Essays on the Economics of Education and Gender

Ipek Mumcu

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Institute for Social and Economic Research

University of Essex

September 2018

Summary

This thesis contains three studies that examine the role of education in skill accumulation in early childhood.

Chapter 1 examines the effect of elementary school teachers' beliefs about gender roles on student achievement. We exploit a natural experiment where teachers are assigned randomly to schools and students are allocated to teachers randomly. We show that girls who are taught for longer than a year by teachers with traditional gender views have lower achievement, and this effect is amplified with longer exposure to the same teacher. We find no effect on boys. The effect is partly mediated by teachers transmitting traditional beliefs to girls.

In Chapter 2, we analyze the effect of school starting age on children's learning outcomes and noncognitive skills through the first two years of primary school. We exploit a sharp date of birth cutoff for school entrance to identify the school starting age effect. We test whether the effect of entrance age varies with the child's readiness for school. We find that being young at school entrance has a negative and significant effect on verbal acquisition in the first year of schooling. In contrast, we find little evidence for entry age effect on social behavior growth and reduction of externalizing behavior.

In Chapter 3, I investigate the relationship between the noncognitive skills and aspirations in elementary school students using unique data on occupational and educational aspirations reported by students and teachers. I find a significant

gender difference in the type of skills that predict educational and occupational aspirations: While competitiveness and challenge seeking appear as two key predictors for boys, patience and self-confidence emerge as the only critical skills that predict aspirations for girls. I also find that teachers are more likely to have high aspirations for students who show inclination toward competitive and challenging tasks.

Declarations

No part of this thesis has been submitted for another degree.

Chapter 1 is co-authored with Professor Sule Alan and Associate Professor Seda Ertac. Chapter 2 is co-authored with Professor Tom Crossley and Associate Professor Nazli Baydar. Chapter 3 is exclusively mine.

Chapter 1 has been previously submitted for publication in The Review of Economics and Statistics (forthcoming, 2018).

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Tom Crossley for his continuous support, excellent guidance, encouragement and advice he has provided throughout my time as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so carefully. I am also indebted to my second supervisor Dr Birgitta Rabe for her helpful comments and guidance.

I would also like to thank Prof. Sule Alan and Assoc. Prof. Seda Ertac for their guidance and collaboration. It has been a great experience for me to work with them and be involved in each stage of research with their valuable experience. Their wisdom and knowledge have been a great source for my desire for academic research. I am also indebted to Assoc. Prof. Nazli Baydar for her collaboration in Chapter 2 with helpful suggestions and remarks.

I feel lucky to have started this journey and have met so many exceptional people along the way. Their friendship allowed me to see the world and life in an ever-expanding way. I thank you all for being a part of my life with your invaluable contribution to my passion for knowledge, arts and culture. Most of all, I would like to thank my family for their continuous love and support throughout my entire life. I thank them for encouraging me to pursue my dreams.

Lastly, I am grateful to Economic and Social Research Council for the financial support during my PhD study.

Contents

List of Figures

List of Tables

Chapter 1

Gender Stereotypes in the Classroom and Effects on Achievement

1.1 Introduction

Stereotypes about gender are pervasive in most societies. These views tend to rigidly define the innate capabilities and attitudes of each sex, and social roles that are deemed appropriate for men and women. To the extent that they influence the actual choices and outcomes of individuals, such beliefs may in large part contribute to gender-achievement gaps as well as the underrepresentation of women in top executive positions, STEM careers, and in leadership. As ample evidence suggests, such gender inequality, factually confirming and perpetuating traditional gender role beliefs, can be quite persistent (Bertrand and Hallock, 2001; Fortin, 2005; Bertrand, 2011; Blau et al., 2013).

The formation of gender role beliefs and conforming behaviors and attitudes likely begins very early in childhood, within the family, as families have the earliest, most

direct impact on children's beliefs and preferences (Bisin and Verdier, 2001).[1] Once a child starts school, factors that contribute to the formation of beliefs and attitudes become broader and more complex. In addition to their families, children now interact with their peers in a more structured environment and perhaps more importantly, with another adult, the teacher. Teachers' views toward gender roles may affect students' attitudes, behaviors and outcomes both directly and indirectly. First, a teacher's beliefs may influence students' achievement outcomes by influencing students' own beliefs: the teacher may express his/her views about gender-appropriate roles in the classroom, and because he/she is a significant authority figure, students may adopt and internalize what their teacher says. These beliefs may in turn influence girls' academic aspirations, their interest in male-stereotyped topics such as math, and their motivation to study for as well as the level of stereotype threat and anxiety they may experience in subjects in the male domain (Spencer et al., 1999).

A more direct mechanism is the teacher interacting differently with girls and boys. A teacher with strongly traditional gender role beliefs may think that acquiring academic skills is not as important for girls, since they are unlikely to put them into practice later in life. Such a teacher may reflect these beliefs in actual classroom practices, by giving different types of feedback to girls and boys, selectively answering/dismissing questions, or focusing on boys when teaching (Sadker and Sadker, 2010). Biases on the part of teachers can also manifest through discrimination in grading (either against or in favor of girls), and this can affect student achievement and choices (Lavy, 2008; Lavy and Sand, 2015; Terrier, 2015). In addition to directly influencing learning, such teaching practices on the part of biased teachers can affect long term outcomes by affecting the development of girls' non-cognitive skills as well.[2] A very progressive teacher, on the other hand,

[1] It has been documented that transmission of gender attitudes from mothers affects daughters' as well as daughter-in-laws' labor force participation and human capital (Farre and Vella, 2013; Johnston et al., 2013). Olivetti et al. (2016) find that women's work hours are positively affected both by the work behavior of their own mother and their peers' mothers.

[2] It is well-known that non-cognitive skills in childhood are predictive of many important

may exert extra effort to engage students in subjects that are typically considered in the domain of the opposite sex and try to break stereotypical attitudes in the classroom.

In this paper, we study the effect of teachers' beliefs about gender roles on their students' achievement outcomes, using rich data from a large-scale field study involving approximately 4000 3rd and 4th grade students and their 145 teachers. In order to identify these effects, we exploit the unique institutional features of our study site, Turkey. The educational system in Turkey provides us with a natural experiment with three main components: First, stratified by gender and pre-school education, state elementary school students are allocated to their teacher in first grade randomly. Second, teachers are appointed to schools centrally by the Ministry of Education based on the need for teachers and as such, they are prevented from self-selecting into catchment areas and schools before acquiring a considerable number of years of service. Finally, the general practice is such that students have the same teacher for the entire elementary school period, from grade one to four, and this is disrupted mainly by teacher rotations and, to a lesser extent, family relocations. This disruption provides us with variation in the number of years a student in a school is taught by the same teacher, allowing us to identify the mediating effect of length of exposure to a teacher with particular gender role beliefs. We provide details on these institutional features in Section 2.

The data reported in this paper were collected by the authors as part of a large field study, with the specific goal of exploring the role of the elementary school teacher in shaping children's beliefs and affecting their achievement outcomes. As such, the dataset includes a rich set of variables on student, family and teacher characteristics, which were collected by physically visiting the classrooms several times. Having access to a rich set of teacher quality indicators was our primary

outcomes over the life cycle (Almlund et al., 2011).

motivation in our data collection effort. This is because teachers' gender role beliefs are likely correlated with teaching quality, rendering the identification of the effect of these beliefs on achievement outcomes difficult.[3] A particular strength of our data is that detailed information on teachers with respect to their daily classroom practices, teaching styles, and their pedagogical approach to teaching as well as indicators of personal effort are collected through surveys.

We find that teachers' gender role beliefs have quite different effects on girls and boys. Girls taught by teachers with traditional views about gender roles for more than one year have lower performance in objective math and verbal tests, and this effect is amplified in size with longer exposure to the same teacher. If the teacher has been teaching the student for two to three years, a one standard deviation increase in teacher stereotypes leads to a 0.12 and 0.06 standard deviation decrease in math and verbal test scores, respectively. This negative effect becomes 0.21 for math and 0.11 for verbal test scores if the student is taught by the same teacher for the entire duration of elementary school (four years). We find no statistically significant effect of teachers' gender role views on boys' test scores.

We then explore various channels through which teachers' beliefs may affect girls' test scores. Our statistical mediation analysis shows that about 17% of the effect of a traditional teacher on girls' math test scores is coming from girls' gender roles beliefs being influenced by their teacher's gender role beliefs. Other potential mechanisms notwithstanding, our results suggest that teachers' influence on girls' beliefs on gender roles may be an important indirect channel. To the extent that these beliefs predict important real life outcomes such as choice of study major and occupation, we conjecture that the importance of this channel extends well beyond test scores.

[3]There is a large literature in economics that studies the effect of teacher quality on educational attainment (e.g. Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Hanushek, 2011; Harris and Sass, 2011; Chetty et al., 2014). See also Hanushek and Rivkin (2006), Schwerdt and Wuppermann (2011).

The role of teacher gender has been an important focus in the education literature, and it has been shown that having a female teacher may affect outcomes such as math performance, STEM grades and graduation rates on the part of female students (Bettinger and Long, 2005; Dee, 2007; Hoffmann and Oreopoulos, 2009; Carrell et al., 2010; Antecol et al., 2014). It has also been shown that the student-teacher gender (mis)match can influence a teacher's perceptions of the student (Dee, 2005). The effect of teachers' beliefs and attitudes with respect to gender roles, however, has received less attention. As mentioned above, a recent set of papers document effects of gender biases as reflected in discrimination in grading on student achievement and choices, with differing results (Lavy and Sand, 2015; Terrier, 2015). The former paper finds (in Israel) that boys are over-assessed, with negative effects on girls' achievement and future math course choices, while the latter finds (in France) that girls are favored in grading in math, and this increases girls' propensity for choosing a science track in high school. In addition, Lavy (2008) documents an anti-male bias in grading and Robinson and Lubienski (2011) also find that teachers rate girls more favorably than cognitive scores would suggest. The current paper differs from these studies in that we measure teachers' gender role beliefs directly rather than using grading biases, use variation in the duration of exposure to the teacher, and control for teaching quality and styles, which can be correlated with both teachers' gender attitudes and students' achievement. The paper contributes to the literature on teacher effects on achievement by showing that teachers' beliefs and attitudes are important in determining achievement outcomes and gender gaps in those outcomes, as well as in shaping the beliefs and attitudes of students. Our data, comprising teacher and student characteristics which are typically not available, allow us to construct a continuous measure of gender stereotypes to facilitate non-parametric as well as parametric identification. The unique educational setting allows us to estimate the mediating effect of the length of contact with a particular type of teacher. Our results highlight that the classroom environment, in particular the type of teacher,

is an important part of a child's social environment and already starts influencing children's performance and beliefs at the elementary school level. The results broadly suggest that gender-equal classroom practices, implemented early on by teachers with progressive views, could prevent gender gaps in achievement that likely cause multiplicative effects on academic persistence, occupational selection and labor market outcomes later in life.

1.2 Background

The Turkish 12-year compulsory education is based on a two-tier system, where both public and private schools are under the oversight of the National Ministry of Education. As Turkey has moved from low income to middle income status over the last 15 years, the majority of the middle- and upper-class parents prefer to send their children to private schools. Our study sample covers 3rd and 4th grade students in state-run (public) elementary schools in particularly needy areas of Istanbul. As such, it primarily represents Turkey's lower socioeconomic segment, with limited variation with respect to socioeconomic status.

In studying the impact of teachers' beliefs on the actual outcomes of students, one faces a fundamental selection issue, that is, students in a given school may be allocated to teachers in a non-random manner. This happens, for example, when a particular type of parent selects a particular type of teacher, a teacher known to be better or appearing to have similar beliefs and attitudes as the parent. If gender role beliefs somehow proxy unobserved teacher quality, for example if more progressive teachers are also more likely to use modern teaching methods or adopt a more constructive approach, or they are simply more intrinsically motivated and care more about their students' achievement, such selection compromises identification. Our setting circumvents this selection issue.

After the registration of all first-graders (school-starters) in a given academic year,

school administrators randomly allocate the students to teachers through publicly held draws in the presence of parents. Classroom sizes are not allowed to exceed 50, although a maximum of 30 is typically preferred. Draws are stratified based on gender and pre-school attendance to ensure balance in gender and school-preparedness in each classroom. Therefore, contrary to the private school system, there is no room for parents to choose their child's teacher in the state system. Of course, a parent may decide to send their child to a school that is not in the catchment area; however, acceptance of the student to a non-catchment area school is subject to the capacity of that school and priority is given to catchment area residents. Sending the child to a school that requires transportation is costly and relocations for educational purposes are extremely rare in this socioeconomic group. This, along with centrally managed teacher appointments, ensures that exposure to the same teacher is largely independent of teacher and student quality. Once students are allocated to classrooms in grade 1, re-mixing in later grades is extremely rare, which means that students remain with the same classmates until they graduate from elementary school, unless the family moves.

Despite the random allocation of students to teachers, if our gender stereotype construct is correlated with some unmeasured aspect of teacher quality, it would still be difficult to interpret our results as the causal effect of teacher's beliefs on student achievement. To isolate the effect of beliefs as much as possible, we collected very detailed information on teachers. In addition to demographic characteristics, these include the teacher's teaching philosophy, pedagogical approaches, classroom practices and indicators of effort and care for student achievement. We explain how we construct summary measures based on this information in Section 3.2.

The final issue to account for in studying the effect of beliefs on actual outcomes is the fact that such effects, if they exist, may take a long time to surface. It is plausible that the longer the exposure to the same teacher, the larger and more

persistent the effects may be. In many countries, elementary school students are taught by a different teacher each year, making it difficult to detect teacher effects. However, this is not the case in our study site. Except for involuntary rotations, re-appointments and retirement, a teacher teaches the students allocated to him/her from grade 1 to grade 4, after which those students move on to middle school. Because of the strictly centralized allocation of teachers and subsequent re-appointments and rotations (explained below), we have substantial exogenous variation in the length of time a given student has spent with the same teacher, which gives us a unique opportunity to study the role of the length of exposure in moderating impacts.

1.2.1 Allocation of Teachers to Schools

Although we exploit only the within-school variation to estimate the effects, it is important to provide a brief account of the way teachers are allocated, rotated and re-appointed (centrally) in our study site. This is because the specific features of this system will provide support for our exogeneity assumption with respect to the time spent with the same teacher, which allows us to identify the mediating effect of exposure. After completing the degree requirement, the current practice in the public system is that all teacher candidates take a nationwide civil servant examination and those above a cutoff score are placed in a pool to be appointed to a public school in need.[4] A new teacher has typically no say in which city, let alone district or school she will be appointed to. It is generally very difficult to be appointed to one's preferred city before 5 to 10 years, except for pure luck. In 2015, among over 300,000 new teachers, only 40,000 were appointed. The situation leaves no bargaining power to teachers as every year an increasing number of teachers remain unappointed, waiting for the next round of appointments.

[4]Private schools, despite being subject to the curricular requirements of the Ministry of Education, enjoy autonomy in implementing their own teacher selection process, and are not subject to the scrutiny of the Ministry in this regard.

Once appointed, teachers begin to collect service points that are assigned to their school. Each school has a score assigned to it by the Ministry, with schools in deprived and dangerous areas having higher scores than those located in well-off cities, districts within cities and catchment areas within districts. A teacher mechanically earns the points assigned to her school, for every year she teaches. There is no other way for a teacher to accumulate service points other than by simply teaching. These points are very important for teachers, as they determine their chances of being re-appointed to the city of their choice, or the district of their choice if they are already in a city they like.

After his/her appointment, a teacher can be re-appointed to another school (generally within the same city) if i) there appears to be an excess supply of teachers at her current school and she has the lowest service points among her colleagues (involuntary rotation), ii) her re-appointment request is honored.[5] A classroom may lose its teacher because of retirement and resignations but the most common reason is involuntary rotation due to excess supply and re-appointment to another school based on teacher request. Note that when a teacher is re-appointed to a new school, she is allocated to a classroom which is in need of a teacher. As this classroom can be of any grade, such moves contribute to the variation we observe in the length of exposure to a given teacher from the point of view of the student.

While teachers who want to move (as our schools are in relatively remote and deprived areas, most in our sample say they would like to, once they accumulate sufficient points) do so mainly to work in the district of their choice, the centralized system makes it very difficult for them to self-select into catchment areas and schools conditional on district. Such self-selection becomes possible only for a

[5]Teachers cannot ask to be re-appointed before completing at least 4 years (over 6 years in actual practice) of service in their current school. Requests to be re-appointed are honored if i) there is a school in need in the preferred district and ii) the teacher has higher service points than her competitors who have the same location preference. As working in high-SES catchment areas is more desirable for most teachers, there tends to be a high teacher turnover in low-SES district schools such as the ones that comprise our sample. For an Istanbul teacher, even with a long tenure in the profession, it is extremely hard to be appointed to the generally desired (high-SES) districts.

teacher with very high service points, usually having taught beyond 25 years or more than the usual amount of time working in high-point areas such as eastern Turkey. While we base our identification strategy (conservatively) on within-school variation through the use of school fixed effects, it is important to re-iterate that teacher sorting within district based on any metric other than service years, which we control for in our regressions, is largely ruled out in this system. In our results section, we show that teachers who have been teaching a class for a longer and shorter time are largely similar in terms of the rich observables we have such as demographics, qualifications and teaching styles.

1.3 Data

Our data were collected as part of a large-scale field project, which has been underway since 2013. The project aims to study the behaviors, attitudes and outcomes of students in conjunction with the behaviors and attitudes of teachers. All student data were collected by the authors of the paper by physically visiting all classrooms multiple times.[6] We took great care to ensure that the teacher was not present when the students worked on the tests and filled in the questionnaires.

