Equalizing Inputs, Enduring Gaps: Examining Changes in Levels and Correlates of Gender Gaps in Noncognitive Skills Over Time

Alex Johann*

Michigan State University

January 11, 2022

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

*Email: ajohann@msu.edu. Website: www.alexjohann.com

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.

1 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 so-

ciology (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 describes the data and measures I use, section 3 goes over my results, and section 4 concludes.

2 Data

2.1 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 have 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 struc-

[1] Results 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

ture), 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 8.1. Weighted descriptive statistics are reported for these two samples in Table 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 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 educated parents and two parent households, and a smaller proportion of teenage mothers.

[2]Described further in the next subsection.

[3]Described further in the next subsection.

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

[5]Following Bertrand and Pan (2013) I define two *biological* parents as the base group for family structure, rather than any two parent households. Results are robust to the use of two parent households as the base category instead.

2.2 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 activities[6] 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 disturbs ongoing activities." Self-control is constructed from "four items that indicate

[6]Measures are whether: read to child $\geq$ 3 times per week, child has $\geq$ 20 books, child reads $\geq$ 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.).

[7]Measures 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.

[8]Both 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.

[9]Parent-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.

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.

3 Results

3.1 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 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 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 A1. 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 2 shows and tests the differences in gender gaps between the two cohorts directly. As Table 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 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

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

column shows the p-value of a joint F-test of the null that the gender gap in each measure across 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 Scores[11] in the two cohorts in Appendix Figure A2. 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.

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

[12] While 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.

3.2 Factor Analysis

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

$$X = \Theta\Lambda' + U \tag{1}$$

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 $corr(X) - U$.[13]

Table 3B 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 3B shows estimates of Λ from equation 1 and the uniqueness column shows estimates of U from equation 1. The scores column is calculated from the factor loadings

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

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

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 3B 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 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 3A, 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^2s in Appendix Table A1. These four measures are eighth grade suspensions, grade retention, math test scores, and reading test scores[16]. 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 A1 provides evidence that much of the correlation the five noncognitive measures collectively has with these four grade 8 behavioral and cognitive measures remains when com-

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

[16]Eighth grade measures are only available for the 1998 cohort

bined 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 1 and Table 2 both display analyses for Latent Noncognitive Skill, and the results show that, as the weights that 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.

3.3 The Diminishing Influence of Predictors

While Figure 1 and Table 2 establish that gender gaps across most noncognitive skills have remained, Table 4 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 4. This first column of predicted gender gaps closely mirrors the findings in Table 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 4 display Oaxaca-Blinder decompositions for latent

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

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 and 3 below. In equation 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 3 shows how I can decompose the gender gap in fifth grade latent noncognitive skill using equation 2. I have shown both possible Oaxaca-Blinder specifications in panels A and B.

$$y_i = \beta_0 + \beta_f f_i + \beta_G(f_i \times X_i) + \beta_B((1 - f_i) \times X_i) + u_i \tag{2}$$

Let $\overline{X}_G = \overline{f_i \times X_i}$ and let $\overline{X}_B = \overline{(1 - f_i) \times X_i}$. Equation 3 shows the specification in panel A, which is derived by adding and subtracting $\beta_G \overline{X}_B$ and factoring out.[18]

$$E(y|j = G) - E(y|j = B) = \underbrace{\beta_f + [\beta_G - \beta_B]\overline{X}_B}_{\text{Unexplained}} + \underbrace{\beta_G[\overline{X}_G - \overline{X}_B]}_{\text{Role of X's}} \tag{3}$$

This Oaxaca-Blinder decomposition in Table 4 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.

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

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

instead of Equation 3.

3.4 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 5 and results for parental inputs in Table 6.

Table 5 looks at different levels of family background characteristics by gender and cohort. While gender[19] is randomly assigned at conception, but 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 5 I show 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 5 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 4.

Table 6 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 A2 reproduces Table 6 for all components of the two indices. Like Bertrand and Pan (2013) and

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

[20] Parental 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.

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 dviations 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 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.

3.5 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 4 in Appendix Table A3, holding constant race and kindergarten school locale across gender and cohort. I estimate Appendix Table A3 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 2A. Figure 2A 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 7A tests whether the gender gap for each quintile has changed significantly individually or jointly in fall-kindergarten, third grade, and fifth grade. As Figure 2A and Table 7A 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 2B. Like Figure 2A, I produce Figure 2B 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 7B 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 2A and Table 7A: 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.

4 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.

5 References

References

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2019), "Family disadvantage and the gender gap in behavioral and educational outcomes." *American Economic Journal: Applied Economics*, 11, 338–81.

Autor, David H, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2020), "Males at the tails: How socioeconomic status shapes the gender gap." *NBER Working Paper*.

Baker, Michael and Kevin Milligan (2016), "Boy-girl differences in parental time investments: Evidence from three countries." *Journal of Human Capital*, 10, 399–441.

Becker, Gary S, William HJ Hubbard, and Kevin M Murphy (2010), "The market for college graduates and the worldwide boom in higher education of women." *American Economic Review*, 100, 229–33.

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

Bibler, Andrew (2018), "Household composition and gender differences in parental time investments." *Available at SSRN 3192649*.

Chetty, Raj, Nathaniel Hendren, Frina Lin, Jeremy Majerovitz, and Benjamin Scuderi (2016), "Childhood environment and gender gaps in adulthood." *American Economic Review*, 106, 282–88.

Cornwell, Christopher, David B Mustard, and Jessica Van Parys (2013), "Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school." *Journal of Human resources*, 48, 236–264.

Demaray, Michelle K, Stacey L Ruffalo, John Carlson, RT Busse, Amy E Olson, Susan M McManus, and Amy Leventhal (1995), "Social skills assessment: A comparative evaluation of six published rating scales." *School Psychology Review*, 24, 648–671.

Deming, David J (2017), "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics*, 132, 1593–1640.

DiPrete, Thomas A and Jennifer L Jennings (2012), "Social and behavioral skills and the gender gap in early educational achievement." *Social Science Research*, 41, 1–15.

Duncan, Greg J, Chantelle J Dowsett, Amy Claessens, Katherine Magnuson, Aletha C Huston, Pamela Klebanov, Linda S Pagani, Leon Feinstein, Mimi Engel, Jeanne Brooks-Gunn, et al. (2007), "School readiness and later achievement." *Developmental psychology*, 43, 1428.

Elliott, Stephen N, Frank M Gresham, Terry Freeman, and George McCloskey (1988), "Teacher and observer ratings of children's social skills: Validation of the social skills rating scales." *Journal of psychoeducational Assessment*, 6, 152–161.

Fortin, Nicole M, Philip Oreopoulos, and Shelley Phipps (2015), "Leaving boys behind gender disparities in high academic achievement." *Journal of Human Resources*, 50, 549–579.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko (2006), "The homecoming of american college women: The reversal of the college gender gap." *Journal of Economic perspectives*, 20, 133–156.

Heckman, James J, Jora Stixrud, and Sergio Urzua (2006), "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics*, 24, 411–482.

Jacob, Brian 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, 589–598.

Lindqvist, Erik and Roine Vestman (2011), "The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment." *American Economic Journal: Applied Economics*, 3, 101–28.

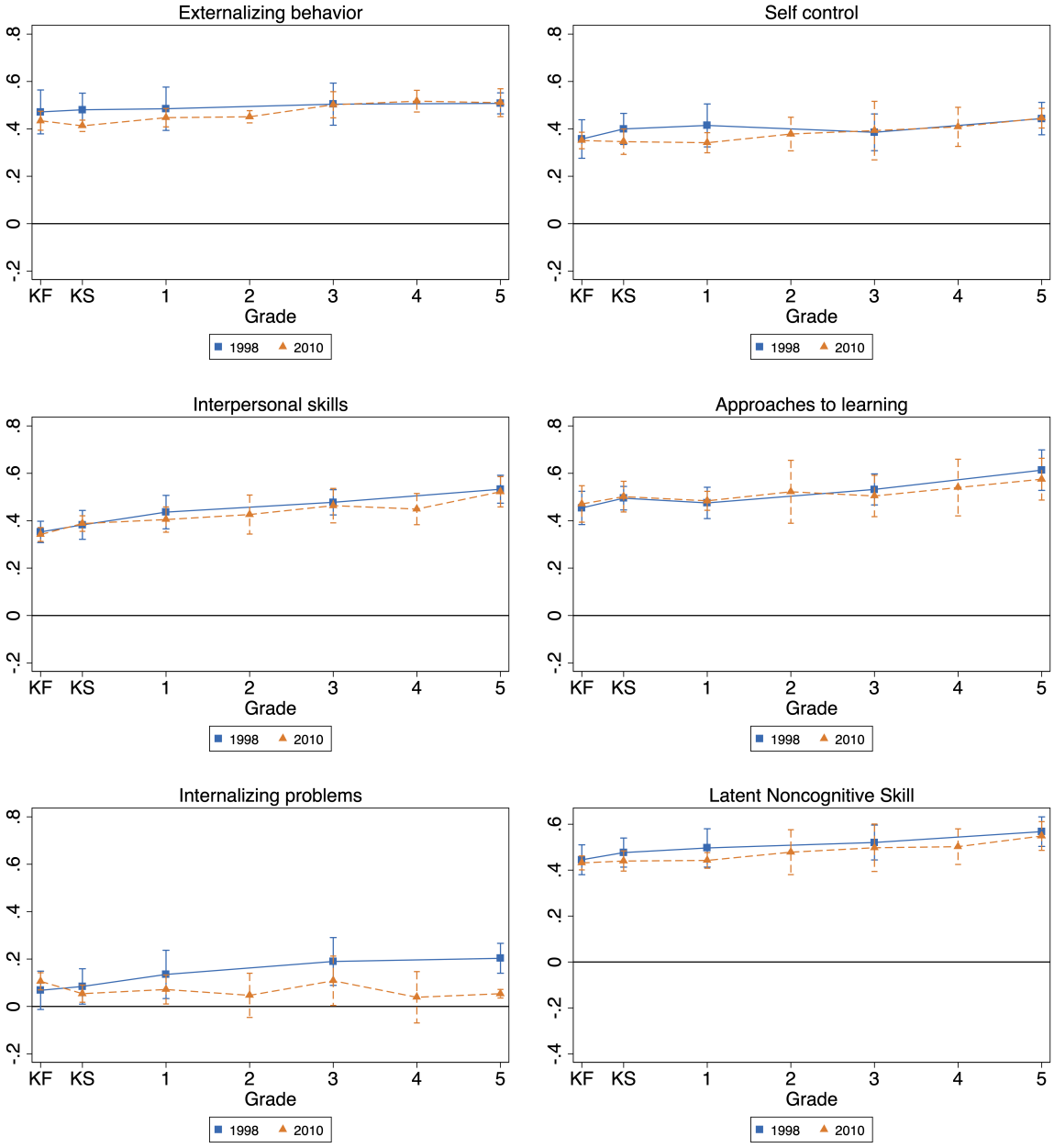
Neidell, Matthew and Jane Waldfogel (2010), "Cognitive and noncognitive peer effects in early education." *The Review of Economics and Statistics*, 92, 562–576.

