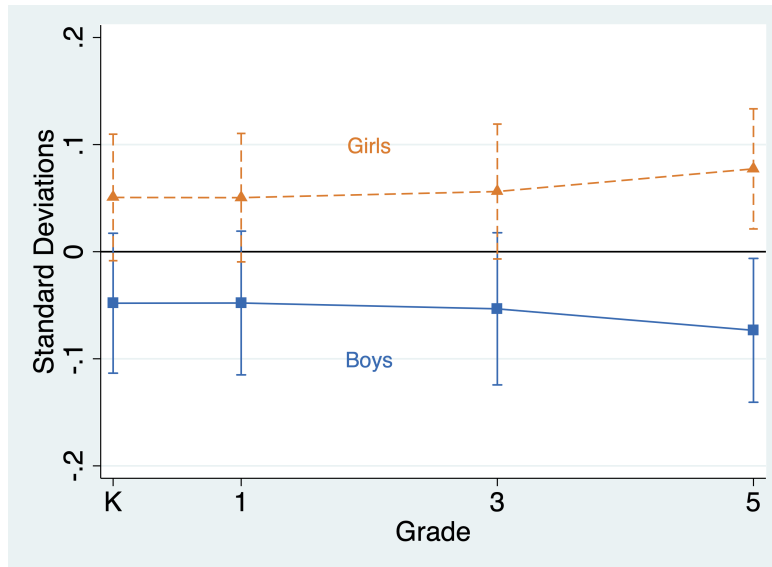


Figure 3.A.1 Parental Investment by Grade, HOME Index

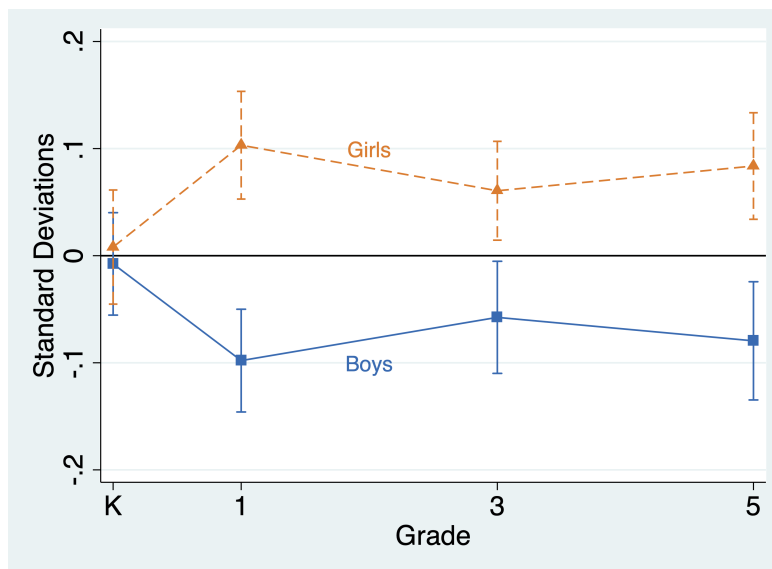


Figure 3.A.2 Parental Investment by Grade, Other Index

APPENDIX B TABLES

Table 3.B.1 Summary Statistics of Means, by Sample and Weighting

Variable	Analysis Sample		Full Sample Original Weights
	IPWs	Original Weights	
White	0.55	0.64	0.55
Black	0.15	0.11	0.15
Hispanic	0.21	0.17	0.22
Asian	0.03	0.03	0.04
Other race	0.05	0.05	0.05
Male	0.51	0.50	0.51
Female	0.49	0.50	0.49
SES Quintile [†] : 1st	0.18	0.13	0.18
SES Quintile [†] : 2nd	0.21	0.19	0.21
SES Quintile [†] : 3rd	0.22	0.22	0.22
SES Quintile [†] : 4th	0.20	0.23	0.20
SES Quintile [†] : 5th	0.19	0.24	0.19
Mother's Education [†] : Less than HS	0.11	0.07	0.11
Mother's Education [†] : High School	0.25	0.22	0.25
Mother's Education [†] : Some college	0.34	0.34	0.34
Mother's Education [†] : College or greater	0.30	0.37	0.30
Lives in City [†]	0.34	0.31	0.35
Lives in Suburb [†]	0.38	0.38	0.38
Lives in Rural Area [†]	0.28	0.31	0.28
Observations	63,315	63,315	109,872

[†]As measured in kindergarten

IPWs = Inverse Probability Weights

Table 3.B.2 Distribution of Parental Investment

Investment Variable	Difference	Girls	Boys
HOME Index			
Quintile 1	-0.044** [0.012]	0.230	0.274
Quintile 2	-0.009 [0.006]	0.190	0.199
Quintile 3	0.020* [0.008]	0.293	0.273
Quintile 4	0.014* [0.007]	0.129	0.114
Quintile 5	0.018* [0.008]	0.157	0.139
Joint Test <i>p</i> -value	0.001		
Observations	36,180		
Other Investment Index			
Quintile 1	-0.029** [0.009]	0.197	0.226
Quintile 2	-0.019** [0.007]	0.186	0.206
Quintile 3	-0.000 [0.007]	0.231	0.232
Quintile 4	0.009 [0.006]	0.179	0.170
Quintile 5	0.040** [0.007]	0.206	0.167
Joint Test <i>p</i> -value	0.000		
Observations	36,180		

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets.