Data were collected using a rich battery of tools, which includes surveys, a fluid IQ test and official grade records as well as objective mathematics and verbal tests that we prepared and conducted in the classroom. This endeavor required visiting each classroom multiple times to minimize disruption to daily teaching activities. Because there tends to be about a 20% non-attendance on each day due to sickness or other valid excuses, we do have some missing data on students. Our analysis is based on the teachers and students for whom we have complete information on key variables, forming a dataset with 31 schools, 145 teachers and approximately 4000 students. Our typical teacher is female, university educated,

[6]The project has local IRB approval as well as official state approval.

and has accumulated about 15 years of service as a teacher. Only about 25% of our teacher sample is male, as teaching in elementary school is still predominantly a female profession in Turkey. A little over 70% of our teachers have majored in a program called "class teaching", which is a 4-year university degree in elementary school teaching.

Our typical student in grade 3 (4) is 9 (10) years of age, and on average 70% of all third-graders have been taught by the same teacher for 2-3 years and 30% for one year. The respective percentages for the fourth grade is 55% for more than three years, 24% for two-to-three and 21% for one year.

1.3.1 Student and Family Characteristics

To account for the role of student and family characteristics in determining academic achievement, we collect rich measures of behaviors, attitudes and beliefs, as well as demographic information, information about the home environment, socioeconomic status and family background. For this, we use survey data from the students themselves as well as from their teachers. In particular, teachers fill out an extensive survey for each individual student, which includes questions regarding the attitudes and behaviors of the student within the classroom, the teacher's assessment of the student's attitudes, traits and performance, and his/her assessment of the student's family characteristics such as socioeconomic background. Student surveys also include questions regarding the student's home environment to better capture the socioeconomic status as well as the behaviors and attitudes of the parents.[7]

Our main outcome measure consists of standardized math and verbal (Turkish) tests, which we implement in each classroom in the absence of the teacher. These

[7]We did not attempt to collect these information directly from parents, as our previous experience shows that the response rate of parents is very low and their answers to the surveys questions are usually not reliable. Instead, we rely on the child and the teacher for this information.

tests were prepared (and extensively piloted) by the authors of the paper based on the national curricula. An independent set of teachers were consulted to tailor the questions to each grade (3 and 4). Students' cognitive ability is measured via Raven's Progressive Matrices. We also have access to students' official math, verbal and behavior grades, all given by their own teachers.

1.3.2 Teacher Characteristics

The primary purpose of the paper is to show the effect of the teacher's gender role beliefs on students' achievement outcomes. However, we acknowledge that these beliefs are likely to be correlated with certain underlying teacher characteristics that are likely instrumental for student achievement. For example, without adequately controlling for teacher quality, even in the absence of selection, it is difficult to give the association between beliefs and achievement outcomes causal interpretation due to the plausible correlation between gender role beliefs and quality. While there is consensus that teacher quality matters a lot for achievement over and above student characteristics (cognitive and non-cognitive skills) and family background, it has proven to be very difficult to measure it.[8] This is possibly because teacher quality is multidimensional, often involving unobservables such as teaching styles, effort and care. Acknowledging this difficulty, we collect two sets of additional information from our teachers, with the hope of better capturing the often unobserved components of teaching quality.

First, in addition to their education, experience and study majors, we collect a set of variables that relate to the teaching styles and pedagogical approach of our teachers. Teachers' styles of teaching the class material and interacting with their students, as well as their expectations from the students, are likely to be important factors in student outcomes (Domino, 1971; Schwerdt and Wuppermann, 2011; Bietenbeck, 2014; Hidalgo-Cabrillana and Lopez-Mayan, 2015).

[8] See Carrell and West (2010).

Using item-set questions directed to teachers, we construct four distinct teaching style variables. We call these modern vs. traditional, growth- vs. fixed-mindset, warm vs. distanced, and extrinsic vs. intrinsic motivator. A traditional teaching style is reflected in the teacher dictating to the students what to do in class, and following a rigid structure to each class that is determined by the teacher. What we call a modern approach to teaching, on the other hand, involves the students more in the learning process and aims to induce the children to think critically.[9] Having a growth mindset (Dweck, 2008) is the belief that abilities are malleable and success can be achieved provided that sufficient effort is exerted, regardless of innate characteristics. Such a mindset has been found in the literature to predict academic achievement (Blackwell et al., 2007; Alan et al., 2016). From the perspective of the teacher, we measure growth mindset through questions about the relative importance of innate ability vs. sustained effort for success, e.g. whether or not the teacher agrees that any student could become the best in the class if he/she works hard enough. The warm vs. distanced construct gets at how authoritarian the teacher is in his/her interactions with the students, and how important it is for him/her to establish a close and warm relationship with them. Finally, extrinsic motivator refers to the use of extrinsic rewards in motivating students (such as stickers, small gifts and applause for good performance) and punishment for inducing desired behavior. The full inventory we use to construct each style score is given in the Appendix.[10]

In addition to teaching styles, a crucial variable to control for is teaching effort or how much teachers care for students' achievement. However, the motivation and effort level of teachers are difficult to observe. The educational system we study, where there are no extrinsic incentives for teachers to maintain a high level of teaching, makes intrinsic motivation somewhat easier to measure, since any

[9]Estimating the effects of traditional vs. modern teaching practices on achievement has been an active research topic in the economics of education (e.g. Bietenbeck, 2014).

[10]Some of these questions were adapted from the "Teaching and Learning International Survey" (TALIS) questionnaire (OECD, 2013), whereas others were constructed by the authors.

extra-curricular activity done by teachers reflects voluntary effort.[11] We therefore collect information on teachers' extra-curricular activities that focus on teaching improvement and student achievement through our survey. We believe this is informative of the teacher's (typically unobserved) care and effort in our setting. This is because, as mentioned before, teachers collect service points passively, by teaching only. No other activity or certificate or diploma will matter in collecting service points, which is required for re-appointments, salary increases and retirement benefits. Having said that, there are many certificate and diploma programs as well as conferences and social projects that aim to inform teachers about best classroom practices based on new evidence, with the goal of improving student achievement. Teachers who participate in these programs do so in a voluntary manner, paying participation fees (if any) themselves and sacrificing time during their evenings and weekends. Similarly, teachers do not gain anything other than professional satisfaction by organizing educational class trips, which often cost them money and require considerable effort, mainly because of the lack of parental interest in the socioeconomic segment we cover. We take the reported frequency of these volunteer activities as measures of teacher effort.

1.3.3 Measuring Gender Role Beliefs

We measure the gender stereotypes of both students and teachers using the same questionnaire. This questionnaire includes a battery of item-set questions based on a 4-point Likert scale, with which we construct a "gender stereotype score" for each teacher and each student. Some examples are "It is more important for boys to go to college than girls", "Women cannot play football well even if they try hard", "It is the father's responsibility to earn a living in a family, and it is the mother's responsibility to take care of the children", which are to be answered using the scale of "I strongly agree", "I agree", "I disagree" and "I strongly disagree" (the

[11] Providing extrinsic incentives to teachers based on student achievement has been found to have ambiguous results (e.g. Fryer, 2013).

full set of questions is given in the Appendix). Figure 3.1 shows the distribution of the stereotype scores of children and teachers, with larger numbers representing more traditional views. In both panels we see substantial variation in gender role beliefs, with male students and male teachers generally reporting more traditional views. For female teachers, we observe a clear pattern of piling up at the extremes (very progressive and very traditional) with considerable variation in between. For children, the distributions look fairly normal.

Table 1.1 presents the predictive power of teacher characteristics on the teacher's gender role beliefs. While male teachers seem to hold more stereotypical views about gender roles, this relationship does not reach statistical significance, and once teaching styles and effort are controlled for, it becomes even weaker. Years of experience have no bearing in predicting teachers' gender role beliefs. Several other interesting findings are noteworthy here. First, in terms of on-paper qualifications, teachers with a plain education (class teacher) degree are more likely to hold traditional beliefs about gender. This may be because this degree is less academically demanding and individuals who select into (or are placed because of their university entrance exam performance into) this major may be coming from a more traditional or less affluent background. Second, our teaching style constructs are by far the best predictors of teachers' gender role beliefs. Adding these constructs to the regression increases the R-square substantially (from 4% to 35%), and not surprisingly, a joint test of all style measures having no effect is decisively rejected. Among these style constructs, growth mindset and warmth are the most important factors in determining teachers' gender role beliefs. Third, only one of our effort measures is statistically significant. Finally, the number of years taught in the same class does not predict teachers' gender role beliefs.[12] We

[12]We also estimate a probability model for teaching the same class long-term. Table A.1 in the Online Appendix presents the results. Based on observable teacher characteristics, we do not find any consistent evidence suggesting that the teachers who taught the same class for a long term are a selected group. The only noteworthy exception is teachers with a linguistics degree. We find that they are about 0.53 percentage points more likely to stay in the same school for a long-time (p-value=0.01). We believe that this is due to the excess demand for teachers who can teach a foreign language in addition to regular class teaching in needy schools.

now turn to estimating the effect of teachers' gender role beliefs on the achievement outcomes of students.

1.4 Results

While we were informed by school officials that the students are allocated to teachers within schools randomly, it is still useful to see whether our data attest to that. To do this, we look at the balance of fixed student and family characteristics across types of teachers. We construct two types to facilitate this balance check. Teachers with gender stereotype scores below the median are taken to be "progressive", while those with scores above the median are taken to be "traditional". While we use our continuous measure in our main analysis, this categorization also help us conduct a causal mediation analysis as detailed in Section 5. Table 1.2 presents the mean characteristics of students and families for traditional and progressive teachers. As can be seen clearly, all fixed student characteristics (including fluid IQ) and family characteristics that are unlikely to be affected by teachers' beliefs are balanced across the two types of teachers. The most notable evidence against the possibility of ability sorting is that our measure of IQ (elicited via Raven's progressive matrices) is balanced across the two types of teachers.[13]

1.4.1 Empirical Specification

We use the following empirical model to estimate the effect of teachers' gender role beliefs on students' outcomes:

These teachers are less likely to be rotated by the Ministry upon appointment.

[13]We also performed another check that involves predicting student achievement with only family socioeconomic indicators and looking at the correlation between the predicted values and teacher gender stereotype scores. If there is significant ability sorting, this correlation would be statistically significant. In both math and verbal and for both genders, we find no significant correlation between predicted test scores and teachers' gender views (p-values for math: girls=0.95, boys=0.16; for verbal: girls=0.39, boys=0.72). These findings provide supportive evidence that allocation of students to teachers is indeed random.

$$y_{iks} = cons + \alpha_1 Exposure_{iks} + \alpha_3 GRB_{ks} + \alpha_4 Exposure_{iks} * GRB_{ks} + \mathbf{X_{1,iks}}\beta + \mathbf{X_{2,iks}}\gamma + \mathbf{X_{3,ks}}\theta + \delta_s + \varepsilon_{iks}$$

where y_{iks} is the standardized test score for student i, who is being taught by teacher k in school s. The variable $Exposure$ captures the number of years student i has been taught by teacher k in school s. The variable GRB_{ks} is the continuous (standardized) score that measures the gender role beliefs of teacher k, with larger numbers representing more traditional beliefs. The interaction term allows for a differential effect of the teacher's beliefs on student outcomes with respect to the length of exposure to the teacher. Matrix $\mathbf{X_1}$ contains student characteristics such as age (in months), cognitive ability (as measured by the Raven IQ test), student's own gender role beliefs, student mindset, behavior score assigned by the teacher and an academic self-confidence measure. Matrix $\mathbf{X_2}$ contains family characteristics and socioeconomic indicators, and $\mathbf{X_3}$ contains teacher characteristics such as gender, experience as a teacher, education, study major, teaching styles and effort. Finally, δ_s denotes school fixed effects.

We divide the exposure variable into three groups: Children who have been taught by the participating teacher for at most one year are labeled as "1-year exposure", those who have been taught for more than one year and at most three years are labeled as "2-3 year exposure" and those who have been taught for more than three years (at most four years) are labeled as "4-year exposure".[14] As mentioned before, we have substantial variation in exposure due mainly to teacher relocation and to a lesser extent, family relocation.[15] Note that only fourth-grade students

[14]Because of the small sample size with respect to teachers in two-year exposure, we are not able to divide "2-3 year" further. We provide disaggregated estimation results in the Appendix (Figure A.1).

[15]About 13% of the students have been exposed to the same teacher less than their classmates. We consider them as re-locators. We were informed that the newcomers are allocated to classrooms in a random manner. Unreported regressions reveal that while they seem to be

can be taught by the same teacher for more than three years in our sample, therefore our results regarding long-term exposure relates to fourth graders.

Given the random allocation of students to teachers, the (conditional) exogeneity of length of exposure, and the fact that we allow for school fixed effects, the coefficient estimates α, which are the estimates of interest, can be interpreted as causal effects. Despite our efforts of collecting very detailed information on teachers, we are cautious about the possibility that teacher gender role beliefs may still be capturing some unmeasured aspect of teacher quality. However, our i) gender-differential results presented in the next subsection, and ii) mediation analysis in Section 5 largely mitigate this concern.

1.4.2 Gender Role Beliefs of Teachers and Student Achievement

We estimate the empirical model presented in Section 4.1 separately for girls and boys. In addition to being of direct interest, looking at the effect of the beliefs separately for each gender also allows us to answer the question of whether beliefs still capture some unmeasured aspect of teacher quality. If, although we control for many important teacher characteristics, beliefs still proxied teacher quality, we would expect to estimate similar effects on both genders unless such omitted characteristics have differential effects on boys and girls. We argue that this is unlikely to be the case and revisit the issue in Section 5. Table 1.3 presents the results by suppressing the coefficient estimates of student, family and teacher characteristics. Table A.2 in the Online Appendix gives the full results, which shows almost all cognitive and noncognitive ability measures we have are highly predictive of math and verbal test scores for both boys and girls. For math

more likely to come from very low SES, their cognitive and non-cognitive skills, including their math and verbal test scores do not appear to be different from the rest of the sample. Results are available upon request.

scores for example, a one standard deviation increase in the Raven (IQ) score is associated with 0.35 (0.23) standard deviations increase in math scores for girls (boys). Another important finding is that students' own gender role beliefs are also strong predictors of test scores for both genders: a one standard deviation increase in the gender stereotype score (going toward more traditional views) leads to about a 0.14 (0.12) standard deviation decrease in math scores girls (boys) and 0.12 standard deviation decrease in verbal scores for both boys and girls.

We now turn to the question of whether the teacher's beliefs affect girls' and boys' outcomes differently. What is clearly seen in Table 1.3 is that the teachers' gender role beliefs affect math and verbal test scores only for girls. The impact on math test scores is of considerable size, particularly when the girls have been taught by the same teacher for a long time (four years). A one standard deviation increase in teachers' gender stereotyped beliefs lowers girls' test scores in mathematics by about 0.21 standard deviations. The effect for an exposure of 2-3 years is smaller: a one standard deviation increase in teachers' gender stereotyped beliefs lowers girls' test scores in mathematics by about 0.12 standard deviations. While the equality of coefficients for 4-year and 2-3 year exposure is not rejected for either gender, we estimate a statistically significant effect of 4-year exposure to the same teacher relative to 1-year exposure for girls. No such effect is present for boys. Remarkably similar findings are obtained for the verbal scores (columns 3 and 4). Again, the impact of the teacher's stereotyped beliefs on girls' verbal test scores in the long term is of considerable size (0.06 and 0.11 standard deviations for 2-3-year and 4-year exposure, respectively) and statistically significant at the 5% level.

When we test the effect of teachers' gender stereotypes for each exposure length across boys and girls, for math, we reject equality only for the 4-year exposure group (p-value=0.066) but for verbal, girls have a significant short-term advantage that is lost as they are exposed to the gender-biased teacher for a longer time.

These results suggest that traditional gender role beliefs on the part of the teacher have a detrimental effect on girls' performance in both mathematics and verbal tests. However, the effects become visible after they spend some years with the same teacher. No such effect is present for boys. Finally, boys' math scores are significantly positively affected by long-term exposure to the same teacher, regardless of the teacher's gender role beliefs.[16] For girls in math, this relationship is weak and is reversed by being exposed to a teacher who holds traditional beliefs.

Since our measure of gender role beliefs is a continuous construct, it would be informative to present the functional relationship between test scores and teacher's beliefs in a non-parametric fashion. For this, we relax our assumption of a linear parametric model and modify our empirical model as follows:

$$y_{iks} = cons + \mathbf{X_{1,iks}}\beta + \mathbf{X_{2,iks}}\gamma + \mathbf{X_{3,ks}}\theta + \delta_s + f(GRB_{ks}) + \epsilon_{iks}$$

Here, while all student, family and teacher characteristics enter the model linearly, we allow for test scores to be a non-parametric function of the teacher's gender role beliefs (GRB). We estimate this model separately for boys and girls for each exposure length. Recall that larger numbers of GRB indicate more traditional (stereotyped) beliefs. Figure 1.2 depicts the results for math test scores. Our findings from the linear models clearly re-emerge for girls in these pictures. Looking at 4-year and 2-3 year exposure results, one can see the decreasing and fairly linear relationship between the gender stereotypes of teachers and girls' math test scores. For boys on the other hand, we observe a rather non-linear relationship where at the very extreme (most progressive teachers) they exhibit similar patterns as girls: boys' math scores are higher under extremely progressive teachers, however the relationship breaks down as the teacher becomes more conservative. It appears that except for the case of an extremely progressive teacher, boys may

[16]Related to this result, Hill and Jones (2018) find that repeat student-teacher matches have a significantly positive effect on student achievement in similar (3rd to 5th grade) elementary school students, pointing to the benefit of staying with the same teacher.

even be benefiting from a teacher's traditional gender role beliefs (note the slight positive relationship, not considering the extremes). For 1-year exposure, the relationship is virtually flat for both boys and girls, with again some evidence of both genders benefiting from a very progressive teacher.

As for the verbal scores, Figure 1.3 depicts the negative functional relationship between the teacher's beliefs and girls' verbal test scores. With again the exception at the corner (most progressive teachers), the relationship is flat for boys. Overall, our results suggest a significantly gender-differential effect of the teacher's gender role beliefs on student achievement. Under both parametric and non-parametric specifications, we estimate a declining and fairly linear relationship for girls under 4-year exposure to the same teacher, while no obvious (statistically significant) pattern of relationship emerges for boys. We now turn to investigate the sensitivity of our results to various issues raised earlier.