Raver, CC, Pamela W Garner, and Radiah 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*, 121–147, Paul H Brookes Publishing.

Tourangeau, Karen, John Burke, Thanh Le, Siu Wan, Margaret Weant, Eugene Brown, Nancy Vaden-Kiernan, Elizabeth Rinker, Richard Dulaney, Kirsten Ellingsen, Brandon Barrett, Ismael Flores-Cervantes, Nicholas Zill, Judith Pollack, Donald Rock, Sally Atkins-Burnett, and Samuel Meisels (2001), *ECLS-K Base Year Public-Use Data Files and Electronic Codebook*. U.S. Department of Education, Washington, DC: National Center for Education Statistics. (NCES 2001-029).

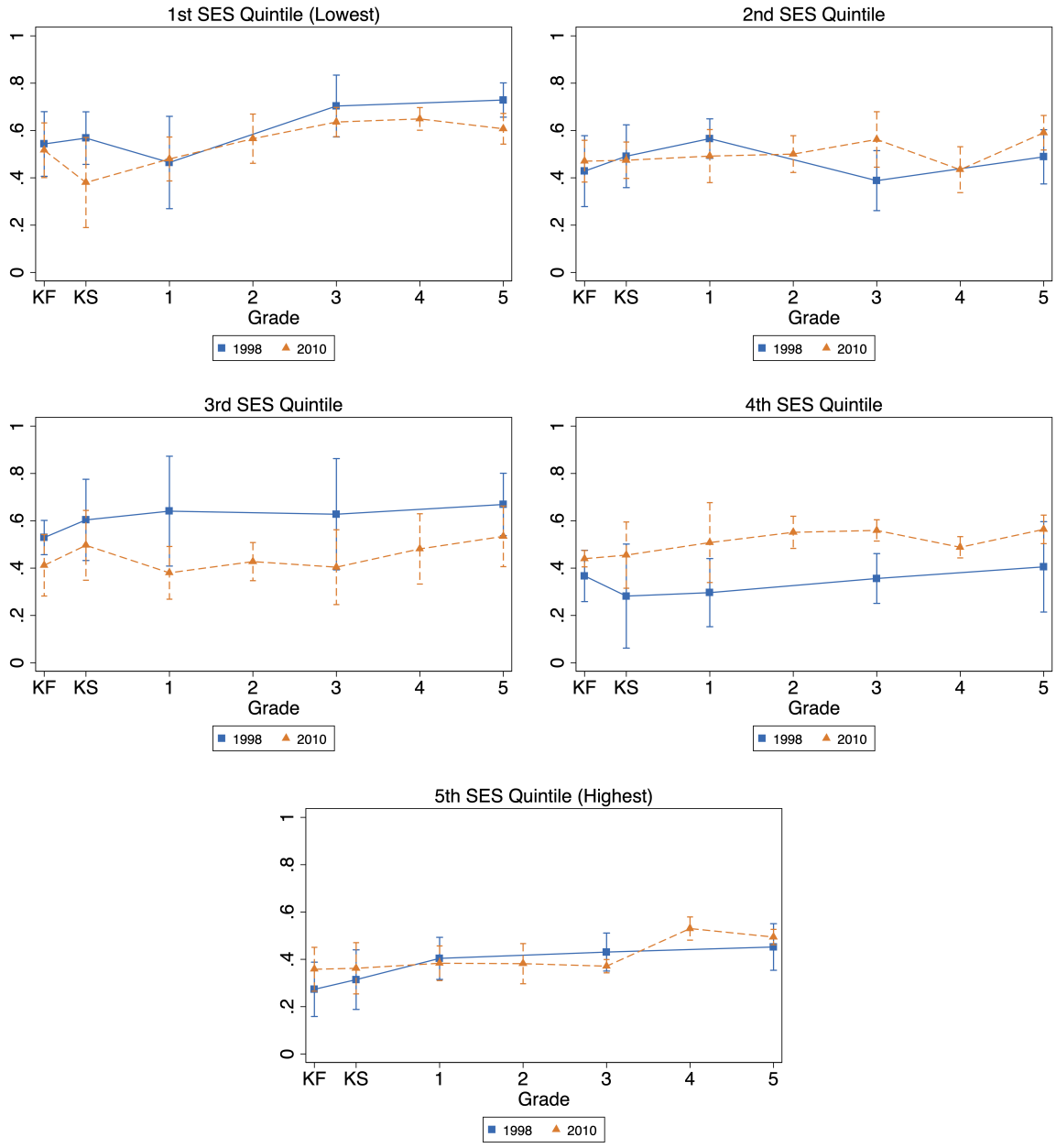
6 Figures

Figure 1: Female-Male Gaps in Teacher Ratings of Noncognitive Skills



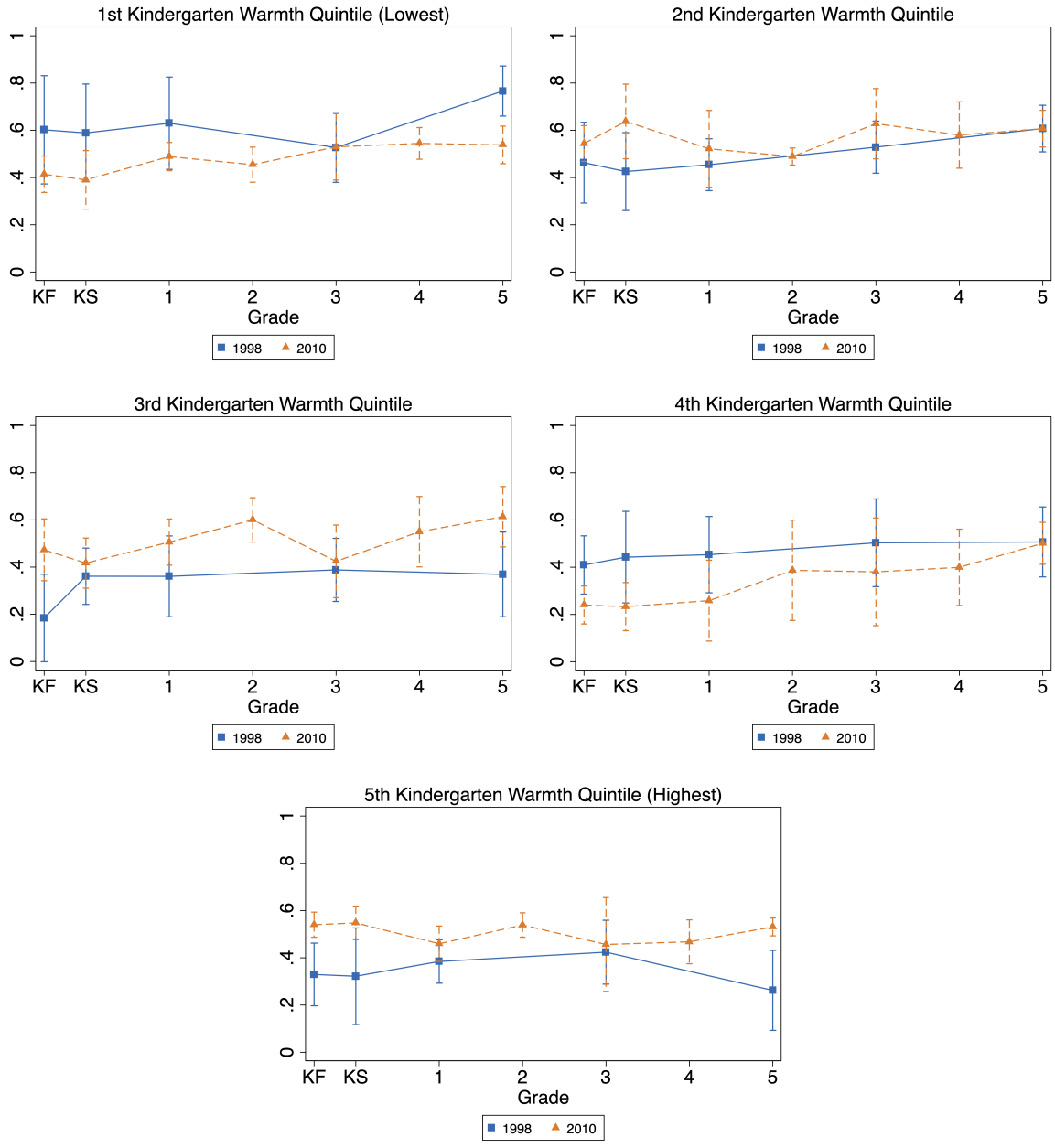
Notes: Each graph shows the coefficient on a female dummy from a regression of each respective teacher-reported noncognitive skill in each respective grade on a female dummy variable. KF refers to the fall of kindergarten, KS refers to spring of kindergarten. Teacher ratings are standardized to have 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 2A: 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 2B: 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.

7 Tables

Table 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 ≥ 20 and < 30	0.56	0.59
Age first birth ≥ 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: Changes in Female-Male Gaps in Teacher Ratings of Noncognitive Skills

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

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

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 3A: 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 3B: Factor Loadings, Scores, and Uniqueness

	Eigenvalue	Proportion Explained	
	2.879	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 4: Oaxaca-Blinder Decomposition of Fifth Grade Gender Gaps: Latent Noncognitive Skill

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

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

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 5: Kindergarten Family Background Characteristics, By Gender and Cohort

	2010 Cohort			1998 Cohort			Diff-in-Diff
Variable	Girls	Boys	Difference	Girls	Boys	Difference	
1st SES quintile (lowest)	0.144	0.141	0.002	0.151	0.172	-0.022	0.024
	[0.021]	[0.024]	[0.007]	[0.007]	[0.024]	[0.021]	[0.022]
2nd SES quintile	0.213	0.193	0.020**	0.203	0.219	-0.016	0.036**
	[0.013]	[0.008]	[0.007]	[0.013]	[0.011]	[0.011]	[0.013]
3rd SES quintile	0.223	0.237	-0.013+	0.199	0.209	-0.010	-0.003
	[0.009]	[0.005]	[0.007]	[0.008]	[0.022]	[0.027]	[0.028]
4th SES quintile	0.204	0.218	-0.013**	0.227	0.190	0.036**	-0.050**
	[0.016]	[0.014]	[0.005]	[0.009]	[0.008]	[0.013]	[0.014]
5th SES quintile (highest)	0.215	0.212	0.004	0.221	0.210	0.011	-0.007
	[0.009]	[0.014]	[0.008]	[0.013]	[0.011]	[0.012]	[0.014]
F-test jointly zero *p*-value			0.000			0.008	0.000
Age first birth < 20	0.219	0.210	0.009	0.239	0.253	-0.014	0.023
	[0.018]	[0.012]	[0.009]	[0.020]	[0.012]	[0.017]	[0.019]
More than 20 years old	0.781	0.790	-0.009	0.761	0.747	0.014	-0.023
	[0.018]	[0.012]	[0.009]	[0.020]	[0.012]	[0.017]	[0.019]
F-test jointly zero *p*-value			0.308			0.429	0.241
Single mom	0.181	0.187	-0.005	0.193	0.200	-0.007	0.001
	[0.005]	[0.013]	[0.009]	[0.011]	[0.009]	[0.008]	[0.012]
Both biological parents	0.728	0.729	-0.001	0.697	0.689	0.007	-0.008
	[0.005]	[0.016]	[0.018]	[0.011]	[0.012]	[0.012]	[0.021]
Other family structure	0.090	0.084	0.006	0.110	0.111	-0.001	0.007
	[0.007]	[0.006]	[0.010]	[0.007]	[0.011]	[0.008]	[0.013]
F-test jointly zero *p*-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 6: Kindergarten Parental Inputs, By Gender and Cohort