Columns 3 and 4 show the percentage of girls and boys, respectively, who fall into each quintile within their grade. Column 1 Shows the difference between columns 2 and 3. The second-to-last row of each panel, "Joint Test *p*-value, shows the *p*-value from a joint test of the all five of the differences in quintiles being equal to zero.

Table 3.B.3 Effects of Parental Investment on Non-Cognitive Skills

	(1)	(2)	(3)	(4)	(5)	(6)
Parental Investment	0.086** [0.016]	0.046** [0.017]	0.026+ [0.015]	0.000 [0.014]	-0.005 [0.009]	-0.050 [0.040]
Parental Investment x Female	-0.040* [0.020]	-0.040* [0.020]	-0.023 [0.018]	-0.015 [0.017]	-0.000 [0.009]	0.050 [0.041]
Observations	27,135	27,135	26,751	26,604	26,385	26,303
Controls	N	Y	Y	Y	Y	Y
School Fixed Effects	N	N	Y	Y	Y	Y
Value-Added	N	N	N	Y	Y	Y
Measurement Error Correction	N	N	N	N	Y	Y
Individual Fixed Effects	N	N	N	N	N	Y

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets.

Table 3.B.4 Effects of Parental Investment on Cognitive Skills

	(1)	(2)	(3)	(4)	(5)	(6)
Parental Investment	0.295** [0.015]	0.137** [0.014]	0.097** [0.012]	0.036** [0.009]	0.001 [0.002]	-0.005 [0.007]
Parental Investment x Female	-0.052** [0.020]	-0.044* [0.019]	-0.027+ [0.016]	-0.008 [0.012]	-0.001 [0.002]	0.004 [0.008]
Observations	27,042	27,042	26,660	26,540	26,324	26,209
Controls	N	Y	Y	Y	Y	Y
School Fixed Effects	N	N	Y	Y	Y	Y
Value-Added	N	N	N	Y	Y	Y
Measurement Error Correction	N	N	N	N	Y	Y
Individual Fixed Effects	N	N	N	N	N	Y

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets.

Table 3.B.5 Testing Alternative Methods of Dimension Reduction

	Main Model	Factor Analysis	Components	
			Externalizing	Self Control
Non-Cognitive Skills				
Parental Investment	-0.005 [0.009]	0.024 [0.018]	-0.006 [0.018]	0.035* [0.017]
Parental Investment x Female	-0.000 [0.009]	-0.031 [0.022]	-0.019 [0.023]	-0.046* [0.020]
Observations	26,385	20,569	26,385	26,385
Cognitive Skills				
			Math	Reading
Parental Investment	0.001 [0.002]	0.003 [0.012]	0.027* [0.011]	0.010 [0.015]
Parental Investment x Female	-0.001 [0.002]	-0.004 [0.017]	-0.014 [0.015]	-0.003 [0.019]
Observations	26,324	20,529	26,324	26,307

** p<0.01, * p<0.05, + p<0.1. Robust standard errors in brackets.

Table 3.B.6 Testing VA vs. Cumulative, Non-Cognitive

	Main Model	Factor Analysis	Components	
			Externalizing	Self Control
Grade 3				
Parental Investment	-0.008 [0.013]	-0.062 [0.046]	-0.029 [0.032]	-0.013 [0.037]
Parental Investment x Female	0.001 [0.014]	0.001 [0.061]	-0.047 [0.038]	-0.041 [0.047]
Joint Test <i>p</i> -value		0.000	0.000	0.000
Observations	8,421	4,544	8,106	8,106
Grade 1				
Parental Investment	0.002 [0.012]	0.006 [0.037]	0.032 [0.033]	0.071 ⁺ [0.037]
Parental Investment x Female	-0.012 [0.013]	0.013 [0.061]	-0.040 [0.039]	-0.027 [0.046]
Joint Test <i>p</i> -value		0.000	0.000	0.000
Observations	8,669	4,916	8,450	8,450
Kindergarten				
Parental Investment	0.013 [0.011]	0.026 [0.033]	-0.010 [0.031]	0.027 [0.029]
Parental Investment x Female	-0.011 [0.012]	-0.030 [0.045]	-0.007 [0.038]	-0.017 [0.035]
Observations	8,664	5,485	8,664	8,664

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets.

Table 3.B.7 Testing Alternative Methods of Dimension Reduction, Cognitive

	Main Model	Factor Analysis	Components	
			Math	Reading
Grade 3				
Parental Investment	0.003 [0.002]	0.003 [0.024]	0.034 [0.021]	0.008 [0.023]
Parental Investment x Female	-0.003 [0.002]	0.016 [0.032]	-0.017 [0.024]	-0.012 [0.028]
Joint Test <i>p</i> -value		0.000	0.000	0.000
Observations	8,405	4,539	8,091	8,093
Grade 1				
Parental Investment	-0.002 [0.003]	0.008 [0.032]	0.008 [0.022]	0.046 ⁺ [0.024]
Parental Investment x Female	-0.001 [0.003]	0.013 [0.046]	-0.019 [0.029]	-0.042 [0.033]
Joint Test <i>p</i> -value		0.002	0.000	0.006
Observations	8,631	4,903	8,416	8,400
Kindergarten				
Parental Investment	0.001 [0.003]	-0.012 [0.023]	0.051* [0.020]	-0.004 [0.021]
Parental Investment x Female	-0.002 [0.003]	0.012 [0.034]	-0.034 [0.025]	0.015 [0.028]
Observations	8,651	5,477	8,651	8,651

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Robust standard errors in brackets.