1.4.3 Robustness

The behavior at the extreme (very progressive teachers) is noteworthy. Given the similar (positive) effects of such teachers on the test scores of both boys and girls, it may be that some omitted aspects of teacher quality are proxied well with extreme progressiveness. In Table A.3, we re-estimate Table 1.3 by excluding very progressive teachers, in order to see how sensitive our results are to these particular teachers. For this, we exclude teachers whose gender stereotype score is lower than the 10th percentile (15 teachers, two of them male). As can be seen in the table, the results for girls, especially for math scores remain very strong, although we lose some precision for verbal results.

Even though our identification relies on within-school variation through the use of school fixed effects, we conduct another robustness check that is related to teacher sorting into schools. Recall that the institutional structure leaves very little room

for self-selection of teachers into catchment areas/schools and our sample consists of generally "undesirable" schools. However, although still difficult, teachers who have accumulated high service points (those with higher number of years of service) might be able to self-select into relatively more desirable schools. Given that working in a catchment area of one's choice is generally ruled out before 20 years of service except purely by chance, we re-estimate our linear model by excluding the teachers who have more than 20 years of service in the teaching profession. This excludes 24 teachers from our sample. Table A.4 presents the results for boys and girls separately. Results are both qualitatively and quantitatively very similar to our full sample results.

Another concern one might have is that a teacher's beliefs may reflect what he/she observes in the class. Suppose that in a given classroom or cohort, boys are indeed better academically than girls. If the teacher bases his/her beliefs on this particular cohort, our results would reflect this reverse causality rather than the effect of the teacher's beliefs on achievement. Our rich data, however, allows us to address this issue. Our teacher survey includes a question where we ask the teacher whether he/she has observed boys or girls to be better at math (or equal) in his/her experience as a teacher. When we exclude the teachers who report boys to be better (only 7 teachers), our results remain the same (see Table A.5).[17]

Responses to the question of which gender tends to be better at math also reveal that the teachers in our sample do not maintain stereotyped beliefs about mathematical ability across gender. 56% of our teachers report that they have observed girls to be better at math and about 39% report that both genders are equally good, with only about 5% thinking boys are better.[18] The lack of a stereotype

[17]Our results also hold when we entirely exclude this question from our gender role belief construct and base the measure on other domains of gender stereotypes than math performance.

[18]In our data, the unconditional performance of girls and boys in objective math test is similar; however, the dummy for male becomes strongly and positively significant in explaining math performance once we control for other student characteristics. As for verbal performance, the unconditional performance of girls is significantly higher but this advantage turns statistically insignificant once we control for student characteristics. All these hold true for math and verbal grades as well.

about math ability is also evident in our findings regarding grades. As can be seen in Table 1.4, we observe absolutely no effect of teachers' gender role beliefs on students' grades. The absence of an effect on grades suggests that the effects we estimate on objective achievement scores do not reflect reverse causality, i.e., they are not coming from teachers' factual beliefs about ability (based on their observations over the years or in their current classroom).[19] In the next section, we explore a potential mechanism that may lead to these results.

1.5 A Causal Mediation Analysis

Recall that Table A.2 shows that various student characteristics, which may be affected by teachers' gender role beliefs, are highly predictive of test scores and therefore may be potential mediators of the effects we estimate. An obvious one is students' own gender role beliefs. If girls adopt the biased beliefs held by their teacher, this may diminish their ambitions, aspirations, and motivation towards academic tasks, reducing their achievement. Another mediator may be self-confidence. Our measure of self-confidence is derived from a survey item designed to measure students' beliefs on their math performance ("In math, I am: very good/good/mediocre/not very good/not good at all"). A traditional teacher may potentially affect girls' confidence in mathematics by either directly voicing beliefs about girls' capabilities or praising/focusing on boys more in math. Finally, another potential mediator could be the students' mindset on achievement, i.e. whether students have a "growth mindset" that highlights the importance of effort or a "fixed mindset" that emphasizes innate abilities. Gender-biased teachers, who hold fixed views of what each gender can and cannot do, may influence the achievement mindset of students, particularly girls. This shift towards a fixed

[19]The absence of an effect on grades despite the effect on objective tests may also point to the fact that grades tend to reflect non-cognitive skills and good behavior in addition to pure exam performance, especially in elementary school (e.g. Brookhart, 1993; McMillan et al., 2002; Borghans et al., 2016; Jackson et al., 2016). Such effects may also potentially explain findings of grading biases in favor of girls (e.g. Terrier, 2015).

mindset may in turn lead to lower motivation and performance, as has been shown in the literature (Blackwell et al., 2007; Alan et al., 2016).

In addition to these indirect channels, or alternatively, teachers' gender role beliefs may affect student achievement directly. A teacher with strongly traditional gender role beliefs, who thinks that it is more important to get boys to do well in school, may adopt classroom practices that reflect these beliefs; for example, asking questions to and answering questions from girls and boys differently, providing more feedback to boys, and generally focusing academic attention more on boys while praising girls for gender-consistent behavior such as compliance and obedience (Dweck et al., 1978). These practices may impede girls' learning directly, without necessarily affecting their own gender role beliefs.

In order to establish whether and how large a part of the effect on test scores is coming through these potential mediators, we perform a statistical mediation analysis. For this, we use an extension of the potential outcomes framework developed by Imai et al. (2010) to estimate causal mediation effects. To make the analysis feasible and facilitate straightforward interpretation, we use a binary teacher gender stereotype score to serve as a binary treatment indicator. Teachers with scores below the median are taken as "progressive", while those with scores above the median are taken as "traditional".[20] Recall that conditional on school, being exposed to a particular type of teacher is random in our setting.

While the random assignment to a type of teacher is sufficient to identify the total effect, additional (strong) assumptions are required to identify the average causal mediation effect (ACME) and the average direct effect (ADE). Imai et al. (2010) show that ACME and ADE can be nonparametrically identified under the "sequential ignorability" assumption, which constitutes two sequential conditions. The first one states that given the pre-treatment confounders, treatment assign-

[20]Doing this analysis with a continuous treatment variable is not trivial. Also the interpretation of the results would be very difficult.

ment is independent of the potential outcomes and potential mediators. The second one states that the mediators are independent of the potential outcomes conditional on pre-treatment confounders and the treatment assignment. While we make use of our rich data on numerous student, family and teacher characteristics that potentially affect both the mediators and the outcome, the latter is still a very strong assumption.

To estimate the average effects (ACME and ADE), we proceed in several steps. First, we posit and fit regression models for the mediator (say, students' own gender role beliefs) and the outcome of interest (test scores). The mediator model includes the treatment dummy (traditional teacher) as well as any relevant covariates. The outcome is modeled as a function of the mediator and the treatment dummy, as well as all covariates. Based on the fitted mediator model, we then generate two sets of predicted mediator values for each girl, one under a progressive teacher and the other under a traditional teacher.

We then use the outcome model to impute potential outcomes. For each girl, we first obtain the predicted value of the outcome corresponding to the traditional teacher and the predicted mediator value for the treatment condition (obtained in the previous step). We then generate the predicted counterfactual outcome, i.e. the outcome where the treatment indicator is still set to 1 (traditional teacher) but the mediator is set to its predicted value under the progressive teacher (also obtained in the previous step). Finally, we compute the average causal mediator effect by averaging the differences between the predicted outcome under the two values of the mediator across observations in the data.

Table 1.5 presents the effects of teacher beliefs on the three potential mediators we consider. Pooling all exposure lengths, we estimate that a traditional teacher increases girls' gender stereotyped beliefs by about 0.20 standard deviations (p-value=0.001). The relationship is not statistically different from zero for boys (p-value=0.96). We estimate no effect on self-confidence and mindset for either

boys or girls. Table 1.6 presents the average causal mediation effect (ACME), average direct effect (ADE) and total effect for both math and verbal test scores for each gender. Overall, we estimate that a traditional teacher lowers girls' math scores by about 0.16 standard deviations. About a 0.03 standard deviation of that (17%) comes from girls' gender role beliefs being affected by their teacher's gender role beliefs. The rest of the effect, not mediated by beliefs, may be due to the direct effect of factors such as lower academic attention on girls by traditional teachers.[21]

The results on verbal scores is quite interesting. The total effect of the teacher's gender role beliefs on verbal performance is not statistically different from zero in this specification; however, transmission of the teacher's gender role beliefs to female students leads to an approximately 0.02 standard deviation decline in verbal scores, making the total effect smaller than ADE. Both ACME estimates (math and verbal) are statistically significant (see the 95% confidence intervals). It should be noted here that these numbers are just direct effects of level shifts in gender role beliefs. It is quite possible that changes in these beliefs affect performance through indirect influences on girls' perceived or true production function. For example, a girl who holds biased beliefs may have lower motivation in a mathematical performance task. As expected, all estimates are not statistically different from zero for boys.[22]

Note that our analysis show that self-confidence is not a potential channel. This finding along with the finding of a significant effect that is mediated by girls' gender role beliefs points to the role of potential indoctrination about what is expected of a woman, which may lower girls' academic motivation or ambitions. That is, rather than lowering girls' self-confidence about their capabilities, tra-

[21] When we exclude short-term exposure (as we find no effect in this case), we lose considerable precision in the mediator model and this results in a lower percentage (about 14%) of the total effect being mediated.

[22] We also performed this analysis using gender roles in the family as a potential mediator and ruled it out. Results are available upon request.

ditional teachers may emphasize appropriate roles for them in the society. If traditional teachers emphasize traditional gender roles whereby girls do not need to be as ambitious as boys in the academic domain (because they won't need to use these skills as much), this may manifest in lower academic motivation in girls, although their beliefs about their capabilities do not necessarily go down. In fact, the set of questions in our survey about appropriate gender roles (e.g. the proper division of labor within the family) are responsible for the effect that comes from student beliefs. Among those, item set questions such as "it is the father's responsibility to earn money for the household", "it is natural for girls to help more than boys in household chores" are highly strong mediators when considered in isolation. We should note that the traditional teacher may also place less academic attention on girls, which may have a strong "direct" effect on their learning that is not mediated through student beliefs.

One alternative explanation of our differential results across gender would be a differential response of girls and boys to teaching quality.[23] Although we have a large set of controls for teacher characteristics, if gender role beliefs still capture an unmeasured aspect of teacher quality and girls' achievement is more responsive to this, similar patterns would emerge. Our data, however, provides suggestive evidence against this. Table A.2 shows that boys' achievement is at least as responsive to teacher characteristics as girls'. Coefficient estimates on teacher characteristics do not suggest that girls are in any way more responsive to quality, styles and approach. Along with the result that teacher gender role beliefs are transmitted to girls more strongly, these results give us confidence that our findings are coming from the teacher acting on biased gender views and conveying these beliefs to children, rather than an unmeasured aspect of teacher quality (correlated with teacher gender role beliefs) affecting girls differentially.

[23]Deming et al. (2014)show that at the high-school level, girls respond to attending a better school with higher grades and taking more courses to prepare for college.

1.6 Conclusion

We exploit a natural experiment to show that teachers' gender role beliefs have a significant impact on girls' math and verbal test scores. Our unique setting allows us to identify the effects moderated by the duration of teacher contact with students. Controlling for student, family and teacher characteristics, we show that girls whose teachers maintain more traditional (progressive) views about gender roles have lower (higher) performance in objective math and verbal tests, and this effect is amplified with longer exposure to the same teacher. For boys, we find no significant effect.

The large dataset we use, collected with the purpose of answering the research question we pose in this paper, allows us to control for a host of teacher, student and family characteristics that are crucial for identifying the effect of gender role beliefs on achievement. The results show that controlling for the teacher's own gender and other characteristics, teachers' beliefs about gender roles affect the test scores of their female students, both in mathematics and verbal tests. It is striking that even without any apparent biases or discrimination in grading, teachers' traditional gender role beliefs still affect girls' achievement outcomes negatively. Our mediation analyses show that a non-trivial portion of the effect comes from the teacher transmitting his/her traditional gender role beliefs to girls. These results indicate that the personal views of the elementary school teacher may play an important role in mitigating or widening gender-achievement gaps, particularly in countries where pervasive gender inequality has been found to contribute to differences in math performance across gender (Guiso et al., 2008). Given that our sample comes from the low socioeconomic tier, our results are also generalizable to vulnerable segments of societies, where patriarchal gender roles are particularly imposing and where improving achievement is a policy imperative (Heckman, 2006).

Two caveats are worth mentioning. First is the fact that our data is a cross section. Panel data with some baseline information on students before they were exposed to a particular teacher would of course be ideal, especially to pin down heterogeneous effects of teacher types. The second one is the external validity of our results. To circumvent the issue of ability sorting of students, we exploit our unique country setting and choose our sample from lower socioeconomic strata (relatively deprived areas of Istanbul). In this group, teachers are prevented from self-selecting into schools and students are randomly allocated to teachers. While giving us a clean identification of the effects of teacher types on achievement, this choice may prevent us from generalizing our findings to the population. Future work should focus especially on these two issues.

Given the importance of the childhood period for long-term choices and outcomes, the results suggest that the type of teacher a child is assigned to in elementary school may have long-lasting consequences. In particular, improved math scores of girls may lead to reductions in gender gaps in the labor market, given the evidence that math performance and math education predict future income (Paglin and Rufolo, 1990; Joensen and Nielsen, 2009). The implication for educational policy is that achieving gender equality in teaching practices and attitudes early on, possibly by training teachers to raise awareness of such biases and their effects, could have substantial value for preventing inefficient gender gaps in achievement, occupational selection and labor market outcomes.