	2010 Cohort			1998 Cohort			Diff-in-Diff
Variable	Girls	Boys	Difference	Girls	Boys	Difference	
Kindergarten HOME index	0.034	-0.033	0.067+	0.082	-0.082	0.163**	-0.096+
	[0.026]	[0.025]	[0.036]	[0.030]	[0.031]	[0.043]	[0.056]
Kindergarten Warmth index	-0.002	0.002	-0.004	0.058	-0.058	0.116**	-0.120*
	[0.025]	[0.024]	[0.035]	[0.027]	[0.035]	[0.044]	[0.056]
Spanked child last week, kindergarten	0.149	0.169	-0.020	0.260	0.276	-0.015	-0.004
	[0.009]	[0.009]	[0.013]	[0.013]	[0.014]	[0.019]	[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 7A: 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	-0.060	-0.135**
	[0.086]	[0.074]	[0.049]
2nd SES quintile	0.050	0.169*	0.122+
	[0.088]	[0.081]	[0.073]
3rd SES quintile	-0.116	-0.232+	-0.144
	[0.079]	[0.139]	[0.094]
4th SES quintile	0.084	0.204**	0.166
	[0.056]	[0.053]	[0.105]
5th SES quintile (highest)	0.099	-0.048	0.060
	[0.077]	[0.046]	[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 7B: 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	-0.012	-0.232**
	[0.130]	[0.101]	[0.077]
2nd Kindergarten Warmth quintile	0.084	0.092	-0.001
	[0.105]	[0.095]	[0.059]
3rd Kindergarten Warmth quintile	0.285*	0.042	0.254*
	[0.113]	[0.106]	[0.119]
4th Kindergarten Warmth quintile	-0.158*	-0.120	-0.000
	[0.071]	[0.150]	[0.086]
5th Kindergarten Warmth quintile (highest)	0.178*	0.003	0.265**
	[0.086]	[0.125]	[0.085]
Joint F-test of no change *p*-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.

8 Online Appendix

8.1 Analysis Sample Creation

The full K-5 longitudinal panel sample described in Section 2.1 is defined as observations with non-missing and non-zero panel weights.[21] 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.

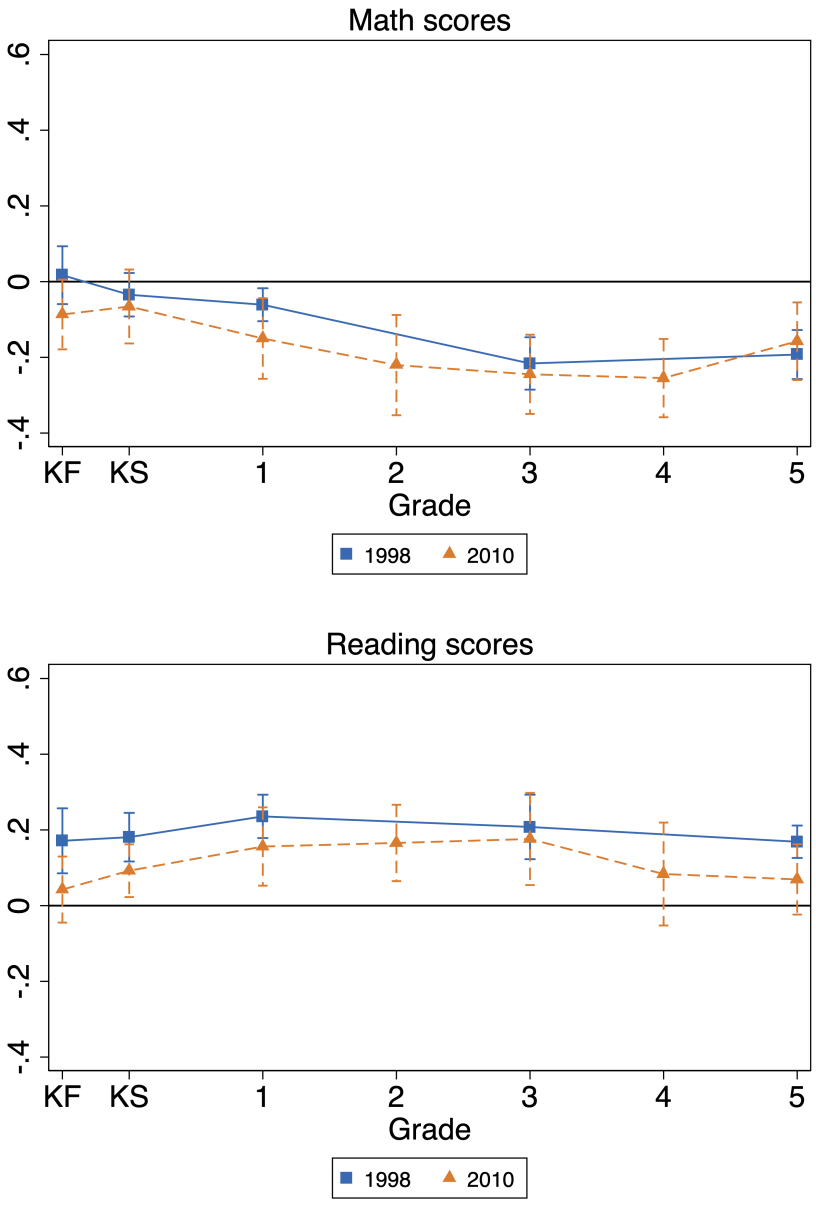
To investigate the possibility of nonrandom item nonresponse reflected by the first and third step, I reran my analysis using weights adjusted by the inverse probability of appearing in the final analysis sample among all longitudinal panel observations. These weights are calculated using a logit regression 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

[21] The 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.

probability weights. All results were then rerun with these alternative weights, with no notable differences. Tables and figures created using these weights are available upon request.

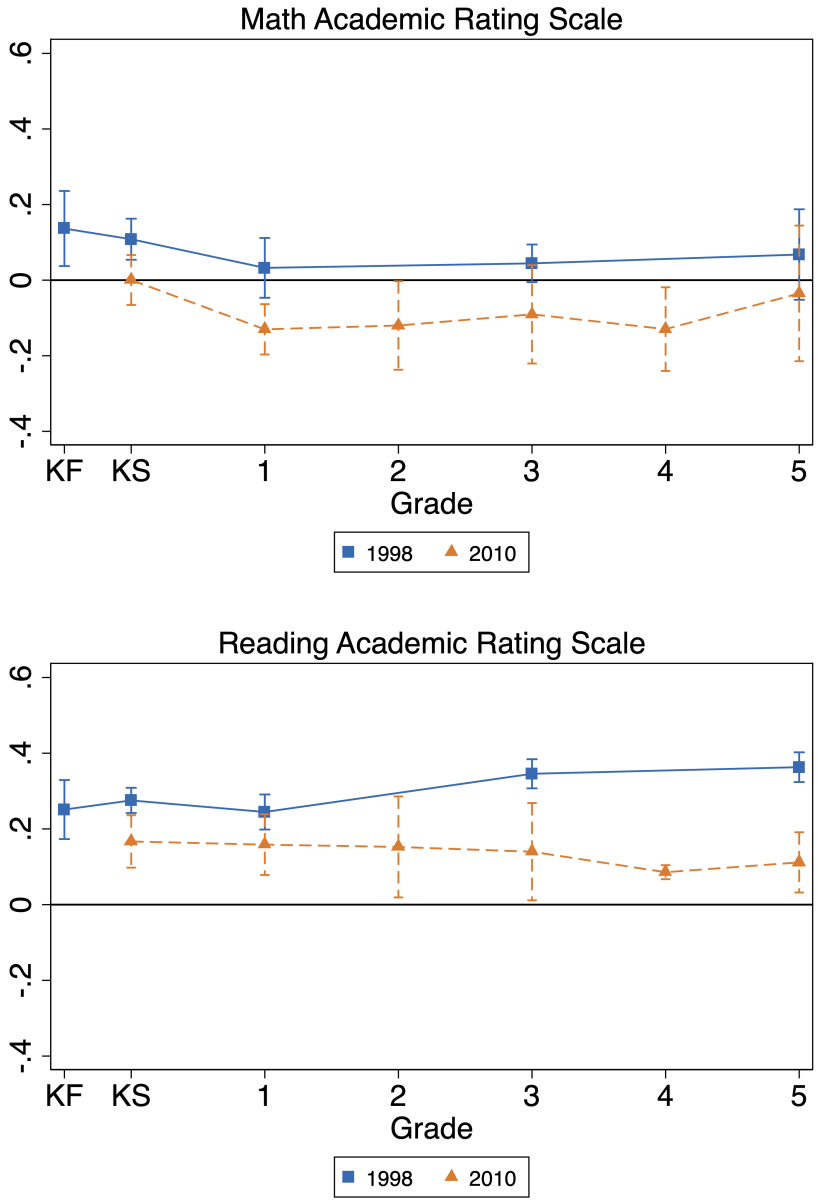
8.2 Appendix Figures

Figure A1: Female-Male Gaps in Test Scores



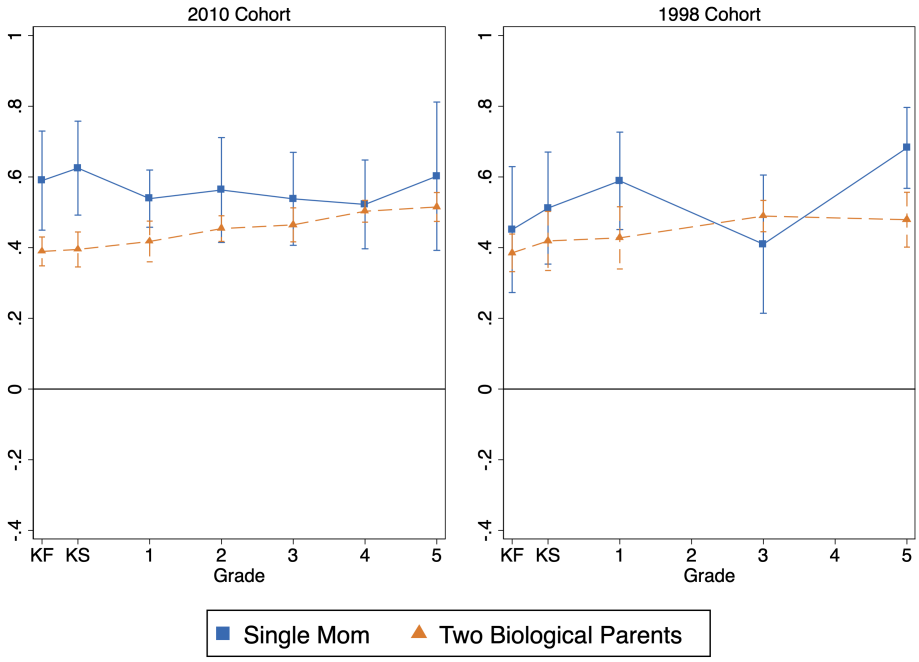
Notes: Each graph shows the coefficient on a female dummy from a regression of each test score in each respective grade on a female dummy variable. KF refers to the fall of kindergarten, KS, refers to spring of kindergarten. Please refer to the text for sample restrictions. Observations are weighted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