Figures

Figure 1.1: Distribution of Gender Role Beliefs

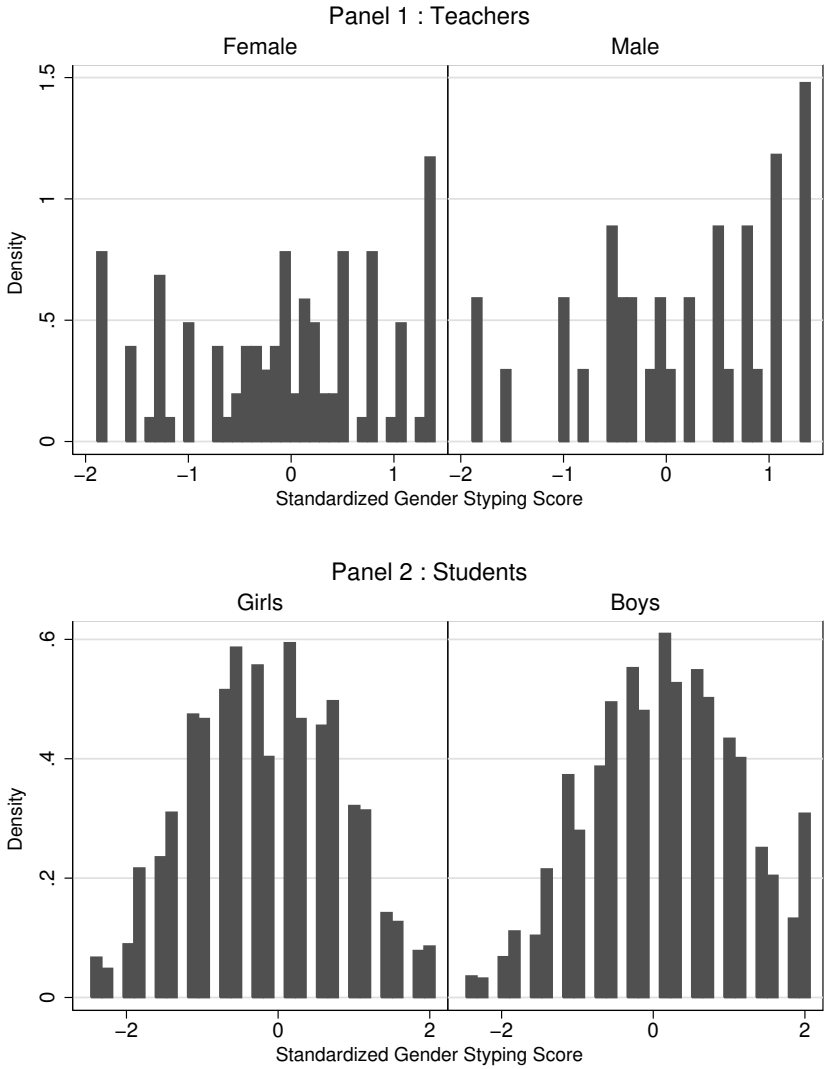
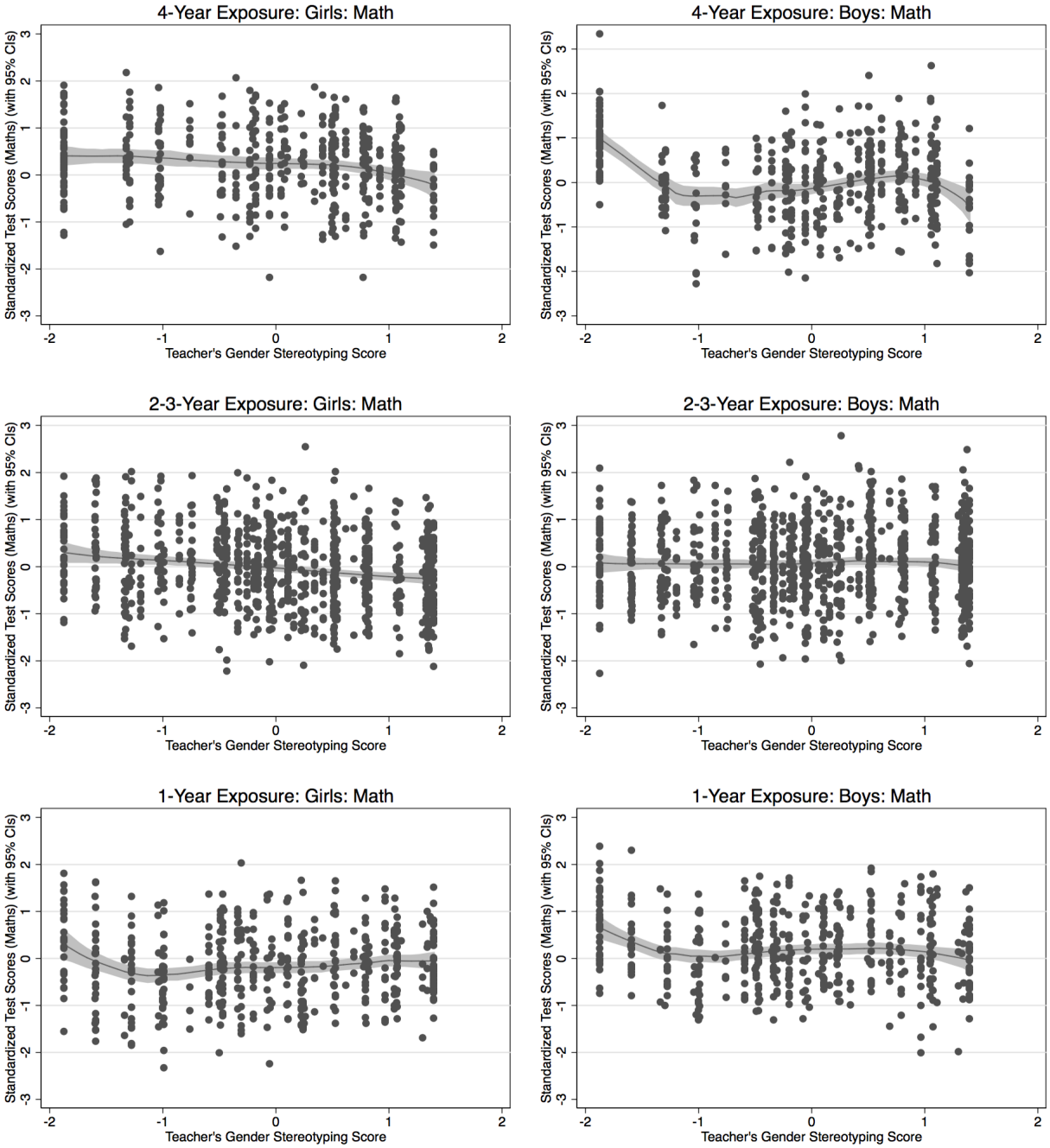
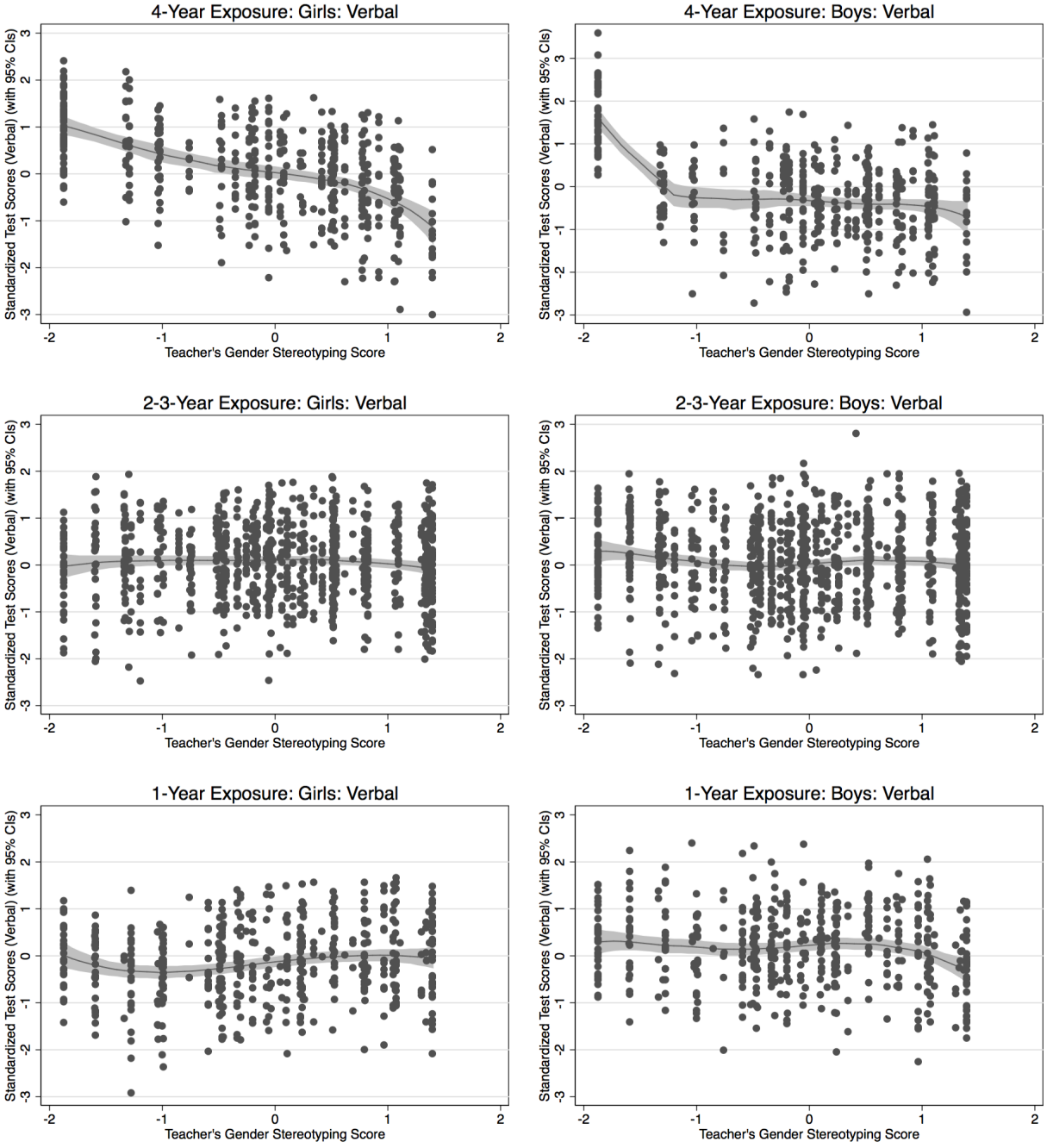


Figure 1.2: Teacher Gender Stereotyping and Math Test Scores: Non-parametric



Figures plot the non-parametric estimates (and 95% confidence bands) of the effect of teacher's role beliefs on math test scores for girls (column 1) and for boys (column 2). All student, family and teacher characteristics enter the model linearly, and school fixed effects are included.

Figure 1.3: Teacher Gender Stereotyping and Verbal Test Scores: Non-parametric



Figures plot the non-parametric estimates (and 95% confidence bands) of the effect of teacher's role beliefs on verbal test scores for girls (column 1) and for boys (column 2). All student, family and teacher characteristics enter the model linearly, and school fixed effects are included.

Tables

Table 1.1: Predictors of Teachers' Gender Role Beliefs

Male	0.241	0.251	0.269	0.101	0.124
	(0.18)	(0.18)	(0.19)	(0.17)	(0.17)
University Degree		-0.167	-0.173	-0.168	-0.171
		(0.31)	(0.32)	(0.31)	(0.29)
Graduate Degree		-0.506	-0.503	-0.457	-0.470
		(0.42)	(0.42)	(0.38)	(0.37)
Years of Experience		-0.007	-0.007	0.004	0.004
		(0.01)	(0.01)	(0.01)	(0.01)
Number of Terms in the Same Class		0.028	0.029	0.015	0.029
		(0.04)	(0.04)	(0.03)	(0.03)
Education Degree			0.336	0.409	0.595***
			(0.34)	(0.25)	(0.21)
Linguistics			-0.203	-0.271	-0.188
			(0.39)	(0.33)	(0.32)
Natural Sciences			0.271	0.141	0.230
			(0.40)	(0.19)	(0.23)
Social Sciences			-0.149	-0.235	-0.191
			(0.30)	(0.23)	(0.24)
Growth Mindset				-0.188***	-0.181***
				(0.03)	(0.03)
Extrinsic Motivator				0.033	0.016
				(0.04)	(0.04)
Modern Approach				-0.013	0.000
				(0.03)	(0.02)
Warm Approach				-0.101***	-0.096***
				(0.03)	(0.03)
Number of Extra_C Programs					0.005
					(0.01)
Number of Volunteer Activities					-0.041***
					(0.01)
N	145	145	145	145	145
R-Squared	0.01	0.03	0.04	0.35	0.39

The dependent variable is the teacher's standardized gender stereotype score. It is constructed in a way that larger values indicate more traditional gender role beliefs. Heteroskedasticity-robust standard errors are in parentheses.

Table 1.2: Balance Across Teacher Types

	Fixed Student Characteristics		
	Progressive	Traditional	P-Value
Male student	0.51	0.51	0.862
Age (in months)	109.5	109.8	0.739
IQ (Raven Score)	0.09	0.07	0.628
	Family Socioeconomic Indicators		
	Progressive	Traditional	P-Value
Working mother	0.30	0.26	0.130
Computer at home	0.75	0.75	0.675
Family gender roles	2.30	2.31	0.813
Low SES	0.34	0.36	0.760
Medium SES	0.44	0.43	0.847
High SES	0.22	0.21	0.845

The table presents mean values of fixed student characteristics (upper panel) and family socioeconomic indicators for progressive and traditional teachers. Progressive (traditional) teachers are defined as those whose gender role beliefs are below (above) the median score. IQ is measured (and standardized to have mean zero and variance 1) via Raven's Progressive Matrices. Binary indicators of whether the mother is working, whether there is a computer at home, and gender roles in the family are reported by the child. The latter is a question based on a 4-item scale that asks how much the father takes part in household chores. Family income/wealth level (SES) is reported by the teacher based on a 1-5 item scale and low, medium and high SES indicators are constructed based on these.

Table 1.3: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.000	-0.055	0.054	-0.094*
	(0.06)	(0.05)	(0.04)	(0.05)
2-3 Year Exposure	0.022	0.058	0.026	0.033
	(0.07)	(0.06)	(0.07)	(0.08)
4 Year Exposure	0.117	0.193***	0.015	0.006
	(0.08)	(0.07)	(0.08)	(0.07)
2-3 Year Exposure*Teacher G-Styping	-0.120*	0.001	-0.110**	0.020
	(0.06)	(0.06)	(0.05)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.211**	-0.016	-0.162**	-0.026
	(0.08)	(0.07)	(0.07)	(0.07)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.229	0.792	0.428	0.442
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.067	0.992	0.037	0.764
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.014	0.831	0.032	0.719
P-value: 1 Year E*G-Sty[Girls=Boys]	0.393		0.004	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.157		0.689	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.066		0.834	
N	1870	1943	1873	1946
R-Squared	0.32	0.34	0.26	0.26

Dependent variables are standardized test scores. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Table 1.4: Heterogeneous Effects of Teacher Gender Role Beliefs on Grades

	Math Grade		Verbal Grade	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.060	0.082	0.144**	0.158*
	(0.10)	(0.12)	(0.07)	(0.09)
2-3 Year Exposure	0.035	0.134	0.094	0.141
	(0.10)	(0.11)	(0.09)	(0.10)
4 Year Exposure	-0.058	0.061	0.003	-0.020
	(0.11)	(0.12)	(0.10)	(0.11)
2-3 Year Exposure*Teacher G-Styping	0.069	0.004	-0.057	-0.100
	(0.10)	(0.12)	(0.06)	(0.10)
4 Year Exposure*Teacher G-Styping	0.028	-0.026	-0.039	-0.053
	(0.12)	(0.14)	(0.09)	(0.11)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.536	0.681	0.770	0.525
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.506	0.977	0.368	0.314
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.813	0.852	0.670	0.644
P-value: 1 Year E*G-Sty[Girls=Boys]	0.808		0.852	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.454		0.535	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.672		0.988	
N	1594	1652	1594	1652
R-Squared	0.42	0.45	0.37	0.42

Dependent variables are standardized grades given by the teacher. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs distanced, extrinsic vs intrinsic motivator, traditional vs modern and growth vs fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Table 1.5: Mediator Model: The Effect of Teachers' Beliefs on Students' Beliefs

	Gender Role Beliefs		Self Confidence		Growth Mindset	
	Girls	Boys	Girls	Boys	Girls	Boys
Traditional Teacher	0.199***	-0.003	0.091	0.080	-0.003	-0.074
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.07)
School Fixed Effects	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓	✓	✓
N	1888	1967	1888	1967	1888	1967

Dependent variables are standardized scores of students' gender role beliefs, self-confidence and growth mindset. The binary variable "Traditional Teacher" takes the value 1 if the teacher's beliefs are above the median score and zero otherwise. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Table 1.6: Potential Channels for the Effects on Test Scores: Causal Mediation

PANEL 1: Math Test Scores			
	G. Role Beliefs	Self-confidence	Growth Mindset
ACME	**-0.028**	0.011	-0.000
	[-0.046, -0.011]	[-0.004, 0.028]	[-0.004, 0.003]
ADE	-0.136	-0.136	-0.136
	[-0.241, -0.021]	[-0.241, -0.021]	[-0.241, -0.021]
TOTAL	-0.164	-0.125	-0.136
	[-0.269, -0.045]	[-0.232, 0.008]	[-0.241, -0.021]
Percentage Mediated (%)	**17%****	-8.5%	0.12%

PANEL 2: Verbal Test Scores			
	G. Role Beliefs	Self-confidence	Growth Mindset
ACME	**-0.024**	0.006	-0.001
	[-0.041, -0.010]	[-0.002, 0.019]	[-0.011, 0.010]
ADE	0.048	0.048	0.048
	[-0.060, 0.165]	[-0.060, 0.165]	[-0.060, 0.165]
TOTAL	0.024	0.054	0.048
	[-0.083, 0.144]	[-0.053, 0.172]	[-0.059, 0.165]
Percentage Mediated (%)	30%	9.3%	-0.32%

ACME: Average causal mediation effect, ADE: Average direct effect. G.R.: Gender role. Estimates (standard deviation effects) and 95% confidence intervals are obtained via Imai et al. (2010). The estimation sample is restricted to girls only. Number of simulations is 1000. **: significant at 5%.

Appendix

A.1 Additional Tables and Figures

Table A.1: Probability of Teaching the Same Class Long-Term

Teacher G-Styping	-0.000	0.007	-0.009	0.013
	(0.04)	(0.04)	(0.05)	(0.05)
Male		-0.086	-0.041	-0.052
		(0.10)	(0.10)	(0.10)
Years of Experience		0.004	0.001	-0.001
		(0.01)	(0.01)	(0.01)
Education Degree		-0.009	-0.006	-0.046
		(0.17)	(0.16)	(0.19)
Linguistics		0.391*	0.565***	0.528***
		(0.22)	(0.15)	(0.15)
Natural Sciences		-0.115	-0.088	-0.106
		(0.18)	(0.17)	(0.14)
Social Sciences		-0.190	-0.158	-0.155
		(0.13)	(0.13)	(0.13)
Growth Mindset			-0.026	-0.021
			(0.02)	(0.02)
Extrinsic Motivator			-0.017	-0.010
			(0.02)	(0.02)
Modern Approach			0.032***	0.028**
			(0.01)	(0.01)
Warm Approach			0.001	0.004
			(0.02)	(0.02)
Number of Extra_C Programs				0.004
				(0.01)
Number of Volunteer Activities				0.021
				(0.01)
N	145	145	145	145
Pseudo-R-squared	0.00	0.05	0.09	0.11
Significance of model test	1.00	0.28	0.04	0.02

Reported estimates are average marginal effects from logit regressions. The dependent variable "Long-term teaching" is defined as a binary variable, which takes the value 1 if the teacher has been teaching the same class for 4 years in grade 4 and 3 years in grade 3. Heteroskedasticity-robust standard errors are in parentheses. The last row "Significance of model test" gives the p-value for joint significance of all covariates used in the corresponding specification.

Table A.2: Teacher Gender Role Beliefs and Student Test Scores - Details of Table 3 in the main text

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.000	-0.055	0.054	-0.094*
	(0.06)	(0.05)	(0.04)	(0.05)
Medium Term	0.022	0.058	0.026	0.033
	(0.07)	(0.06)	(0.07)	(0.08)
Long-Term	0.117	0.193***	0.015	0.006
	(0.08)	(0.07)	(0.08)	(0.07)
Medium Term*Teacher G-Styping	-0.120*	0.001	-0.110**	0.020
	(0.06)	(0.06)	(0.05)	(0.07)
Long Term*Teacher G-Styping	-0.211**	-0.016	-0.162**	-0.026
	(0.08)	(0.07)	(0.07)	(0.07)
Student Characteristics:				
Age(months)	0.005*	0.004	-0.001	0.001
	(0.00)	(0.00)	(0.00)	(0.00)
Raven Score	0.350***	0.225***	0.298***	0.257***
	(0.03)	(0.02)	(0.02)	(0.02)
Teacher's assesment: well-behaved	0.110***	0.137***	0.151***	0.127***
	(0.03)	(0.02)	(0.03)	(0.02)
Academic Self-confidence	0.129***	0.155***	0.079***	0.115***
	(0.02)	(0.02)	(0.02)	(0.02)
Student G-Styping	-0.138***	-0.116***	-0.119***	-0.124***
	(0.02)	(0.02)	(0.02)	(0.02)
Student GMS	0.018	0.037*	0.085***	0.068***
	(0.02)	(0.02)	(0.02)	(0.02)
Family Characteristics:				
Middle SES	0.108**	0.127**	0.068	0.140**
	(0.05)	(0.05)	(0.05)	(0.05)
High SES	0.186***	0.141**	0.158***	0.197***
	(0.05)	(0.06)	(0.06)	(0.07)
Working Mother	0.017	0.038	0.087*	-0.029
	(0.05)	(0.04)	(0.05)	(0.05)
Computer at Home	0.037	0.152***	0.004	0.105**
	(0.04)	(0.05)	(0.05)	(0.04)
G-Styping at Home	0.034*	0.022	0.024	-0.008
	(0.02)	(0.02)	(0.02)	(0.02)
Teacher Characteristics:				
Male Teacher	-0.033	0.120**	0.081	0.046
	(0.07)	(0.06)	(0.06)	(0.07)
Teacher Qual - 2 Year College	-0.252*	0.224**	0.111	0.116
	(0.14)	(0.10)	(0.11)	(0.13)
Teacher Qual - Grad S	-0.018	-0.014	0.050	0.010
	(0.10)	(0.10)	(0.10)	(0.07)
Years of Teaching	0.006	-0.002	-0.010**	-0.008*
	(0.00)	(0.00)	(0.00)	(0.00)
Linguistics	0.090	0.201*	0.180	-0.039
	(0.13)	(0.12)	(0.11)	(0.15)
Sciences	-0.100	0.088	0.159*	0.064
	(0.09)	(0.08)	(0.08)	(0.09)
Social Sciences	-0.181	0.071	-0.131	-0.014
	(0.13)	(0.11)	(0.13)	(0.13)
Other	-0.160	-0.011	0.045	-0.002
	(0.12)	(0.12)	(0.10)	(0.09)
Teacher Styles:				
GMS	-0.041***	-0.035***	0.013	-0.021
	(0.01)	(0.01)	(0.01)	(0.01)
Extrinsic Motivation	0.005	-0.018	0.008	-0.011
	(0.02)	(0.01)	(0.01)	(0.01)
Modern Approach	-0.006	0.011	0.010	0.001
	(0.01)	(0.01)	(0.01)	(0.01)
Teacher Warmth	-0.021	-0.033***	-0.003	-0.010
	(0.01)	(0.01)	(0.01)	(0.01)
Teacher Effort:				
Occupational Trainings	-0.000	-0.000	-0.002	-0.005
	(0.00)	(0.00)	(0.00)	(0.00)
Extra-curricular	0.003	0.015***	-0.004	0.012**
	(0.01)	(0.01)	(0.01)	(0.00)
School Fixed Effects	✓	✓	✓	✓
N	1870	1943	1873	1946
R-Squared	0.32	0.34	0.26	0.26

Table A.3: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores, Excluding Very Progressive Teachers

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.027	-0.028	0.069	-0.102*
	(0.08)	(0.06)	(0.05)	(0.06)
2-3 Year Exposure	0.006	0.048	0.009	-0.005
	(0.07)	(0.07)	(0.07)	(0.08)
4 Year Exposure	0.137	0.200***	-0.029	-0.021
	(0.09)	(0.07)	(0.07)	(0.08)
2-3 Year Exposure*Teacher G-Styping	-0.135*	-0.021	-0.124**	0.061
	(0.07)	(0.07)	(0.06)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.231**	-0.007	-0.102	0.044
	(0.10)	(0.09)	(0.09)	(0.08)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.248	0.861	0.782	0.803
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.061	0.763	0.042	0.404
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.028	0.937	0.257	0.556
P-value: 1 Year E*G-Sty[Girls=Boys]	0.467		0.007	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.219		0.781	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.063		0.740	
N	1729	1798	1732	1801
R-Squared	0.33	0.34	0.25	0.26

Dependent variables are standardized test scores. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level. Teachers who scored lower than the 10th percentile (very progressive) are excluded.

Table A.4: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores: Teachers with Less than 20 Years of Service

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	-0.010	-0.029	0.065	-0.090*
	(0.06)	(0.05)	(0.04)	(0.05)
2-3 Year Exposure	-0.016	0.027	-0.011	0.015
	(0.07)	(0.07)	(0.07)	(0.08)
4 Year Exposure	0.089	0.213***	-0.015	-0.013
	(0.09)	(0.07)	(0.08)	(0.07)
2-3 Year Exposure*Teacher G-Styping	-0.122*	-0.005	-0.126**	0.031
	(0.07)	(0.06)	(0.05)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.222**	-0.031	-0.177**	-0.044
	(0.09)	(0.07)	(0.08)	(0.07)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.202	0.686	0.454	0.224
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.063	0.926	0.021	0.642
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.011	0.678	0.024	0.555
P-value: 1 Year E*G-Sty[Girls=Boys]	0.758		0.003	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.038		0.953	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.017		0.728	
N	1695	1755	1698	1758
R-Squared	0.32	0.33	0.25	0.26

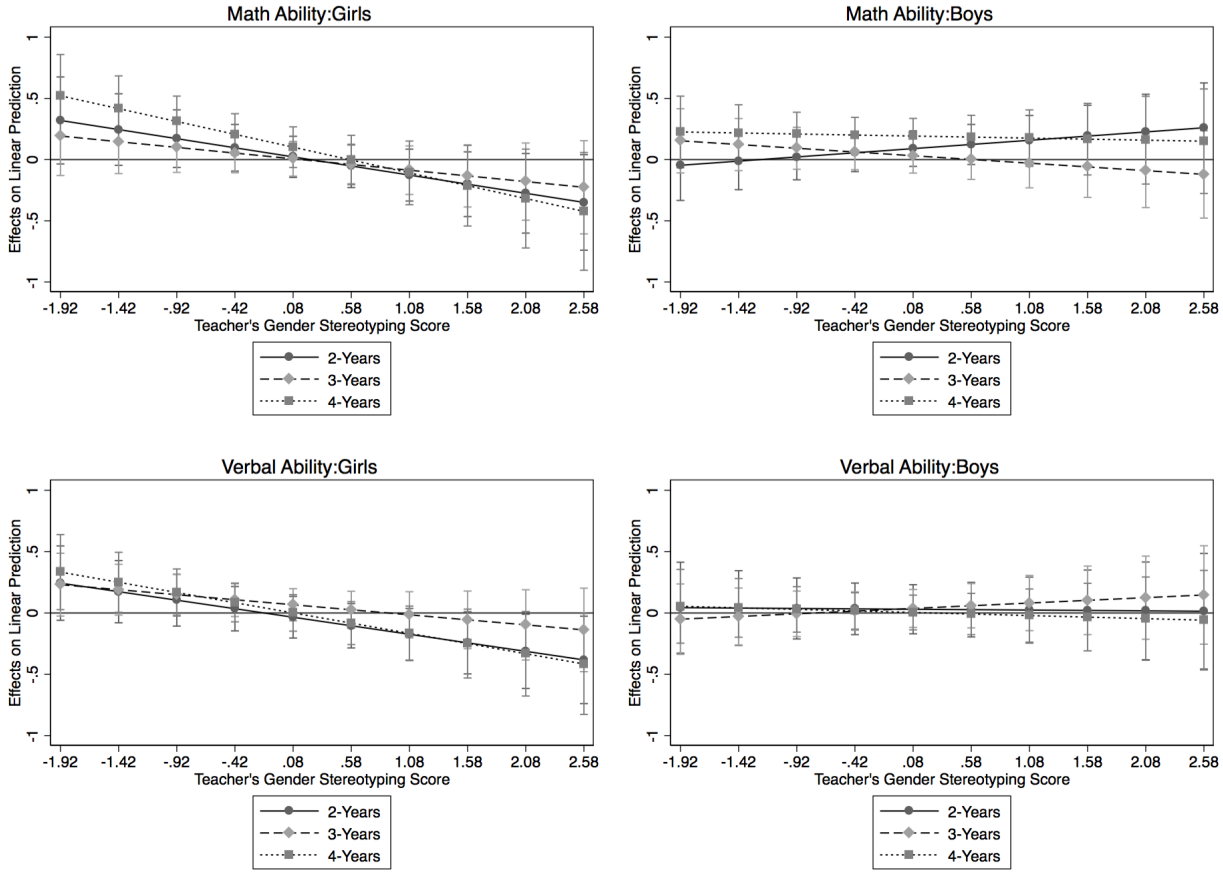
Dependent variables are standardized test scores. Estimated coefficients are obtained by constraining the sample to teachers who have less than 20 years of service. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student- reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Table A.5: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores: Excluding Teachers Who Believe Boys are Better at Math

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.008	-0.057	0.063	-0.090
	(0.07)	(0.05)	(0.05)	(0.06)
2-3 Year Exposure	0.006	0.062	0.036	0.014
	(0.08)	(0.07)	(0.07)	(0.08)
4 Year Exposure	0.098	0.210***	0.017	0.004
	(0.09)	(0.08)	(0.08)	(0.08)
2-3 Year Exposure*Teacher G-Styping	-0.113*	0.008	-0.127**	0.040
	(0.07)	(0.06)	(0.06)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.208**	-0.009	-0.171**	-0.020
	(0.08)	(0.08)	(0.08)	(0.08)
School Fixed Effect	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.215	0.781	0.496	0.329
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.097	0.899	0.029	0.579
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.015	0.904	0.031	0.802
P-value: 1 Year E*G-Sty[Girls=Boys]	0.324		0.007	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.253		0.751	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.084		0.992	
N	1772	1836	1775	1839
R-Squared	0.32	0.34	0.26	0.27

Dependent variables are standardized test scores. Estimated coefficients are obtained by dropping the teachers who reported that, in their teaching experience, they observed that boys were better at math than girls. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Figure A.1: Effects on Test Scores: Linear Predictions



Figures present linear predictions (and 95% confidence bands) obtained from estimating the specification presented in Table 3 in the main text, where teacher stereotyping score is interacted with exposure categories (1,2,3 and 4), rather than 1, 2-3 and 4.

A.2 Student and Teacher Surveys

Questions Used for Constructing Gender Role Beliefs–Students

4-point item scale: completely agree, agree, disagree, completely disagree

1. It is the father's responsibility to earn a living in a family, and it is the mother's responsibility to take care of the children.

2. Being a nurse is not a suitable profession for a man.

3. Men cannot sew well even if they try hard to learn it.

4. Women cannot play football well even if they try hard to learn it.

5. Men are better at math than women.

6. Being a space scientist/astronaut is not a suitable profession for a woman.

7. It is more natural for girls to help with housework than boys.

Questions Used for Constructing Gender Role Beliefs

4-point item scale: completely agree, agree, disagree, completely disagree

1. It is the father's responsibility to earn a living in a family, and it is the mother's responsibility to take care of the children.

2. Being a nurse is not a suitable profession for a man.

3. Men cannot sew well even if they try hard to learn it.

4. Women cannot play football well even if they try hard to learn it.

5. Men are better at math than women.

6. Men generally understand money-related issues better than women.

7. It is much more important for boys to go to university than girls.

8. It is more natural for girls to help with housework than boys.

9. Men have better judgment compared to women, hence they are better leaders.

Questions Used for Constructing Teaching Styles

4-point item scale: completely agree, agree, disagree, completely disagree

Growth vs. Fixed Mindset

1. Intelligence is a fixed trait. One cannot change how smart he/she is.

2. People can improve their intelligence regardless of their innate level.

3. Only very few people can excel in arts, music and sports, as innate ability is required to be successful.

4. Working hard does not make you successful in a task unless you are talented.

5. If a student works hard enough, he/she can be the best in the class.

Extrinsic vs. Intrinsic Motivator

6. Punishment is necessary to attain a disciplined and ordered classroom environment.

7. I often reward students (applauding, giving stars etc.) to elicit the outcomes and behaviors that I aim for.

8. I often punish students (grounding them on the breaks, making them sit alone etc.) to elicit the outcomes and behaviors that I aim for.

9. Rewarding behaviors or outcomes with material incentives (giving them stars and stickers etc.) prevents students from developing intrinsic motivation.

Modern Approach vs. Traditional Approach in Teaching

10. A noisy classroom is not a problem as long as students are busy with learning.

11. It is important to let students express their ideas regardless of how wrong and absurd they are.

12. I do not like to fall behind on the syllabus due to students' problems and questions or any other reason.

13. It is more efficient to teach students the correct answers directly rather than asking them questions and spending time on their potentially wrong answers.

14. Students should be entitled to choose what activities we do in the class.

15. When a student asks a question about a subject he/she is curious about, I only answer it if it is related to the subject I am covering at that moment. If it is irrelevant, I leave it to a later time not to disrupt the class flow.

Warm vs. Distanced

16. Teachers should keep their distance and be the authority in their relationship with the students, as this is beneficial for the students' development.

17. My educational standards and expectations from students can be described as strict and prescriptive.

18. Inculcating a strict discipline and ability to obey in students during elementary school, despite being difficult, is very beneficial for them further in their lives.

19. Having a warm teacher-student relationship and a classroom environment where students feel comfortable is more important for effective learning than a respect-based teacher-student relationship and a quiet classroom.

Chapter 2

The Effect of School Starting Age and Readiness for School on Verbal Ability Acquisition and Socio-emotional Skill Development

2.1 Introduction

Investment in early childhood is an important channel to influence later life outcomes. Recent empirical evidence documents that cognitive and non-cognitive skills at early ages are strong predictors of educational attainment, labor market performance, adult earnings, social and behavioral outcomes in later ages (Cunha and Heckman, 2008; Cunha et al., 2010). Hence early education can play a critical role in development of these skills. As learning in early ages influences student's academic motivation and their attitudes towards education, it facilitates

student's performances and skill development at later ages to be more effective Heckman (2006). In recent years, an increasing discussion on when children are ready to learn has dominated the policy debate. Chronological age is often set as the standard for assessing school readiness while children's cognitive, emotional and physical maturation is overlooked when considered for eligibility. Compulsory schooling policies across countries also reflect the extent of the discussion on what age is the optimal for starting school. Current rules vary from 5 to 7 seven years of age. The majority of top ten performing countries in 2015's PISA math assessment have age six as their compulsory schooling age [1]. Countries such as UK, Australia and New Zealand, which performs above OECD average but ranked below top ten, implement compulsory schooling age policies as low as five years of age. Scandinavian countries which have persistent reputations for performing better in international assessments have later school start age policies with seven years of age. Although each country has some unique agenda for its education policy, differences in school start age across countries and performance gaps are at the core of public debate in many. Additionally, parents are also concerned that their children are at risk of being permanently disadvantaged when being among the youngest in the classroom. Many parents of children born in months before the cutoff rule take the decision to hold their children out of primary school for a year. This rising trend in school entry age across several states in the US is well documented. Deming and Dynarski (2008) note that the one-third of the increase in the recent years are attributed to the changes in legislative rules of entrance as almost many states have increased the age at which children are allowed to start primary school.

In this paper, we investigate whether the age of entry to formal schooling affects the rate of skill acquisition in early years in elementary school students. We use data from a unique cohort study of child development in Turkey, the Study of Early Childhood Ecologies in Turkey (ECDET). We use the sharp cutoff in enrollment

[1] These countries include Singapore, China, Japan, Estonia and Canada.

eligibility as a source of exogenous variation in school starting age. Cutoff rule imposes a child who was born at the cutoff date to start formal schooling one year later than a child who was born a day prior to the cutoff date. We also account for potential endogeneity due to parental choice of delayed entrance in our estimations of the causal effect of starting age. The decision of delaying entrance may originate from parental preference, parents' assessment of the child's characteristics or other family attributes. Failing to take into account those characteristics in estimation methodology may lead to biased results for the effect of school starting age. A key feature of ECDET is that it measures mothers' perceptions of children's readiness for school. This allows us to look for heterogeneity in the effect of starting school young. We compare the outcomes of four groups of students defined according to their relative age and their school readiness, assessed prior to starting elementary school.

Policies of school-entry age might affect the children's learning outcomes in two ways. First, these policies directly determine the absolute age of starting formal education which introduces children to a new learning environment with standardized materials and teaching practices. Formal education, in that regard is an intervention to the children's maturational development under the available child care and home care arrangements. Since the development of early emotional, intellectual and social skills in pre-school period is an important factor for skill accumulation, the timing of this intervention poses a fundamental question for effective transition to schooling. Second, the school entry policies also determine the relative age of the students to their peers. Children who are born at different months of the year but are subject to start school at the same academic year are placed in the same classroom where the age differences between the oldest and the youngest student can go up to 12 months or more. This relative age at entry could also affect how well the child adopts to school and performs in the core subjects. Academic research on the effect of school start age aims to explain whether this effect is due to one's relative age or absolute age at when introduced to formal

education (Stipek, 2002). Our focus is the effect of relative age arising from the differences in the months of birth.

A growing literature explores the differences in ages when children start school as a potential mechanism for performance gaps. Studies of in-school test scores provide substantial evidence that the relatively older students outperform, on average, their younger peers with higher cognitive test scores and better non-cognitive skills (Puhani and Weber, 2008; Stipek and Byler, 2001; Fredriksson and Öckert 2006)[2]. Moreover, this achievement gap between older and younger students persist in later grades, even though, it diminishes somewhat over time (McEwan and Shapiro, 2008; Elder and Lubotsky 2009; Bedard and Dhuey, 2006)[3]. Elder and Lubotsky argue that the effect stems from endowment differences and skills accumulated in pre-schooling period. A linear growth curve in early childhood implies that children who are relatively old at the school entry have also acquired higher level of cognitive and non-cognitive skills compare to their younger peers up until formal schooling. Heckman and colleagues have extensively studied question of complementarities within skills in human capital accumulation, where initial higher level of skill leads to a higher return to investment in that skill(Heckman et al., 2006; Heckman and Masterov, 2007; Cunha and Heckman, 2008). If there are such complementarities in the production of human capital, then early investment in children's skill development will have large returns because they raise the return to future investments. The empirical evidence further suggests that the late entrance to formal schooling reduces the probability of grade retention and increases higher education participation (McEwan and Shapiro 2008; Crawford et al. 2010). Early entrants are less likely to enroll to a pre-university pro-

[2]This achievement gap among students with different birth month varies across grades, being roughly around 0.5 and 0.9 standard deviation for the reading and math scores in the kindergarten (Elder and Lubotsky, 2009) and 0.3 and 0.2 standard deviation in grade 4 and 8 (Bedard and Dhuey, 2006).