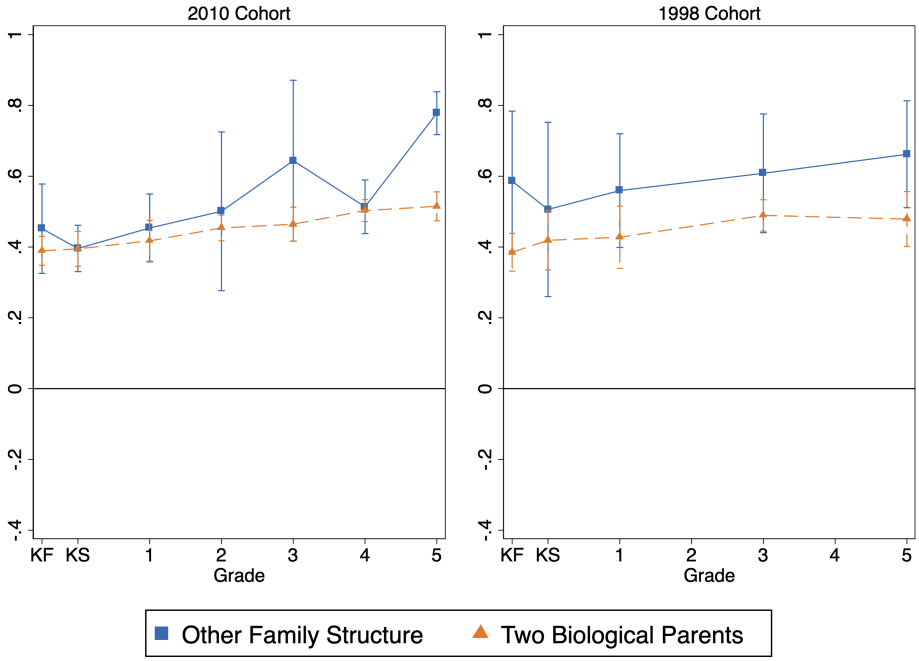
Figure A2: Female-Male Gaps in Teacher Cognitive Evalutions



Notes: Each graph shows the coefficient on a female dummy from a regression of each respective teacher-reported rating of cogntive 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.

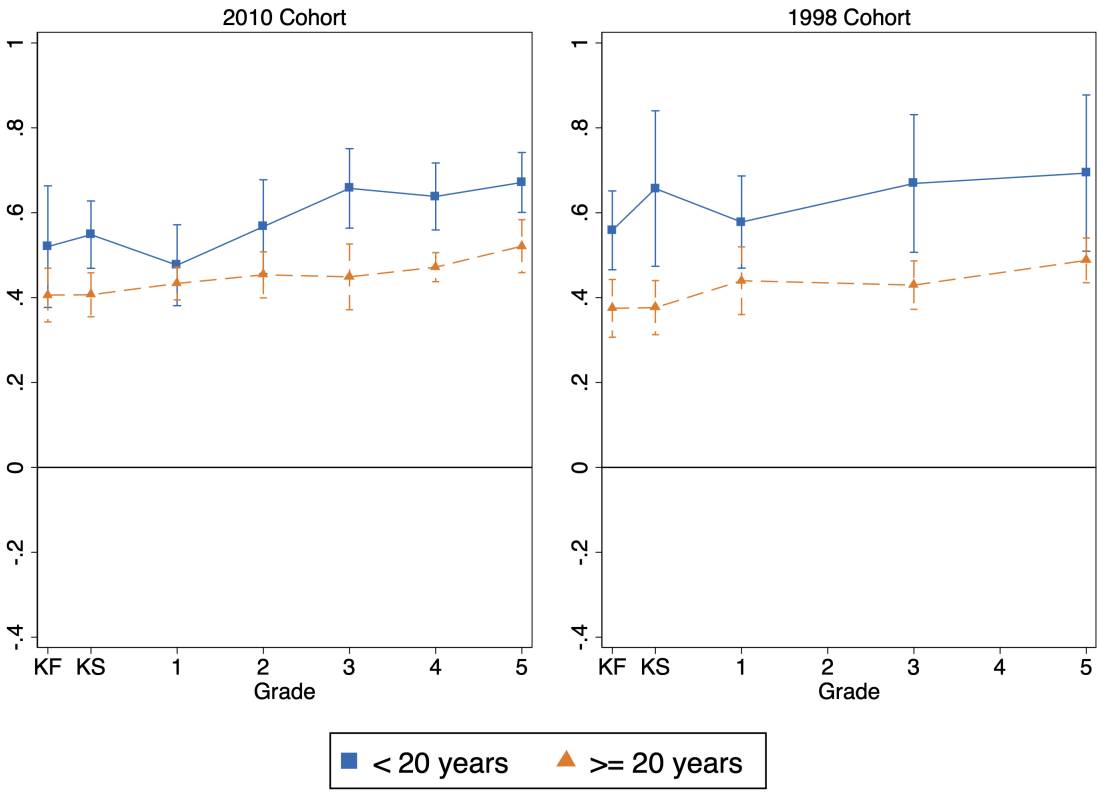
Figure A3A: Female-Male Gaps in Latent Noncognitive Skill, By Kindergarten Family Structure





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 A3B: 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.

8.3 Appendix Tables

Table A1: Adjusted R^2s from Regressions of Outcomes on Latent Noncognitive Skill or Components

	Latent Noncognitive Skill in 5th Grade	All Noncognitive Measures in 5th Grade
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 A2: Kindergarten Parental Inputs, By Gender

Variable	2010 Cohort Girls	2010 Cohort Boys	2010 Cohort Difference	1998 Cohort Girls	1998 Cohort Boys	1998 Cohort Difference	
Kindergarten HOME index	0.034	-0.033	0.067+	0.082	-0.082	0.163**	-0.096+
	[0.026]	[0.025]	[0.036]	[0.030]	[0.031]	[0.043]	[0.056]
Read book to child 3+ times per week	0.883	0.868	0.015	0.837	0.794	0.043*	-0.028
	[0.008]	[0.008]	[0.012]	[0.011]	[0.013]	[0.017]	[0.021]
Child has $\geq$20 books around house	0.874	0.866	0.008	0.887	0.857	0.030*	-0.023
	[0.009]	[0.008]	[0.012]	[0.009]	[0.011]	[0.015]	[0.019]
Visited the library	0.618	0.571	0.047**	0.572	0.550	0.021	0.026
	[0.012]	[0.012]	[0.017]	[0.015]	[0.015]	[0.021]	[0.027]
Gone to a play/concert/show	0.426	0.410	0.016	0.422	0.346	0.076**	-0.060*
	[0.012]	[0.012]	[0.017]	[0.015]	[0.014]	[0.021]	[0.027]
Visited art/musuem/historical site	0.338	0.344	-0.006	0.308	0.297	0.011	-0.017
	[0.012]	[0.011]	[0.016]	[0.014]	[0.014]	[0.020]	[0.025]
Child reads outside school 3+ times per week	0.249	0.009	0.241**	0.139	-0.309	0.448**	-0.207**
	[0.021]	[0.024]	[0.032]	[0.026]	[0.034]	[0.042]	[0.053]
Have home computer child uses	0.761	0.764	-0.003	0.602	0.563	0.038+	-0.041
	[0.011]	[0.011]	[0.015]	[0.015]	[0.015]	[0.021]	[0.026]
Child engages in other outside school activity	0.753	0.740	0.012	0.683	0.629	0.054*	-0.041
	[0.011]	[0.011]	[0.016]	[0.014]	[0.015]	[0.021]	[0.026]
Kindergarten Warmth index	-0.002	0.002	-0.004	0.058	-0.058	0.116**	-0.120*
	[0.025]	[0.024]	[0.035]	[0.027]	[0.035]	[0.044]	[0.056]
Warm, close times together	0.962	0.950	0.012+	0.956	0.947	0.009	0.002
	[0.005]	[0.005]	[0.007]	[0.007]	[0.008]	[0.011]	[0.013]
Child likes me	0.981	0.970	0.011*	0.979	0.968	0.012	-0.001
	[0.003]	[0.004]	[0.005]	[0.004]	[0.007]	[0.008]	[0.009]
Always show child love	0.931	0.934	-0.004	0.870	0.862	0.008	-0.011
	[0.006]	[0.006]	[0.008]	[0.010]	[0.010]	[0.015]	[0.017]
Express affection	0.992	0.985	0.007+	0.986	0.976	0.010*	-0.004
	[0.002]	[0.003]	[0.004]	[0.002]	[0.004]	[0.005]	[0.006]
Being parent harder than I thought (reverse)	0.413	0.420	-0.007	0.522	0.513	0.009	-0.015
	[0.012]	[0.012]	[0.017]	[0.015]	[0.015]	[0.021]	[0.027]
Child does things that bother me (reverse)	0.916	0.902	0.014	0.912	0.882	0.030*	-0.016
	[0.006]	[0.007]	[0.010]	[0.008]	[0.011]	[0.014]	[0.017]
Sacrifice to meet child's needs (reverse)	0.715	0.751	-0.036*	0.767	0.754	0.013	-0.050*
	[0.012]	[0.010]	[0.016]	[0.013]	[0.013]	[0.018]	[0.024]
Often feel angry with child (reverse)	0.988	0.985	0.003	0.984	0.985	-0.002	0.004
	[0.002]	[0.003]	[0.004]	[0.004]	[0.004]	[0.005]	[0.007]
Spanked child last week, kindergarten	0.149	0.169	-0.020	0.260	0.276	-0.015	-0.004
	[0.009]	[0.009]	[0.013]	[0.013]	[0.014]	[0.019]	[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 A3: 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*	0.016	0.023	0.038+	0.023*	0.014	0.009
	[0.015]	[0.011]	[0.021]	[0.023]	[0.011]	[0.029]	[0.036]
Lower kindergarten Warmth index	-0.040+	-0.048	0.008	-0.022*	-0.166**	0.145**	-0.137**
	[0.022]	[0.034]	[0.014]	[0.011]	[0.030]	[0.029]	[0.032]
Spanked child last week, kindergarten	-0.178**	-0.192**	0.014	-0.083	-0.154**	0.072	-0.057
	[0.035]	[0.011]	[0.039]	[0.056]	[0.046]	[0.051]	[0.064]
Single mom	-0.279**	-0.343**	0.064	-0.156**	-0.286**	0.130	-0.066
	[0.103]	[0.030]	[0.112]	[0.035]	[0.068]	[0.081]	[0.136]
Other family structure	-0.314**	-0.547**	0.233**	-0.320**	-0.428**	0.108	0.125
	[0.057]	[0.047]	[0.039]	[0.062]	[0.083]	[0.083]	[0.092]
Age first birth < 20	0.002	-0.110*	0.112	-0.214*	-0.353**	0.138	-0.027
	[0.039]	[0.044]	[0.080]	[0.095]	[0.048]	[0.123]	[0.147]
1st SES quintile (lowest)	-0.397**	-0.402**	0.005	-0.262*	-0.400**	0.138	-0.134
	[0.065]	[0.063]	[0.041]	[0.127]	[0.046]	[0.122]	[0.127]
2nd SES quintile	-0.330**	-0.333**	0.003	-0.325**	-0.269**	-0.056	0.059
	[0.055]	[0.068]	[0.064]	[0.094]	[0.047]	[0.096]	[0.113]
3rd SES quintile	-0.261**	-0.255**	-0.005	-0.090+	-0.278**	0.188**	-0.193*
	[0.086]	[0.059]	[0.059]	[0.048]	[0.031]	[0.051]	[0.079]
4th SES quintile	-0.099**	-0.147*	0.047	-0.129+	-0.074	-0.054	0.102
	[0.020]	[0.056]	[0.043]	[0.066]	[0.046]	[0.092]	[0.102]
SES F-test of jointly zero, *p*-value			0.345			0.000	0.000

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

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 A4A: Changes in Gender Gaps Between Cohorts
The Role of Family Structure at Kindergarten

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

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 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 A4B: 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	-0.003	-0.016
	[0.087]	[0.093]	[0.106]
More than 20 years old	0.039	0.019	0.040
	[0.052]	[0.045]	[0.042]
Joint F-test of no change *p*-value	0.649	0.842	0.636

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 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.

8.4 Results for Two Orthogonal Latent Factors

8.4.1 Introduction

As described in Section 3.2 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 3B 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 3B, 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 externaliz-

ing behavior. Comparing to the factor scores in Table 3B, 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 A1), which shows the R^2s 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.

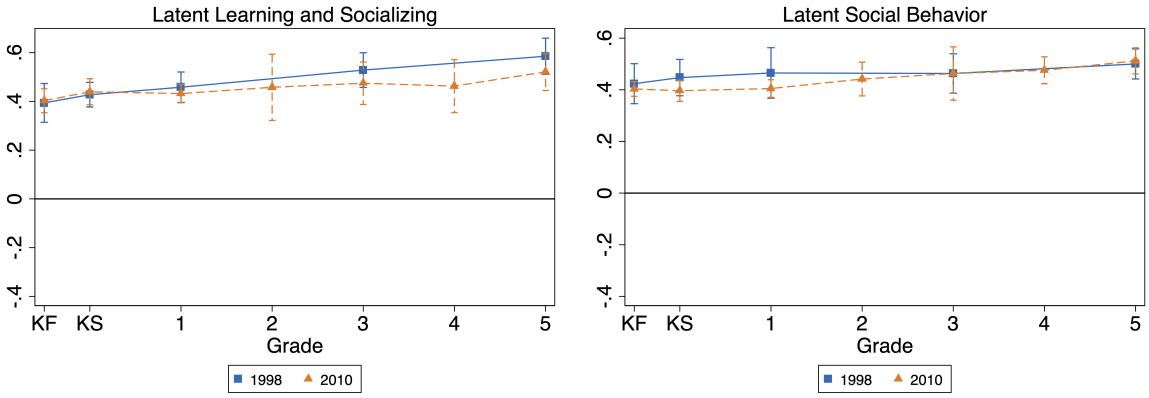
Table B1: Factor Loadings, Scores, and Uniqueness

	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.

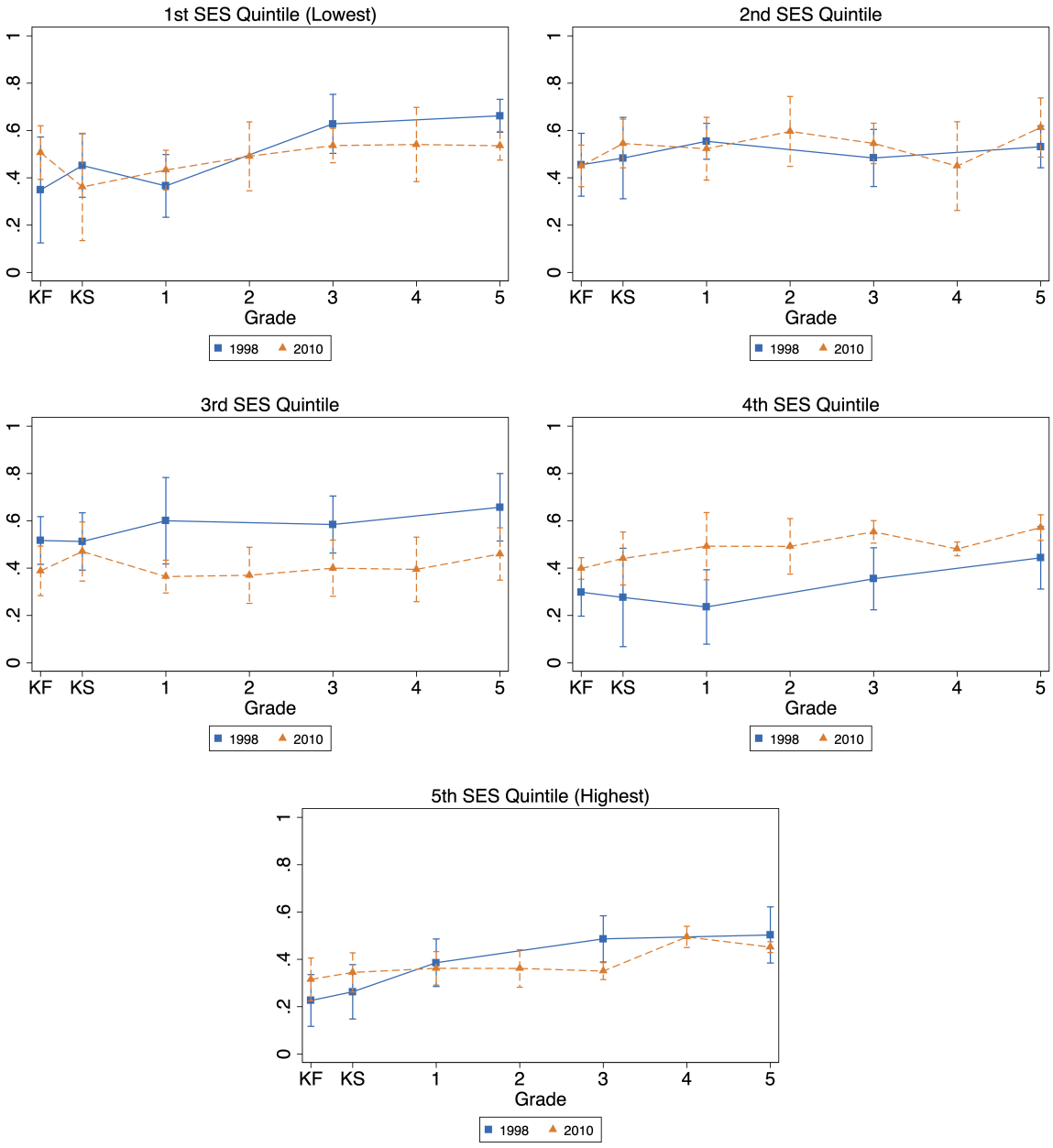
8.4.2 Figures

Figure B1: Female-Male Gaps in Teacher Ratings of Noncognitive Skills



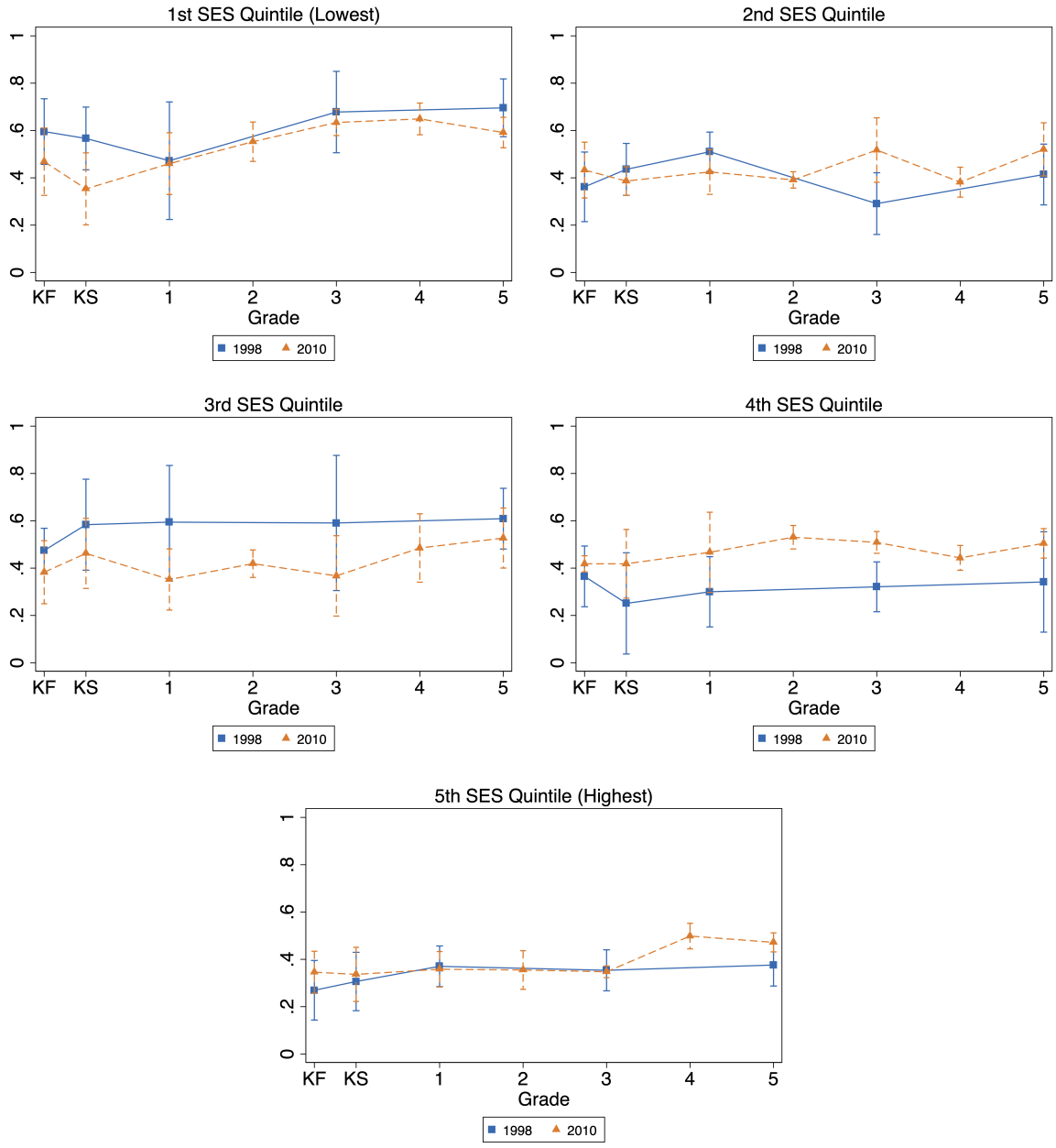
Notes: Each graph shows the coefficient on a female dummy from a regression of each respective teacher-reported noncognitive skill in each respective grade on a female dummy variable. KF refers to the fall of kindergarten, KS refers to spring of kindergarten. Teacher ratings are standardized to have 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 B2A.A: Female-Male Gaps in Latent Learning and Socializing, By SES at Kindergarten