[3]Other country-level studies, reporting positive outcomes on performances on late entrants include Black et al. (2011) for Norway, Fredriksson and Öckert (2006) for Sweden, McEwan and Shapiro (2008) for Chile, Puhani and Weber (2008) for Germany, Crawford et al. (2010) for England, Smith (2009) for Canada and Elder and Lubotsky (2009), Datar (2006)for the United States.

gram and take the university admission exam Bedard and Dhuey (2006) while being among the oldest at the entrance increases the probability of attending the most academic educational track Puhani and Weber (2008). However, there is contradicting evidence on persistence of the effects of starting age on long-term outcomes. Fredriksson and Öckert (2006) find that the effect of delayed entrance leads to higher adult wages, whereas Dobkin and Ferreira (2010) and Black et al. (2011) find small negative or no significant effect of later school entrance on earnings. This is consistent with the idea that opportunity cost of one year experience in labor market fades away in later ages.

The discussion on the effect of age at entry does not extensively cover the developmental trajectories in non-cognitive skills. Few studies on non-cognitive outcomes find that entering school one year later reduces the risk of negative social behaviors as children show fewer symptoms of internalizing behaviors, hyperactivity and temperament in early grades of primary school (Mühlenweg et al., 2012;Datar and Gottfried, 2014). Early non-maternal care is also suspected to have detrimental effects for children's development of non-cognitive skills during the early years of childhood, during which effective engagement of mother or main caregiver is essential. Belsky, in series of paper, claim that provision of daycare during infancy with more than 20 hours per week raises the levels of aggression, noncompliance and results in overall poorer socio-emotional skills for children aged between 3 and 8 (Belsky, 1986,1986,1988;Belsky and Rovine, 1988). Even after including family and background factors that are likely to be powerful predictors of the developmental trajectories some studies report the presence of adverse effect of day care. Children who stay in day care longer exhibit less social competence, more externalizing behavior and more adult-childhood conflict at age 4 to 5(NICHD,2003). On the other hand, Children who enter child-care before age 1 perform high in school with better school adjustment skills and social competence, at age 8 and 13 (Andersson, 1992).One concern with the findings of these studies is their incapability to identify a causal relationship. NICHD(2003) argue that incorporating

background factors for family characteristics increases the plausibility of a casual link, but may not be sufficient to address it fully. To provide a reliable inference of the provision of the non-maternal care, Baker et al. (2005) exploit the exogenous expansion in the day care policies and conclude that children are worse off in a range of behavioral outcomes such as aggression, hyperactivity, anxiety and social development. Although these findings do not directly reflect the effect of age at entrance to education, they are subject to similar mechanism of channelling the schooling effect. While being exposed to day care longer constraints the quantity of the mother-child interaction, being an early entrant in primary school induces an early switch of environment from home to a different learning environment.

Analysis on early school outcomes is essential in understanding the risks associated with school entry policies for young starters. However, studies using the outcomes measured while children are still in school suffer from a methodological limitation in separating the effect of entrance age from age at measurement and duration of schooling. For example, when outcomes of students are collected at the same grade, meaning length of schooling is fixed across all subjects, the chronological age observed at measurement differs due to differences in the months of birth. Hence any estimated effect of month of birth gives the combined effect of relative age at school entrance and chronological age. Empirical results that in favor of a strong advantage for late starters is claimed to reflect the effect of chronological age instead of the effect of age at school start.The older students may simply perform better in tests because they are more mature by the time of the test. Similarly, when children are tested at the same age, older starters have lower years of schooling compared to the younger ones, creating undesirable dispersion in total time spent in school . Controlling for the age at measurement in fact causes the effect of entrance age to be combined with the effect of length of schooling. When this is the case, the direction of the relationship is more ambiguous as the older entrants spent less time in school[4].

[4]Identification problem is similar to Age-Period-Cohort analysis.

Datar (2006) studies the changes in test scores of students in the US to resolve this methodological problem while accounting for the variation both in the entrance age rules and month of births. Their estimation strategy utilizes on the direct association between difference in test score gains and differences in starting age that is independent of age at measurement. In this study, we adopt a similar methodological strategy to identify the age effect on the verbal ability scores and behavioral outcomes. We show the effect of age on verbal ability is linear prior to school start, which we refer to as natural maturational process.

We find that children starting school young make less progress in verbal ability relative to their older counterparts through the first year of school. Young entrants, regardless of their readiness for school, are among the least performing students. Delayed entry increases the verbal ability growth by 0.4 standard deviation for unready and 0.7 standard deviation for ready children. Although, there are statistically significant differences in the rates of verbal ability acquisitions in the first grade, the group disparities are eliminated in the second grade. The analysis for non-cognitive skills produces no substantial evidence to support the progress in prosocial behavioral in first and second grade. We find that old entrants achieve a higher rate of reduction in their externalizing behavioral outcomes and the effects are more profound in second grade.

The rest of the paper proceeds as follows. In Section 2 we outline our empirical strategy. Section 3 gives details of our data, ECDET survey, and how key measures were constructed. Results are presented in Section 4 and Section 5 concludes.

2.2 Econometric Framework

We take advantage of the available longitudinal survey which allows us to track individual characteristics and outcomes across time. We organize the data by child and by year of schooling, rather than time or age. The latent ability outcome k,

for child i in school year s is denoted by y_{is}^k. We expect that the growth in a cognitive or non-cognitive measure, $\triangle y_{i1}^k$, from the start of the school year to beginning of the next reflects a natural maturation process with age, the gains from schooling and shocks. Our examination of prior-to-school data on verbal ability score (Figure 2.1) and the prior literature (Datar, 2006) suggests that the maturational process is linear[5]. However, we allow it to be heterogeneous across students, reflecting latent ability in measure k.

We assume that conditional on an individual specific maturation rate, shocks are independent over time. Thus we adopt the following form for the growth of measure k, for child i through the s$^{\text{th}}$ year of school.

$$\triangle y_{is}^k = \alpha_i^k + \theta_{i,s}^k + \beta_s y_{i,s-1}^k + e_{s,i}$$

$$\alpha_i^k = Z_i \gamma_k + u_i$$

where α_i^k captures natural maturation, $\theta_{i,s}^k$ captures the gains from schooling in grade s and $e_{s,i}$ captures the individual and time specific shocks. We also include the initial level of outcome k to control for any predetermined differences in the learning ability of the individuals [6]. Furthermore, we model the natural maturation in cognitive or non-cognitive measure k as a function of background variables

[5]We also test the linearity of the relationship depicted in Figure 2.1 against the alternative hypothesis of non-parametric specification of age on TRLT score based on the Yatchew's semi-parametric regression method (Yatchew, 2003). Yatchew's approach exploits the assumption that nonparametric component of the partial linear models have a smooth functional form. As the dependent variable values become close, differencing tends to remove the nonparametric effect, eventually allowing the application to estimate the residual variance. The test statistics is based on the difference between the sum squared residuals for the linear and non-parametric model specification. Out examination of the prior-to-school data by Yatchew's specification test produces test statistics (V = 0.133 with P>|V| = 0.447) that fails to reject the null hypothesis of linear specification.

[6]Our model specification is an example of cumulative value-added production function. Estimation results under alternative specification where the lagged value of the outcome variable is excluded are presented in Table B.2, B.3, B.4. Comparison of the estimated effects of young, unready and their interaction term reveals that our model specification is robust since estimated effects are not different than of the alternative specification.

Z_i. These background variables include a set of child's characters; as gender, verbal ability, prosocial and externalizing behavior scores at age 4 and inhibitory control, and household characteristics such mother's verbal ability, parents completed education, working status, sibship size, socio-economic status, home environment and measurement gap[7].

Our model investigates the heterogeneity in $\theta^k_{i,s}$ and in particular the impact of starting school young, and of being rated unready for school. Therefore, we specify:

$$\theta^k_{i,s} = \theta^k_{0,s} + \theta^k_{1,s} YNG_i + \theta^k_{2,s} UNREADY_i + \theta^k_{3,s} YNG_i \times UNREADY_i$$

where YNG_i=1 if the child start school young and 0 otherwise: and $UNREADY_i$ =1 if the child was ready for school at one year prior to school start age and 0 otherwise.

Putting this all together the empirical model is:

$$\begin{aligned} \triangle y^k_{is} = Z_i\gamma^k + \theta^k_{0,s} + \theta^k_{1,s} YNG_i + \theta^k_{2,s} UNREADY_i \\ + \theta^k_{3,s} YNG_i \times UNREADY_i + \beta_s y^k_{i,s-1} + e_{s,i} \end{aligned} \tag{2.1}$$

Although the national cutoff rule is widely enforced, the decision of when children are send to school is primarily taken by the parents. In fact some children start school early and some are held back. In Table 2.1, we provide evidence that deviations from the official rule differ by maternal perception of children's readiness for school, indicating the entrance age may be endogenous. However, to be correlated with $e_{s,i}$ the parents must anticipate future shocks to specific

[7]There are differences in the timing of measurement and hence the gap between two measurement. However descriptive statistics in Table 2.4 shows that these differences are not significantly different across children who enter school young and old or rated to be ready and unready.

developmental trajectories when they make the decision. The greater concern is that it is rather correlated with variation in the natural maturation (α_i^j) that are not captured by the observable background variables, Z_i. Any unobserved parent and child characteristic that could influence child's growth of ability measure k would also influence the parental decision of school start age. For example, low ability students may be more likely to both delay entry and perform worse in school, in which case OLS estimates would find children who delayed entry to be performing worse than other children, causing a downward bias in estimation of true effect. Moreover, parents who are involved in their children's education may be more likely to both delay entrance and substitute schooling with other investments in their child that improve school performance. Estimation of entrance age wtihout accounting for the unobserved factor would lead an upward bias in the OLS estimation. Hence, we estimate the models both by Ordinary Least Squares (OLS) regression and by Instrumental Variable (IV) Estimation method where actual school starting age is instrumented by the regulated school starting age.

2.3 Data

The data comes from the Study of Early Childhood Ecologies in Turkey (ECDET). ECDET is a longitudinal survey that studies children's developmental trajectories in early childhood, and identifies the social and environmental factors that influence those trajectories. Children were surveyed annually starting from 36-42 months of age in 2008 throughout the following five years. The survey examined a nationally representative sample of approximately one thousand children and their families from nineteen different provinces in Turkey. Data collection efforts were carried by the group of researchers visiting participants in their place of residency during summer months, June to October, to minimize the probability of attrition due to schooling.

We focus on children who had finished at least the first year of formal schooling at the wave five interview. Our sample includes children who were born between January 2004 and October 2005, who started primary school in academic years 2009-10, 2010-11 or 2011-12. Hence, we use outcomes measures of cognitive and non-cognitive skills that are collected in the third, fourth and fifth wave and other detailed background information on household and family characteristics available in the dataset. For children who have completed second grade by the time of wave five data collection, we use measures of ability and behaviour scores from (i) wave three, measured in summer before they started school, grade 0, (ii) wave four, end of grade 1 and beginning of grade 2 and (iii) wave five, end of grade 2. Hence differences of ability measures between wave three to four is included as change in score in grade 1 and wave four to five as change in score in grade 2.

Our sample covers two academic cohorts of primary school children with 767 observations in total. This is due to the range of participants' birth year. Figure 2.2 provides and overview of the each birth cohorts academic progression over the timeline of data collection. Majority of the children (63%) in the sample started primary education after wave 4 visit, in academic year 2010-2011, whereas the rest (37%) had started school the previous academic year[8]. Despite the majority of the children complying with the cutoff rule of school starting age, the implementation of the rule is imperfect.1.7 percent of the children in the sample started the primary school education one year earlier then they were eligible while 6.1 percent delayed the entrance by one year. Although non-compliance with the cut off rule is the primary factor determining the relative age of entrance, we do not solely rely on this information. We, utilize the observed month of age by first grade entrance to

[8]Enrollment to pre-school day care centers and childcare programs is very rare in Turkey. The rate of 3-5 year olds enrolled in any pre-school program is 27 percent in 2010 , being the lowest participation rate among all OECD countries whose average is 80 percent (OECD, 2015). Our sample characteristics is coherent with the statistics provided by the OECD. The number of children who are enrolled in a pre-primary education institute is 42 (5%) at age 3. The percentage of children enrolled in pre-primary education increases to 36 at age 5. An important factor that might explain the low participation rate in Turkey is the parental preference towards a strong substitute of informal childcare provided by either mother or a non-working elderly in the family.

identify the relatively old entrants within the corresponding cohort.

2.3.1 Dependent Variables

The main learning outcome variable of interest in the analysis is the children's verbal ability, measured by the Turkish Receptive Language Test (TRLT). TRLT is a test that was originally constructed by Berument and Guven (2010) and adopted in our analysis to estimate the vocabulary knowledge of children who are aged between 3-7.

During the test, each participating child is asked to choose the picture, among 4 pictures shown, which represents the word said out loud by the interviewer. The test resembles the widely used Peabody Picture Vocabulary Test, PPVT (Dunn & Dunn, 1981). TRLT is an adaptive test. The item list used in the ECDET survey varies with the age of the children and consists of 159 items in total including two practice questions. The test is started with the items from child's corresponding age level and is proceeded to higher age level items if the child's answers to two thirds of the items of that level were correct. The TRLT is conducted in each wave, enabling us to track the progress in vocabulary knowledge across waves.

The responses to the items in TRLT is used to estimate an overall score for verbal ability. The estimation is done through a three-parameter logistic item response theory (IRT) model to obtain a more accurate estimation of the latent vocabulary ability. The IRT process is a methodology that utilizes the assumption that a subject's probability of answering an item correctly depends on his or her ability as well as item's characteristics. A three parameter IRT logistic model (3PL-IRT) extracts a child's vocabulary ability on a continuous ability scale by making use of the dichotomously scored item responses. It follows a two step estimation method in which the first step is to estimate each item's difficulty, discriminating ability and guessing characteristics by using the responses of entire sample. The

estimated item characteristics are then used for estimating the child's latent verbal ability (Andreassen and Fletcher, 2007). The method enables us to calibrate the item characteristics on the same scale even if the assessment of the test is different for each wave. Similar procedures are commonly used for scoring adaptive tests. The receptive vocabulary ability scores are standardized across all the individual-time observations[9].

A second set of outcome variables in our analysis covers the desirable and socially accepted behaviors, namely prosocial behavior, and problem behaviors that are often disruptive to the child's environment called externalizing behaviors. Information on children's social behavior is collected in all 5 waves by maternal assessment on the Adaptive Social Behavior Inventory(ASBI; Hogan et al., 1992) and Eyberg Child Behavior Inventory (ECBI; Eyberg and Robinson, 1983). The ASBI is designed to provide a measure of children's social competence in early childhood, especially the preschool period. The original Inventory is a 30-item questionnaire. The mothers are asked to rate their children on a 3-point Likert scale with respect to frequency of the behavior. ASBI yields validated three sub-scales: express (e.g., "Understands others' feelings, like when they are happy, sad or mad."), comply (e.g., "Is helpful to other children."), and disrupt (e.g., "Is bossy, needs to have his/her way."). The inventory was modified by the ECDET team (Baydar et al., 2010). The version adopted in the ECDET survey keeps the context of the original 30 items and three sub-scales. However, the response format is changed into a 5-point Likert scale with the intention to make the instrument consistent with other scales in the survey. Further, some expressions in the items are slightly altered when the inventory is translated into Turkish.

The ECBI is widely used to measure negative social behavior in children between the ages of 2 and 17. The scale consists of 36 items that are rated by the mother on two different aspects; frequency and whether the behavior is a problem. The

[9]Estimated latent verbal scores using IRT is provided as part of the data by the survey team instead of the responses to each item.

inventory was adapted by Baydar and colleagues (2010) for the ECDET survey. The adopted version also includes 36 items and maintains the original structure. But as with the modified ASBI, the frequencies of behaviors are rated on a 5-point Likert scales instead of 7-point scales. The items allow for estimation of total behavior problem intensity as well as 3 sub-scales: aggression intensity (e.g., "Fights with peers"), demand for attention intensity (e.g., "Whines"), and conduct problems intensity (e.g., "Argues with parents about rules.")[10].

We use the responses from the ASBI and ECBI to construct relevant behavior scores using factor score models. The explanatory factor analysis, along with the findings of the previous studies on ASBI (Hogan et al., 1992; Taylor et al., 2004) provide evidence that there is a single factor when the express and comply sub-scales are combined. We refer to this factor as prosocial behavior and use the information available by merging the responses from these two subscales to obtain the aggregate behavioral score. A similar analysis for ECBI items suggests that the entire response set gives a total negative social behavior; which we refer to as externalizing behavior.

We are able to follow the children and responses obtained from their mother for each inventory across the five waves. We use the information available from the time and child pairs to construct a latent score from a continuous scale. Yet, a structural formation to account for within subject dependencies in the measurement is important in the estimation process. So we employ a multilevel structured confirmatory factor analysis (Jeon and Rabe-Hesketh, 2012), which is random effects model. The following is the representation of our multilevel version of the latent variable model for child i in wave t:

[10]We only include the aggregate behavior score in our analysis.

$$logit[Pr(Y_{jti} = k|\theta_{ti})] = \beta_j + \theta_{ti}\alpha_j \tag{2.1}$$
$$i = 1, 2, ..., N; t = 1, 2, ..., Z;\ j = 1, 2, ..., K$$

$$\theta_{ti} = \delta_{ti}^{(2)} + \delta_i^{(3)}$$

where probability of rating the corresponding item j on a 5-point scale is modeled through an ordinal logistic function. As in an IRT model, the estimation of the latent ability, θ_{ti}, is based on the set of observed responses Y_{jti}. Moreover, ${\delta^{(2)}}_{ti} \sim N(0, \sigma_t^2)$ captures the time-child level random effect and ${\delta^{(3)}}_i \sim N(0, \sigma_i^2)$ is a child-level random effect. The latter represents the child's mean ability, whereas the former represents the deviation of the child's ability at time t, from the child's mean ability level. Substituting the model for θ_{ti} into the model for y_{jti}, we have three-level ordered logistic random effect model with factor structures. Furthermore, the factor loadings (α) are assumed to be the same at levels 2 and 3.

$$logit[Pr(Y_{ji} = k|\theta_i] = \beta_j + \delta_{ti}^{(2)}\alpha_j + \delta_i^{(3)}\alpha_j$$

We follow our explained measurement model for the underlying latent variables and extract the behavior scores from a continuous scale for all 5 wave collection. For practical purposes, we reversed estimated externalizing score to obtain a positive growth in development trajectories. The scores obtained are standardized with mean zero and standard deviation 1.