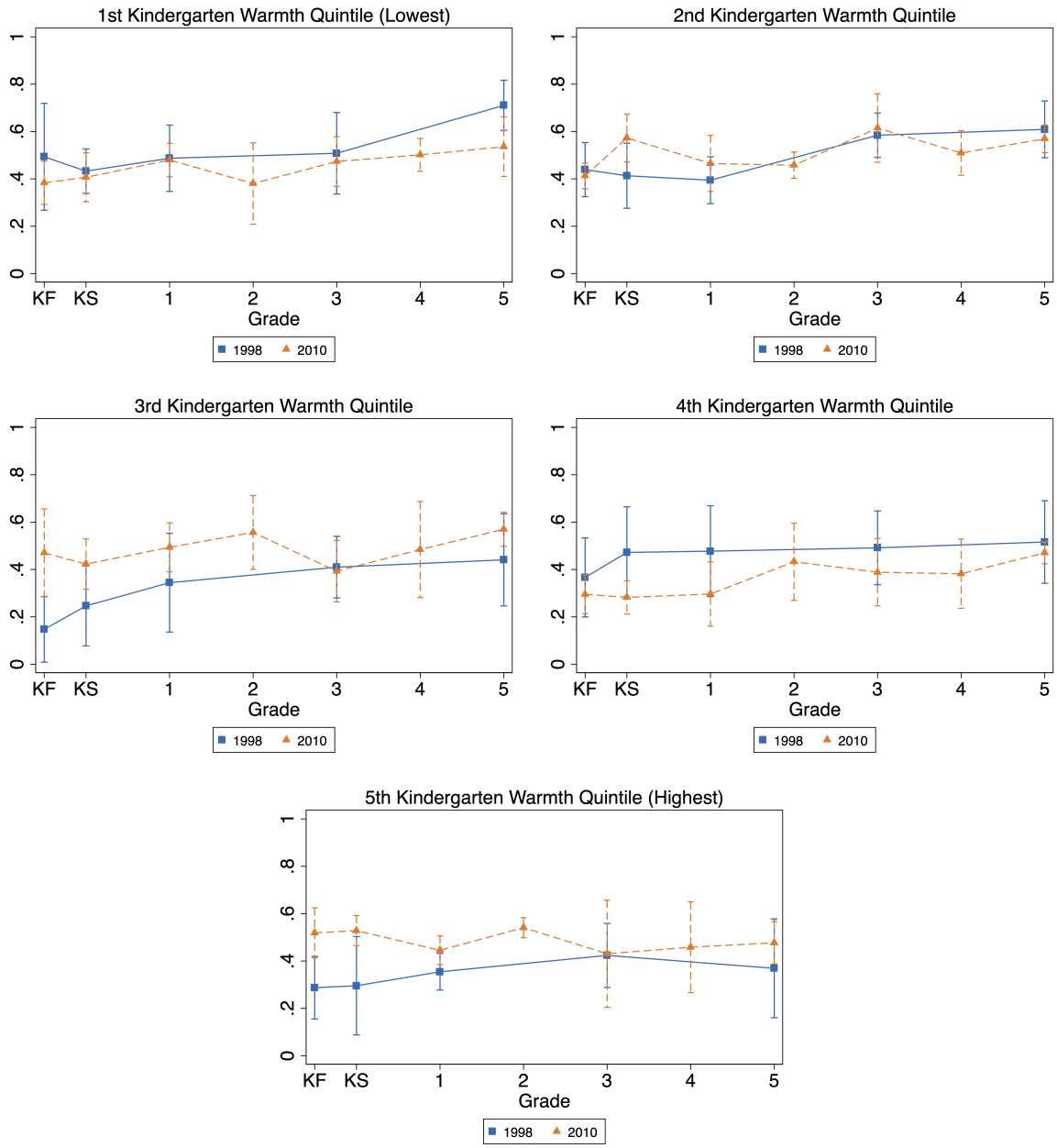
Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by SES quintile interaction term, and a female by 2010 cohort by SES quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female 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 B2A.B: Female-Male Gaps in Latent Social Behavior, By SES at Kindergarten



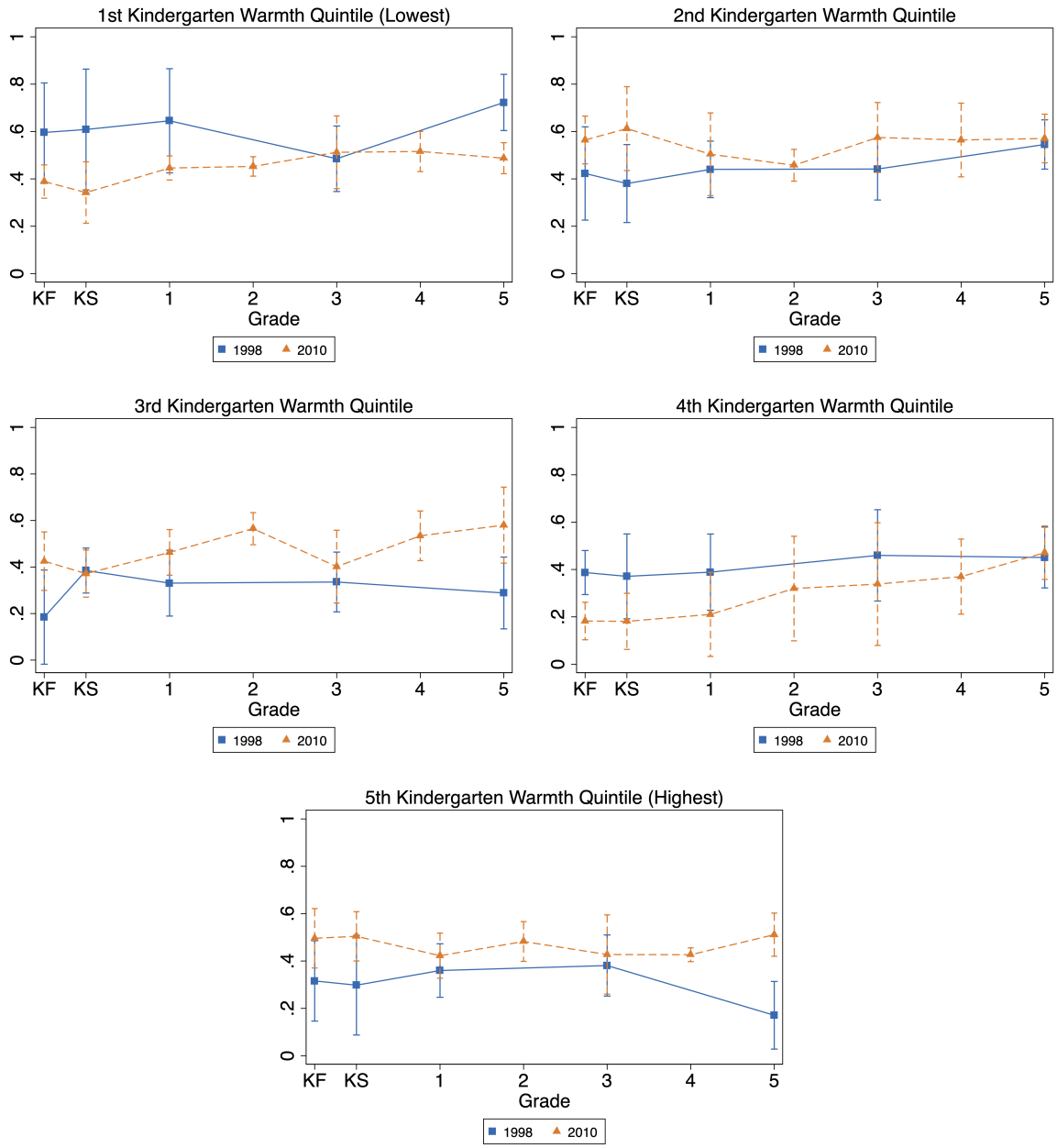
Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by SES quintile interaction term, and a female by 2010 cohort by SES quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female dummy and a female 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 B2B.A: Female-Male Gaps in Latent Learning and Socializing, By Kindergarten Warmth Index



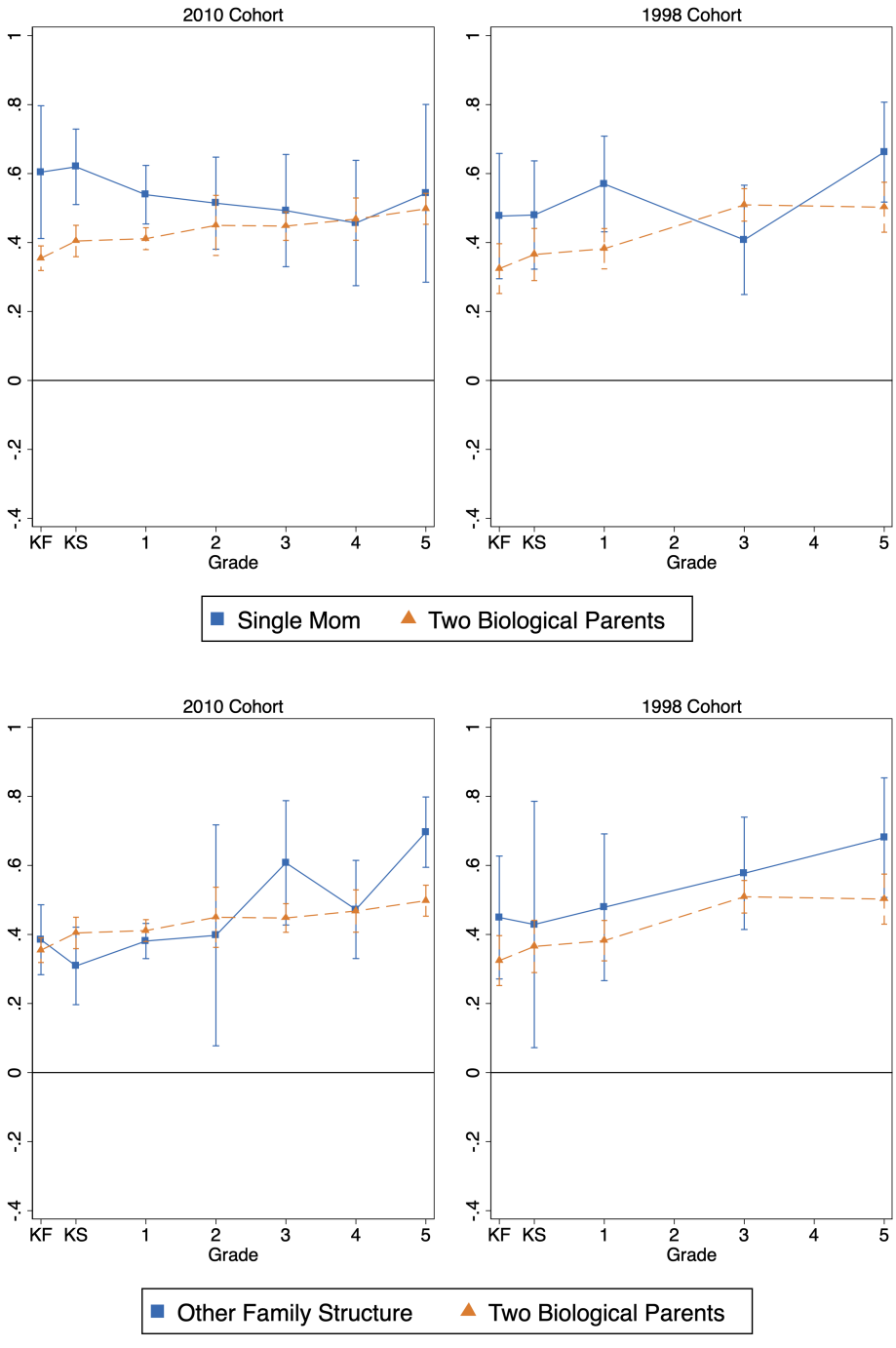
Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by Warmth index quintile interaction term, and a female by 2010 cohort by Warmth index quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female 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 B2B.B: Female-Male Gaps in Latent Social Behavior, By Kindergarten Warmth Index



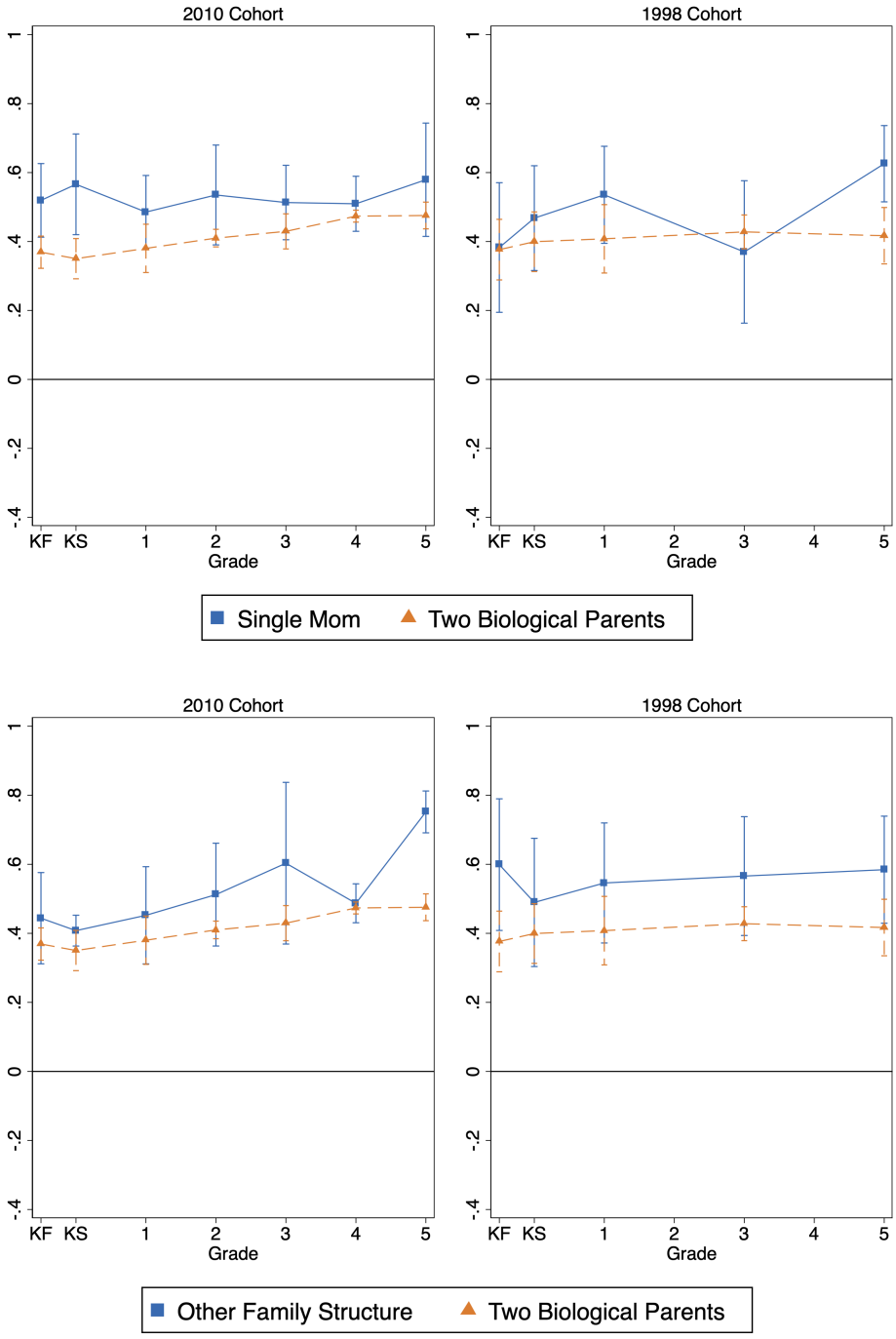
Notes: Each graph shows the sum of the coefficients on a female dummy, a female by 2010 cohort interaction term, a female by Warmth index quintile interaction term, and a female by 2010 cohort by Warmth index quintile interaction term (for the 2010 estimates) as well sum of the coefficients on a female 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 B3A.A: Female-Male Gaps in Latent Learning and Socializing, By Kindergarten Family Structure



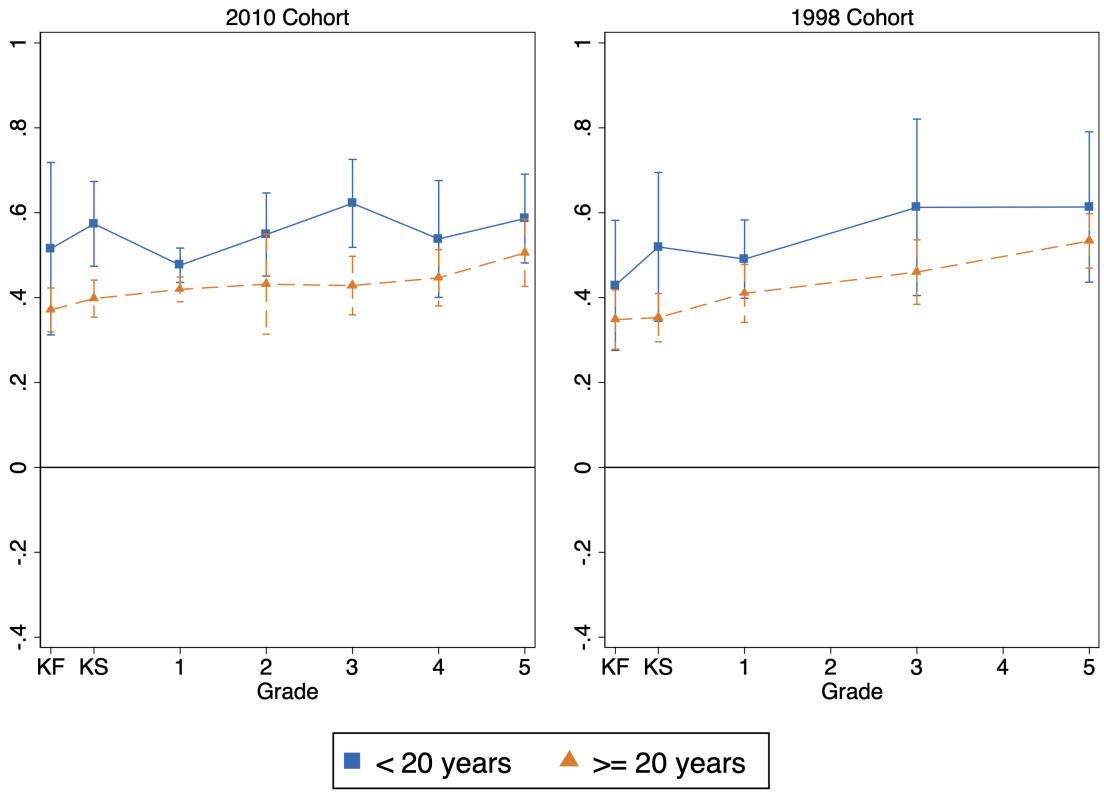
Notes: Estimates for each grade in all four graphs come from one regression of latent learning and socializing in each grade with controls for teen motherhood, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. 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 B3A.B: Female-Male Gaps in Latent Social Behavior, By Kindergarten Family Structure



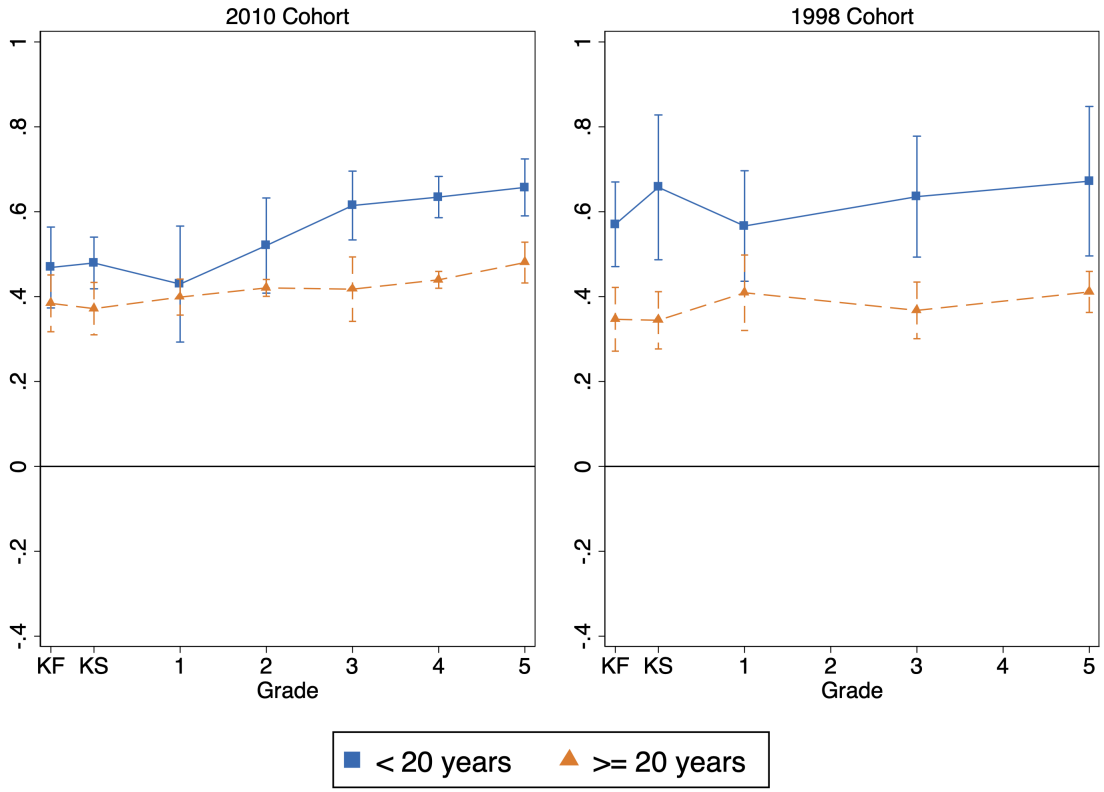
Notes: Estimates for each grade in all four graphs come from one regression of Latent Social Behavior in each grade with controls for teen motherhood, SES at kindergarten, HOME index at kindergarten, Warmth index at kindergarten, spanking at kindergarten, child race, and school locale at kindergarten. 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 B3B.A: 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 B3B.B: 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.