2.3.2 Relative Age and Readiness

The key explanatory variable in our analysis is the indicator variable of child's relative age. We assign each child in our sample to relevant states of old and young according to their school entrance age in months. Children were expected to start school when they were 6 years old in the calendar year[11]. In Turkey, academic year starts in mid-September and the cutoff date for school entrance eligibility is at the beginning of the calendar year. If the cutoff rule is followed, the expected age range of the entrants varies from 80 months to 69 months. Children born in December start primary education a year earlier than children born in January. However, It is not unusual to delay entrance to school in the rural parts of the country and for the children that are born later in the year. In fact, we observe late and early starters in our sample, from 86 months of age to 50 months of age. When constructing the variable, we assign children in our sample in two categories, "young" if they are at most 74 months old at school entry or "old" otherwise. Young children are either born in the second half of the year or start the first grade earlier than they are eligible according to the rule, whereas the opposite is true for old starters. Figure 2.3 displays the distribution of the age in months by the primary school entrance for our sample.

Table 2.2 presents the compliance rates to school entrance cutoff rule for our sample. Compliance with the school starting rule is high for the children who are born in the first half of the year, and irregular enrollments are more evident for the second half. This accumulation of the non-compliers predominantly in the group of "expected young" students indicate that parental concerns regarding school adaptation and performance is consistent with families with relatively young children. The high non-compliance rate could also be a result of school administrations and local authorities' inability to enforce the cutoff rule.

[11] A recent education reform was brought forward in 2012 which increased the compulsory schooling period to 12 years and introduced an earlier entrance age, commonly referred to as "4+4+4" system. Children in our sample was not affected by this new entrance age rule. Majority of the sample has started school in the academic year of 2011-2012.

Second key explanatory variable in our analysis is the child's readiness for school according to maternal assessment when child is at age five. This measure was developed by Baydar et al. (2010) and originally consisted of 106 questions with 7 sub-measures. It was modified to a shortlist of 15 items for the ECDET survey. The items were asked to mothers both in wave 2 and 3 to obtain the age 5 measurement of all children in the sample that have different dates of birth. We include one school readiness score in our analysis, where our criteria in choosing is based on proximity in month of measurement to the 5th birthday. Furthermore, the scores are age standardized to remove any remaining age pattern. The shortlist includes questions such as "my child is able to write the first letter of his/her name, my child is curious about the content of books" that are designed to identify the behavioral and motor skills as well as capacities of the child in terms of reading, writing and learning. The items were rated by mothers on a 5-point Likert scale. Explanatory factor analysis of the inventory displays a high internal reliability and suggests a unidimensional latent trait of school readiness. We detect 13 items from the short list which have factor loadings greater than 0.5 and we include only those items in the measurement model. We constructed a readiness measure through a multilevel confirmatory factor analysis, similar to the model described to estimate latent social behavior skills .

Although, the estimated latent score for school readiness is from a continuous scale, we incorporate this measure in the analysis through an indicator variable. The threshold level (School Readiness Score=0) for the categorization is chosen with reference to the measure's sample distribution, presented in graph 2.4. Children with raw school readiness scores less than 0 are classified as unready. Table 2.3 also presents summary statistics for the indicator variable of school readiness by other background variables. The raw school readiness score exhibits a difference more than 2 standard deviation between children who are unready for school (m=-1.00 , sd=0.62) and ready for school (m=1.33 , sd=0.95).

2.4 Results

We first provide descriptive statistics of our variables for the full sample and by school entry age status and readiness for school in Table 2.4. There are no differences in most of the children and home characteristics between young/old entrants and ready/unready students. Children who start school early are more likely to have working mothers with less completed years of education, compared to children who start school later. They are also more likely to attend pre-school programs at age five. These differences in mother's characteristics by school start age confirm that the entry age is unlikely to be exogenous. Further, differences in outcome variables across old and young entrants these groups are evident in the data. Children who start school old have no differences in their verbal ability compare to young entrants at the beginning of year one, but they are more likely to have higher verbal ability scores at the end of year one. The differences in the means of school entry verbal scores and the first year verbal scores is 0.4 which corresponds to two thirds of the standard deviation of the verbal scores at school entry. The variables for changes in the scores in first and second grades are rescaled to standard deviation of scores at school entry, baseline year, to accommodate comparable interpretation of estimated effects.

2.4.1 Verbal Ability

We begin the analysis with estimates of the relative age at school entry and school readiness on verbal ability growth. Table 2.5 reports the estimation results for model specification in section 2 for changes in TRTL scores during first and second grade. The intercept of the model captures the increase in verbal ability for children who are old at entry and rated ready for school. We choose this group to represent the baseline to accommodate easy interpretation for coefficients on the indicator variables. The coefficients on young and unready are evaluated as

penalties associated with the deviations from the baseline group . The coefficient on the interaction term indicates whether the estimated penalty of age at entry differs across students who are ready and not ready for school and provides an interpretation on the combined effect of entry age and school readiness. We interpret the estimated effect of the interaction term as dif-in-dif effect in our framework of two treatments; age at entry and readiness for school.

A comparison of the results of OLS and IV regressions reveals similar estimated effects, in Table 2.5. This suggest that our specification under OLS estimation method does not suffer from endogeneity arising from selection into school start age groups[12]. We also report the first stage regression results for verbal ability in Table B.1. The coefficients of the instruments on actual levels are significant and considerably high; 0.88 with a standard error of 0.02 for expected young (column 1) and 0.81 with a standard error of 0.04 for expected young and ready (for column 2). Moreover, The test statistics from a LM test of whether the equation is identified, i.e., that the excluded instruments are correlated with the endogenous regressors, indicates that the matrix is full column rank, and the model is identified. The weak instrument is clearly not a concern in our specification. Hence, we refer to results from the IV estimation method for discussing the effects in the rest of this section.

The first grade results suggest that there are associated penalties for starting school young or unready as the estimated coefficients are negative and statistically significant. Students who are young will have smaller growth in their verbal ability in first grade, by almost 0.67 standard deviation of baseline verbal ability score, compare to their old peers. However, the positive and significant interaction term

[12]Interaction term in our specification introduces a second first-stage estimation where the interaction of the exogenous variable and the instrument (expected young) is included as the second instrumental variable. Hausman specification test for endogenity where estimation from the OLS model including the predicted residuals from first stage estimations as covariates to jointly test their significance provides evidence for exogeneity of young and young*unready . F-statistics for verbal ability is 0.20 (Prob > F = 0.8213), for prosocial behavior is 1.11 (Prob > F = 0. 3301), and for externalizing behaviour is 0.79(Prob > F = 0.4710).

indicates that the penalty for being both young and unready is smaller than the sum of the two separate penalties. Second grade results reported in column 3 and 4 do not provide any evidence for the effect of relative age of entry. The estimated coefficients for being young and its interaction with unready are negative but not statistically significant. However, these results are based on the sample of students who completed the second grade by the time of the wave 5 visits. The sample size is considerably smaller than the first grade sample. Therefore, the observed variation in the entrance age attained through variation in month of birth is smaller, resulting in loss of precision in our estimation results. We acknowlegde that the sample of analysis in our study is relatively small, the estimated effects are in line with the findings of the literature. A similar study that explores the learning accumulation in the US finds higher growth rates for older entrants in the 1st year and no effect of age of entrance in the 2nd year (Lubotsky and Kaestner, 2016).

We find statistically significant positive effect of the TRLT level at age 4 on the verbal ability gains in first grade of primary school. This estimate implies that the accumulation of cognitive abilities in early years of schooling is highly dependent on the initial ability levels in child's early life course. Although no causal link between pre-school period skills and ability growth is being proposed in our framework, the significant coefficient support the evidence that early levels of skills generate higher growth. Moreover, skills levels in early childhood may also capture differences in family characteristics and practices that are influential in development of skills.

The estimation results indicate a strong negative effect of TRLT level at primary school entrance. Students with higher level of TRLT at primary school entrance experience a slower progress throughout the first year, by 1 standard deviation, suggesting a mean-reversion effect of education.

In our model, a student belongs to one of the four categories based on their

age at entry and readiness for school; (1) old and ready, (2) old and unready, (3) young and ready, (4) young and unready. Table 2.6 presents the associated increases in student's verbal ability in first and second grade if the student is in that respective category. These effects are simply calculated from the estimated coefficients from IV regressions in Table2.5 in column 2 and 4. We rely on these reported effects to compare the group differences. There is a clear monotonic decrease across groups. Students who are old and ready achieves the highest acquisition in grade 1, by 1.3 standard deviation, whereas the lowest progress is obtained by the children who are young and unready, by 0.5 standard deviation. Moreover, the bottom half of the table reports the p-values of the tests between the estimated effects. The estimated positive difference between old entrants and young entrants is statistically significant both for ready and unready students, first and second row in p-value section. This might suggest that delaying entrance for both categories has important advantages in learning in the first grade. On the other hand, there is no significant difference between ready and unready students if they are the youngest in their cohorts. The disadvantage of not being ready for school could partially be alleviated by being relatively old within the cohort. Yet, the benefit of delayed entrance is the highest for children who are ready for school, due to longer time spent in maturation process.

2.4.2 Social Behavior Skills

Our analysis of the social skill development includes two behavioral skills as outcome variables; prosocial and externalizing behaviors. Table 2.7 shows the estimates of the effect of school starting age for changes in prosocial behavior. Unlike results for verbal score changes, OLS and IV estimation results in first grade produce non-significant coefficients for indicator variables, young and unready. However, the estimated coefficient for interaction term is negative both in OLS and IV results, but statistically significant at 10% level only in IV results. This

indicates that there is slower growth in prosocial behaviors in first grade in school only for students who are both young at entry and unready for school. A comparison of the growth rates in prosocial behavior in Table 2.8 clearly exhibits that the group effects are similar except for young and unready. The increase in the prosocial skills in grade 1 is between 0.8 and 0.9 standard deviation for the first 3 groups, but smaller for young and unready students, by 0.65 standard deviation compare to young and ready. The estimation results do not provide enough evidence to conclude the existence of an effect from school starting age. Similarly, second grade estimation results for estimated effects on growth rate and the differences across groups are not statistically significant.

Table 2.9 and 2.10 reports the estimated effects for analysis of the externalizing behavior outcomes. Note that differencing the reversed scores gives the reduction in externalizing behavior over the course of the corresponding grade. Our objective of the reverse coding the item responses in ECBI is to get scores that provide parallel interpretation for our estimated effect that complies with the previous analysis. Thus, a negative coefficient obtained from the estimation results reflects an increase in the externalizing behavior, whereas a positive coefficient implies an improvement in child's behavioral skills. The estimation results reported in table 2.9 shows that being relatively young at school entry has a negative effect on the rate of reduction in externalizing behavior in first grade. Moreover, this negative effect is larger for students who are ready. Table 2.10 highlights the group differences across old and young students. The rate of change in externalizing behavior scores is significantly slower by 0.6 standard deviation for ready students and by 0.18 standard deviation for unready students. Children rated to be ready for school obtains higher benefit in reducing their negative social behavior if they start school old. This effect is more profound in the second grade results. The estimated age effect is 1.7 standard deviation for children rated to be ready. We find no evidence for the effect of school start age for children who are unready for school.

2.4.3 Gender Heterogeneity

We conduct the analysis for school entry age effects separately by gender. Our aim is to examine whether any practice of of early entrance that leads to being young within the cohort has any differential penalties for students from different gender groups. Although descriptive statistics reported in Table 2.4 suggests that no gender differences in proportion of young and old entrants, girls are much more reported to be ready than unready. We have reasons to suspect that meaningful gender differential effects would be present for verbal ability, but not for non-cognitive skills, from the estimated coefficients of gender in Tables 2.5, 2.7 and 2.9.

Table 2.11 reports the results of subsample analysis for verbal ability and skill acquisition only in first grade as the sample size for second grade outcomes is substantially small to allow us to run these analysis.The first noticeable pattern is that the estimated coefficients for young is consistent with the results observed for the full sample analysis. The effect of being young at entry on the verbal ability acquisition is negative for both gender and significantly higher for girls. Unlike the full sample results, there is no effect associated with being unready for school in subsample of boys. We further examine the estimated group effects presented in Table 2.12. Boys who are old at school entry have faster growth in verbal ability compare to young entrants in first grade with no differential effect of readiness for school. Similarly, girls who are old at entry and ready for school have the highest acquisition and their growth rate in verbal ability is significantly different from other groups. The estimated growth rates in other groups; (1) old and unready, (2) young and ready and (3) young and unready, show no significant differences.

Subsample analysis for prosocial behavior differences in first grade suggests that there is no significant effect for school starting age for either gender. Boys don't

experience any significant changes in their prosocial behavior regardless of their school start age and readiness for school, the differences across these four groups are not statistically significant. Although girls have positive growth rates in their prosocial behavior, only significant difference in their growth rates is between young entrants who are ready and unready. This evidence is not sufficient to conclude that there is differences in the effect of entry age between boys and girls in first year prosocial behavior acquisition. The estimated effects of school start age and readiness for school in externalizing behavior, reported in column 5 and 6 of Table 2.11 shows no differences for boys and girls.

Boys who delay entry to school have higher growth rates of verbal ability in first grade. They also have higher reduction in their negative behaviors if rated ready for school, while no effect on the prosocial behavior. Delaying entry has similar effects for girls. They obtain higher growth rates in verbal ability if ready. There is no effect of school start age on prosocial behavior. Delaying entry increases the rate of reduction in negative behavior both for ready and not ready, but the benefit is higher for girls who are rated be ready for school.

2.5 Conclusion

In this paper, we studied the effect of school starting age on the cognitive and non-cognitive skills development through the first two years of primary school in Turkey. To our knowledge, this is the first study of the effects of school starting age on educational progress in Turkey. We use the sharp date of birth cutoff for school entrance as source of exogenous variation in school starting ages. We exploit the fact that the natural age maturation process for verbal ability is linear – which we document in our data - to break the identity between school starting age, time in school and age at testing. This linearity means that we can isolate an effect of school starting age on growth in verbal ability through the school year.

An important and unique feature of our data is the availability of sophisticated measure of school readiness. This allowed us to investigate whether chronological age per se matters, or just school effects. The latter is an exploration of heterogeneity: whether, for example, the effect of school starting age on subsequent acquisition of verbal ability varies with measured school readiness.

Our findings are that students starting school young make less progress in verbal ability through the first year of primary school, as do students with lower measured school readiness. Student who are both chronologically young and with lower measured readiness make the least progress of all. The profound positive effect of late entrance on the cognitive abilities in first grade disappears in the second grade. Yet, the results provide a different explanation for social behavioral skill acquisition. We do not find evidence for social behavioral skills development in first grade, neither for gains in prosocial behavior nor in reduction of externalizing behavior. However, second grade results that a significant rate of reduction in externalizing behavior is expected for students who are old and rated to be ready for school.

This means that the move to earlier school starting ages may have some negative effects on learning outcomes, and parents are right to be concerned with this. However, the question of how persistent these effects in later grades remain to be at interest as our findings suggest that the educational input might resolve differences at rates of learning after first grade.