8.4.3 Tables

Table B2: Changes in Female-Male Gaps in Teacher Ratings of Noncognitive Skills

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

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

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 B3A: 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**	0.523**	-0.002
	[0.034]	[0.034]	[0.011]
1998	0.585**	0.560**	0.025+
	[0.038]	[0.036]	[0.014]
Difference	-0.064	-0.038	-0.026
	[0.051]	[0.049]	[0.017]
Panel B: Girls' X's, Boys' Betas			
2010	0.521**	0.526**	-0.004
	[0.036]	[0.033]	[0.012]
1998	0.585**	0.538**	0.048**
	[0.041]	[0.039]	[0.017]
Difference	-0.064	-0.012	-0.052*
	[0.055]	[0.051]	[0.021]

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

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 B3B: 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**	0.518**	-0.006
	[0.033]	[0.033]	[0.010]
1998	0.500**	0.489**	0.011
	[0.046]	[0.042]	[0.014]
Difference	0.012	0.029	-0.017
	[0.057]	[0.054]	[0.017]
Panel B: Girls' X's, Boys' Betas			
2010	0.512**	0.519**	-0.007
	[0.030]	[0.029]	[0.015]
1998	0.500**	0.456**	0.044+
	[0.044]	[0.041]	[0.023]
Difference	0.012	0.063	-0.051+
	[0.053]	[0.050]	[0.027]

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

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 B4A.A: 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	-0.090	-0.142**
	[0.115]	[0.071]	[0.049]
2nd SES quintile	0.002	0.056	0.095
	[0.081]	[0.064]	[0.082]
3rd SES quintile	-0.129	-0.189*	-0.205*
	[0.079]	[0.082]	[0.092]
4th SES quintile	0.111*	0.206**	0.135+
	[0.053]	[0.069]	[0.072]
5th SES quintile (highest)	0.102	-0.124*	-0.034
	[0.072]	[0.056]	[0.057]
Joint F-test of equality *p*-value	0.001	0.000	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 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 B4A.B: 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	-0.033	-0.114
	[0.111]	[0.090]	[0.071]
2nd SES quintile	0.079	0.222*	0.129
	[0.097]	[0.092]	[0.088]
3rd SES quintile	-0.090	-0.233	-0.090
	[0.083]	[0.164]	[0.093]
4th SES quintile	0.063	0.184**	0.171
	[0.069]	[0.057]	[0.117]
5th SES quintile (highest)	0.090	0.005	0.113*
	[0.081]	[0.042]	[0.055]
Joint F-test of equality *p*-value	0.268	0.000	0.023

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 B4B.A: 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	-0.035	-0.175*
	[0.126]	[0.102]	[0.080]
2nd Kindergarten Warmth quintile	-0.027	0.030	-0.040
	[0.064]	[0.086]	[0.068]
3rd Kindergarten Warmth quintile	0.323**	-0.017	0.129
	[0.120]	[0.096]	[0.105]
4th Kindergarten Warmth quintile	-0.071	-0.103	-0.046
	[0.094]	[0.108]	[0.092]
5th Kindergarten Warmth quintile (highest)	0.232*	0.007	0.108
	[0.089]	[0.131]	[0.113]
Joint F-test of no change *p*-value	0.084	0.850	0.024

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.

Table B4B.B: 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+	0.027	-0.235**
	[0.113]	[0.106]	[0.070]
2nd Kindergarten Warmth quintile	0.142	0.133	0.025
	[0.114]	[0.101]	[0.074]
3rd Kindergarten Warmth quintile	0.240+	0.066	0.292*
	[0.126]	[0.106]	[0.115]
4th Kindergarten Warmth quintile	-0.205**	-0.121	0.020
	[0.061]	[0.165]	[0.088]
5th Kindergarten Warmth quintile (highest)	0.180+	0.046	0.340**
	[0.106]	[0.104]	[0.090]
Joint F-test of equality *p*-value	0.038	0.651	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 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 B5: 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 B6A: Fifth Grade Latent 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	0.005	0.008	0.008	-0.029	0.036	-0.028
	[0.014]	[0.016]	[0.026]	[0.015]	[0.021]	[0.022]	[0.035]
Lower kindergarten Warmth index	-0.025	-0.043	0.018	-0.013	-0.119**	0.106**	-0.087*
	[0.022]	[0.031]	[0.017]	[0.018]	[0.022]	[0.033]	[0.037]
Spanked child last week, kindergarten	-0.213**	-0.185**	-0.028	-0.107	-0.112**	0.005	-0.033
	[0.036]	[0.013]	[0.030]	[0.066]	[0.031]	[0.059]	[0.066]
Single mom	-0.281**	-0.319**	0.038	-0.178**	-0.291**	0.113	-0.075
	[0.095]	[0.043]	[0.128]	[0.039]	[0.077]	[0.092]	[0.156]
Other family structure	-0.335**	-0.512**	0.178**	-0.235**	-0.377**	0.142	0.036
	[0.056]	[0.034]	[0.060]	[0.062]	[0.056]	[0.087]	[0.104]
Age first birth < 20	-0.012	-0.055	0.043	-0.195**	-0.215**	0.020	0.022
	[0.050]	[0.058]	[0.106]	[0.068]	[0.041]	[0.095]	[0.144]
1st SES quintile (lowest)	-0.442**	-0.478**	0.036	-0.354**	-0.417**	0.063	-0.027
	[0.037]	[0.054]	[0.064]	[0.094]	[0.064]	[0.072]	[0.097]
2nd SES quintile	-0.345**	-0.463**	0.118	-0.332**	-0.295**	-0.037	0.154
	[0.071]	[0.056]	[0.090]	[0.084]	[0.046]	[0.075]	[0.117]
3rd SES quintile	-0.293**	-0.282**	-0.011	-0.150**	-0.287**	0.137**	-0.147*
	[0.035]	[0.053]	[0.058]	[0.043]	[0.034]	[0.036]	[0.070]
4th SES quintile	-0.093**	-0.205**	0.111*	-0.139*	-0.075	-0.064	0.175+
	[0.011]	[0.040]	[0.046]	[0.067]	[0.047]	[0.090]	[0.101]

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

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 B6B: Fifth Grade Latent 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**	0.021+	0.030	0.052+	0.052*	-0.001	0.031
	[0.019]	[0.012]	[0.027]	[0.028]	[0.022]	[0.047]	[0.054]
Lower kindergarten Warmth index	-0.045*	-0.046	0.001	-0.025	-0.178**	0.153**	-0.153**
	[0.021]	[0.032]	[0.012]	[0.016]	[0.033]	[0.026]	[0.028]
Spanked child last week, kindergarten	-0.140**	-0.178**	0.038	-0.060	-0.164**	0.105+	-0.066
	[0.049]	[0.014]	[0.057]	[0.050]	[0.056]	[0.058]	[0.080]
Single mom	-0.251*	-0.324**	0.073	-0.127**	-0.254**	0.128	-0.054
	[0.100]	[0.029]	[0.093]	[0.039]	[0.072]	[0.082]	[0.124]
Other family structure	-0.271**	-0.514**	0.244**	-0.339**	-0.415**	0.077	0.167
	[0.052]	[0.057]	[0.031]	[0.063]	[0.101]	[0.098]	[0.104]
Age first birth < 20	0.011	-0.132**	0.142*	-0.204+	-0.399**	0.196	-0.053
	[0.034]	[0.032]	[0.059]	[0.105]	[0.051]	[0.130]	[0.143]
1st SES quintile (lowest)	-0.332**	-0.317**	-0.015	-0.180	-0.350**	0.170	-0.184
	[0.085]	[0.067]	[0.042]	[0.136]	[0.052]	[0.154]	[0.158]
2nd SES quintile	-0.288**	-0.223**	-0.066	-0.287**	-0.225**	-0.062	-0.004
	[0.042]	[0.075]	[0.058]	[0.094]	[0.050]	[0.105]	[0.117]
3rd SES quintile	-0.216*	-0.214**	-0.002	-0.044	-0.244**	0.200**	-0.201*
	[0.108]	[0.065]	[0.060]	[0.052]	[0.037]	[0.068]	[0.091]
4th SES quintile	-0.093**	-0.098	0.005	-0.109+	-0.066	-0.044	0.048
	[0.033]	[0.065]	[0.044]	[0.063]	[0.056]	[0.095]	[0.104]

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

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 B7A.A: 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	0.098	-0.104
	[0.135]	[0.119]	[0.158]
Both biological parents	0.038	-0.059*	-0.004
	[0.043]	[0.029]	[0.044]
Other family structure	-0.076	0.030	0.022
	[0.104]	[0.122]	[0.105]
Joint F-test of equality *p*-value	0.440	0.435	0.723

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 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 B7A.B: 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	0.163	-0.033
	[0.111]	[0.127]	[0.109]
Both biological parents	0.004	0.002	0.064
	[0.060]	[0.030]	[0.048]
Other family structure	-0.168	0.025	0.172+
	[0.116]	[0.140]	[0.089]
Joint F-test of equality *p*-value	0.205	0.384	0.323

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 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 B7B.A: 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	0.014	-0.028
	[0.130]	[0.110]	[0.111]
More than 20 years old	0.030	-0.029	-0.022
	[0.047]	[0.050]	[0.051]
Joint F-test of equality *p*-value	0.598	0.766	0.961

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 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 B7B.B: 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	-0.009	-0.005
	[0.076]	[0.083]	[0.101]
More than 20 years old	0.045	0.048	0.075*
	[0.056]	[0.045]	[0.036]
Joint F-test of equality *p*-value	0.240	0.538	0.468

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 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.