Figures

Figure 2.1: Pre-school maturation process

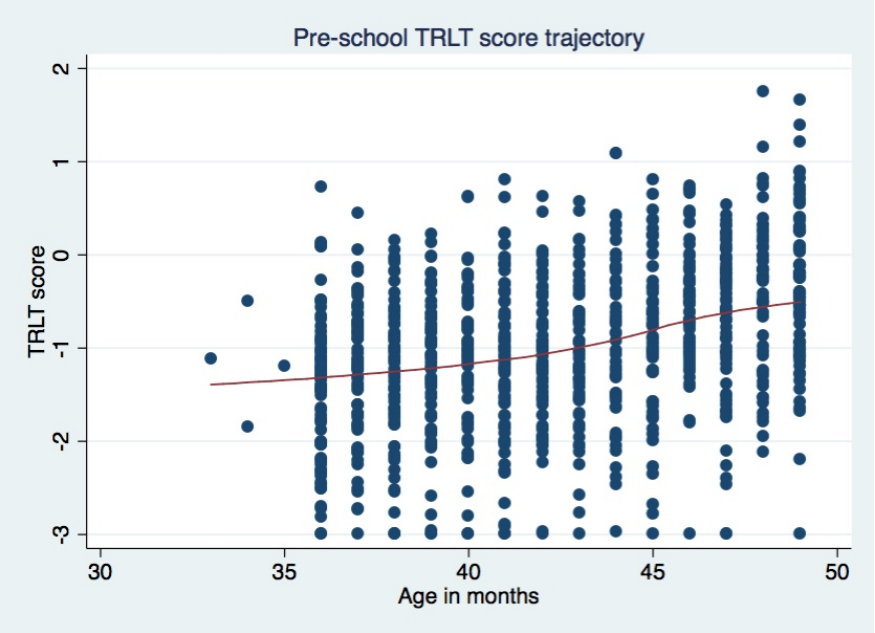


Figure 2.2: Timeline describing the cohorts and data collection

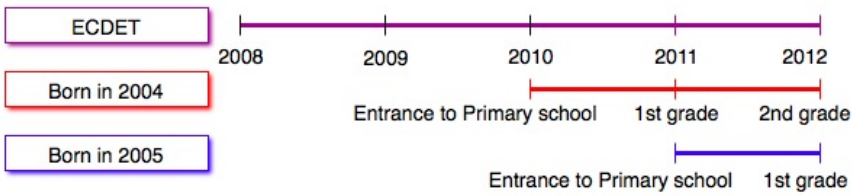


Figure 2.3: Distribution of school starting age

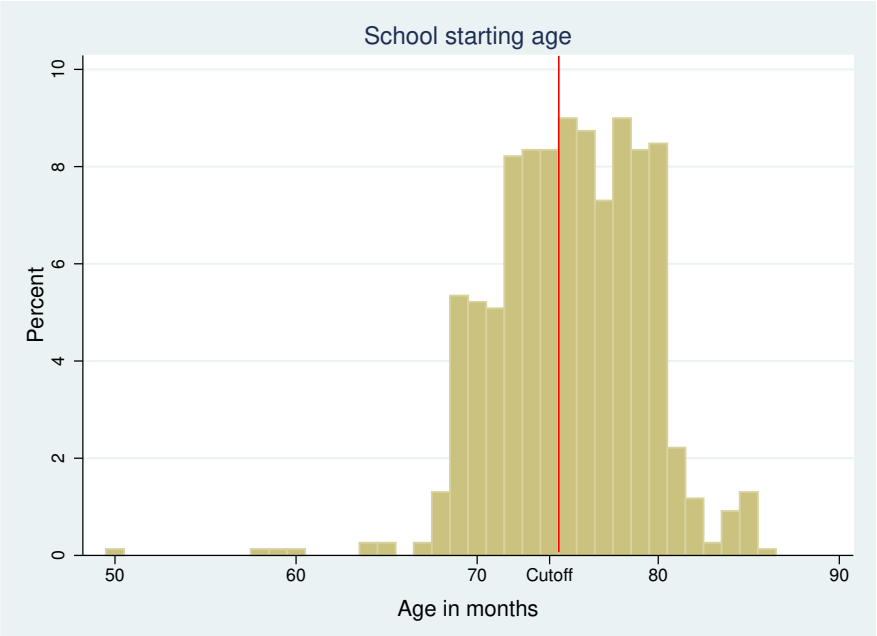
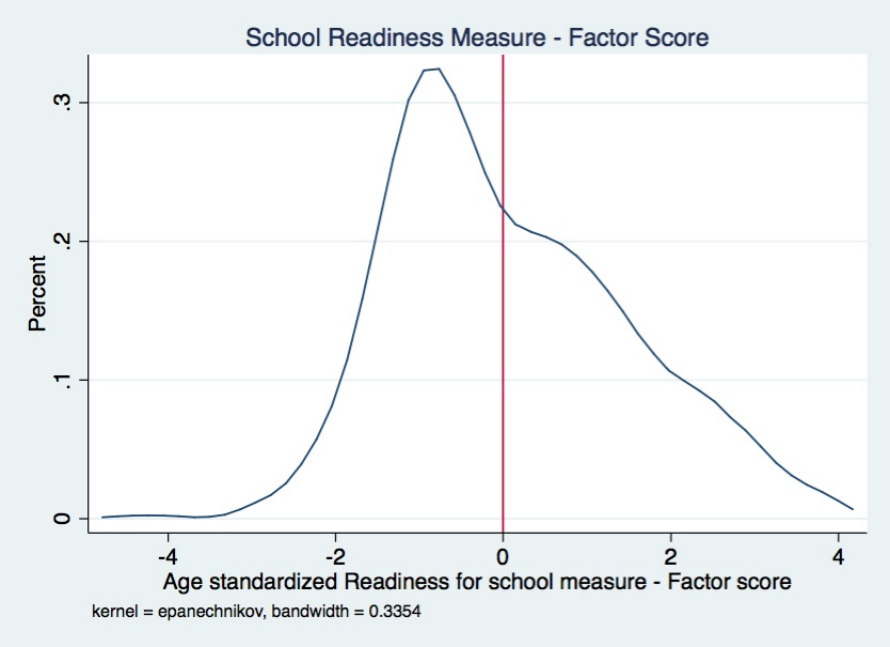


Figure 2.4: Distribution of school readiness raw score



Tables

Table 2.1: Compliance Rates by Readiness for School and Gender - Row Percentages

	Early starter	On time	Late starter
Unready for school (N=412)	1.5	91.0	7.5
Ready for school (N=355)	2.0	93.5	4.5
Female (N=348)	2.6	90.5	6.9
Male (N=419)	0.9	93.6	5.5

Table 2.2: Compliance Rates by Month of Birth

	Early	On time	Late	N
January	3.2	96.8	0.0	93
February	3.6	96.4	0.0	56
March	0.0	100	0.0	61
April	3.1	96.9	0.0	64
May	2.8	97.2	0.0	71
June	0.0	98.5	1.5	67
July	1.4	97.2	1.4	72
August	0.0	86.7	13.4	75
September	1.7	88.1	10.2	59
October	2.1	89.1	8.5	47
November	2.0	78.4	19.6	51
December	0.0	70.6	29.4	51
Total Sample	1.7	92.2	6.1	767

Table 2.3: Summary Statistics for School Readiness

	Unready for school (N=412)	Ready for school (N=355)	Pearson's X^2
School Readiness Score (RFS)			
Mean	-1.00	1.33	
Standard deviation	0.62	0.95	
Gender			10.17
Female			
N	165	183	
Row %	52.6	47.4	
Male			
N	247	172	
Row %	58.9	41.1	
Primary Language			8.74
Turkish			
N	401	354	
Row %	53.1	46.9	
Kurdish			
N	10	0	
Row %	100.0	0.0	
Arabic			
N	1	1	
Row %	50.0	50.0	
Pre-primary education			11.86
Not enrolled at age 5			
N	285	203	
Row %	58.4	41.6	
Enrolled at age 5			
N	125	150	
Row %	45.5	54.5	

Table 2.4: Summary Statistics, by School Start Age and Readiness

	Full Sample	**Old**	**Young**	**Ready**	**Unready**
School starting age in months	75.18(4.10)	78.03(2.36)	71.42(2.59)	74.64(4.15)	75.48(4.04)
Unready for school (0,1)	0.64(0.48)	0.67(0.47)	0.60(0.49)		
Relatively young (0,1)	0.43(0.50)			0.48(0.50)	0.40(0.49)
Test and Behaviour Scores:					
TRLT score at age 4	0.17(0.83)	0.15(0.82)	0.20(0.85	0.64(0.73)	-0.08(0.77)
TRLT score (grade0)	0.40(0.67)	0.41(0.59)	0.39(0.77)	0.75(0.60)	0.20(0.63)
TRLT score (grade 1)	0.80(0.73)	0.93(0.77)	0.62(0.63)	1.08(0.58)	0.64(0.75)
Δ score in grade 1	0.59(1.03)	0.79(1.05)	0.33(0.94)	0.51(1.01)	0.64(1.04)
Δ score in grade 2	0.25(0.81)	0.28(0.80)	0.25(0.81)	0.23(0.91)	0.26(0.72)
Prosocial behaviour at age 4	-0.03(1.02)	-0.07(1.02)	0.03(1.01)	0.31(0.95)	-0.21(1.00)
Prosocial behaviour (grade 0)	0.22(0.99)	0.23(1.01)	0.21(0.97)	0.59(0.98)	0.01(0.94)
Prosocial behaviour (grade 1)	0.27(0.93)	0.29(0.95)	0.25(0.92)	0.55(0.89)	0.12(0.92)
Δ behaviour in grade 1	0.05(1.02)	0.06(1.03)	0.04(1.00)	-0.05(0.99)	0.11(1.02)
Δ behaviour in grade 2	0.01(0.99)	-0.10(1.28)	0.01(0.98)	-0.08(0.93)	0.07(1.03)
Externalizing behaviour at age 4	-0.12(1.01)	-0.15(1.02)	-0.08(0.99)	0.07(0.95)	-0.22(1.02)
Externalizing behaviour (grade0)	0.08(0.94)	0.06(0.93)	0.10(0.96)	0.19(0.88)	0.01(0.97)
Externalizing behaviour (grade1)	0.24(1.00)	0.38(1.01)	0.05(0.97)	0.33(1.01)	0.19(1.00)
Δ behaviour in grade 1	0.17(1.05)	0.34(1.05)	-0.05(1.00)	0.15(1.06)	0.19(1.04)
Δ behaviour in grade 2	0.50(1.00)	1.01(0.93)	0.47(1.00)	0.71(0.99)	0.35(0.99)
Controls:					
Female (0,1)	0.45(0.50)	0.46(0.50)	0.44(0.50)	0.53(0.50)	0.41(0.49)
Household size	5.07(1.92)	5.08(1.92)	5.07(1.92)	4.51(1.64)	5.39(1.99)
Urban (0,1)	0.52(0.50)	0.53(0.50)	0.50(0.50)	0.63(0.48)	0.46(0.50)
Number of kids in the household	2.26(0.98)	2.24(0.99)	2.29(0.95)	1.93(0.84)	2.45(1.00)
Birth Order	1.98(0.98)	1.94(0.98)	2.03(0.98)	1.73(0.84)	2.12(1.03)
Years of education - Mother	5.91(3.48)	6.13(3.44)	5.62(3.53)	7.63(3.72)	4.95(2.94)
Years of education - Father	7.40(3.24)	7.34(3.18)	7.47(3.32)	8.49(3.41)	6.78(2.97)
Mother Working at wave 5	0.17(0.38)	0.14(0.35)	0.21(0.41)	0.23(0.42)	0.14(0.35)
Father Working at wave 5	0.94(0.23)	0.94(0.24)	0.95(0.22)	0.96(0.19)	0.93(0.25)
Socio-economic level	-0.01(0.97)	0.01(0.98)	-0.05(0.95)	0.53(1.04)	-0.33(0.76)
Pre-primary education - age3	0.01(0.11)	0.01(0.12)	0.01(0.09)	0.03(0.16)	0.00(0.06)
Pre-primary education - age4	0.05(0.23)	0.05(0.21)	0.06(0.24)	0.09(0.29)	0.03(0.18)
Pre-primary education - age5	0.36(0.48)	0.25(0.43)	0.51(0.50)	0.47(0.50)	0.30(0.46)
Mother vocabulary knowledge	8.26(4.97)	8.20(5.01)	8.33(4.92)	10.61(5.23)	6.95(4.30)
Age 4 inhibitory control	10.11(7.82)	9.65(7.86)	10.70(7.74)	13.37(7.22)	8.28(7.55)
HOME Language Stimulation (age4)	77.84(22.53)	77.90(21.51)	77.75(23.83)	88.69(14.54)	71.83(23.89)
HOME Academic Stimulation (age4)	47.44(28.53)	46.60(27.70)	48.54(29.59)	62.01(27.20)	39.33(25.93)
HOME Learning Materials (age4)	48.95(32.07)	48.34(32.03)	49.76(32.15)	70.28(26.20)	37.10(28.74)
HOME Punishment (age4)	12.47(15.18)	12.46(15.68)	12.49(14.51)	10.20(13.35)	13.73(15.97)
HOME Responsivity (age4)	63.67(29.16)	65.49(28.39)	61.28(30.01)	74.01(23.63)	57.99(30.35)
Measurement gap (grade1)	11.97(1.65)	12.16(1.70)	11.72(1.55)	11.72(1.72)	12.10(1.59)
Measurement gap (grade2)	12.36(1.76)	13.27(1.16)	12.31(1.77)	12.24(1.72)	12.44(1.79)

Standard Errors are reported in parentheses. TRLT: Turkish Receptive Language Test Scores

Table 2.5: Verbal Ability Acquisition

	First Grade		Second Grade	
	OLS	IV	OLS	IV
Constant	1.237***	1.258***	1.739**	1.824***
	(0.38)	(0.37)	(0.66)	(0.61)
YOUNG	-0.682***	-0.669***	-0.112	-0.199
	(0.14)	(0.15)	(0.14)	(0.14)
UNREADY	-0.383***	-0.360***	0.114	-0.046
	(0.12)	(0.10)	(0.21)	(0.32)
YOUNG*UNREADY	0.343**	0.288**	-0.226	-0.054
	(0.14)	(0.13)	(0.27)	(0.36)
Gender - Girl	-0.124*	-0.122**	-0.069	-0.071
	(0.06)	(0.06)	(0.09)	(0.08)
Verbal Score - Age 4	0.224***	0.224***	0.210**	0.213***
	(0.06)	(0.06)	(0.09)	(0.08)
Prosocial Behavior - Age 4	-0.091**	-0.091**	-0.029	-0.027
	(0.04)	(0.04)	(0.05)	(0.04)
Externalizing Behavior - Age 4	-0.023	-0.024	-0.049	-0.048
	(0.05)	(0.04)	(0.04)	(0.04)
Verbal Score - School Entry	-1.022***	-1.025***		
	(0.10)	(0.10)		
Verbal Score - First Grade			-1.218***	-1.219***
			(0.14)	(0.13)
Family Controls	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓
N	665	665	247	247
R-Squared	0.34	0.34	0.56	0.56

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months.

Table 2.6: Verbal Ability Acquisition - Effect Size

	First Grade	Second Grade
Old & Ready	1.258***	1.824***
	(0.37)	(0.61)
Old & Unready	0.898**	1.778***
	(0.35)	(0.67)
Young & Ready	0.589*	1.625***
	(0.33)	(0.61)
Young & Unready	0.517*	1.525**
	(0.30)	(0.60)
P-value: Ready[Old=Young]	0.00	0.16
P-value: Unready[Old=Young]	0.01	0.39
P-value: Old[Ready=Unready]	0.00	0.89
P-value: Young[Ready=Unready]	0.56	0.30
P-value: Dif-in-Dif	0.03	0.88
N	665	247

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. P-values come from testing equality of coefficients.

Table 2.7: Prosocial Behavior

	First Grade		Second Grade	
	OLS	IV	OLS	IV
Constant	0.786	0.808	0.383	0.224
	(0.53)	(0.51)	(0.63)	(0.54)
YOUNG	0.032	0.079	0.120	0.119
	(0.09)	(0.08)	(0.16)	(0.22)
UNREADY	0.063	0.112	-0.485	-1.143
	(0.09)	(0.09)	(0.60)	(0.98)
YOUNG*UNREADY	-0.238	-0.352*	0.524	1.222
	(0.17)	(0.18)	(0.63)	(0.99)
Gender - Girl	0.140	0.146	0.127	0.105
	(0.09)	(0.09)	(0.12)	(0.11)
Verbal Score - Age 4	-0.038	-0.040	-0.140*	-0.148**
	(0.07)	(0.07)	(0.07)	(0.07)
Prosocial Behavior - Age 4	0.034	0.034	0.175*	0.182**
	(0.06)	(0.06)	(0.09)	(0.09)
Externalizing Behavior - Age 4	0.079**	0.079**	0.126	0.134**
	(0.04)	(0.03)	(0.07)	(0.07)
Prosocial - School Entry	-0.677***	-0.678***		
	(0.04)	(0.03)		
Prosocial - First Grade			-0.620***	-0.618***
			(0.08)	(0.07)
Family Controls	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓
N	698	698	261	261
R-Squared	0.37	0.37	0.36	0.35

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months.

Table 2.8: Prosocial Behavior - Effect Size

	First Grade	Second Grade
Old & Ready	0.808	0.224
	(0.51)	(0.54)
Old & Unready	0.920*	-0.919
	(0.51)	(1.05)
Young & Ready	0.887*	0.343
	(0.53)	(0.52)
Young & Unready	0.647	0.422
	(0.48)	(0.45)
P-value: Ready[Old=Young]	0.32	0.58
P-value: Unready[Old=Young]	0.07	0.20
P-value: Old[Ready=Unready]	0.22	0.24
P-value: Young[Ready=Unready]	0.07	0.50
P-value: Dif-in-Dif	0.06	0.22
N	698	261

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. P-values come from testing equality of coefficients.

Table 2.9: Externalizing Behavior

	First Grade		Second Grade	
	OLS	IV	OLS	IV
Constant	0.199	0.153	1.203	1.699**
	(0.40)	(0.37)	(0.77)	(0.84)
YOUNG	-0.596***	-0.591***	-0.715**	-1.160**
	(0.13)	(0.13)	(0.29)	(0.46)
UNREADY	-0.112	-0.142	-1.101**	-1.470**
	(0.12)	(0.12)	(0.52)	(0.69)
YOUNG*UNREADY	0.340**	0.416***	1.021*	1.432*
	(0.14)	(0.15)	(0.55)	(0.75)
Gender - Girl	0.128*	0.124*	0.118	0.120
	(0.07)	(0.07)	(0.11)	(0.11)
Verbal Score - Age 4	-0.019	-0.018	0.019	0.030
	(0.05)	(0.05)	(0.09)	(0.08)
Prosocial Behavior - Age 4	0.077**	0.076**	0.110	0.116*
	(0.04)	(0.03)	(0.07)	(0.07)
Externalizing Behavior - Age 4	0.165**	0.164***	0.317***	0.318***
	(0.06)	(0.06)	(0.06)	(0.06)
Externalizing - School Entry	-0.602***	-0.598***		
	(0.05)	(0.04)		
Externalizing - School Entry			-0.575***	-0.579***
			(0.08)	(0.07)
Family Controls	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓
N	698	698	262	262
R-Squared	0.31	0.31	0.34	0.34

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months.

Table 2.10: Externalizing Behavior - Effect Size

	First Grade	Second Grade
Old & Ready	0.153	1.699**
	(0.37)	(0.84)
Old & Unready	0.011	0.229
	(0.38)	(0.70)
Young & Ready	-0.438	0.539
	(0.38)	(0.59)
Young & Unready	-0.164	0.502
	(0.38)	(0.57)
P-value: Ready[Old=Young]	0.00	0.01
P-value: Unready[Old=Young]	0.12	0.61
P-value: Old[Ready=Unready]	0.24	0.03
P-value: Young[Ready=Unready]	0.00	0.75
P-value: Dif-in-Dif	0.01	0.06
N	698	262

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. P-values come from testing equality of coefficients.

Table 2.11: Subsample Analysis - by Gender

	Verbal		Prosocial		Externalizing	
	Boys	Girls	Boys	Girls	Boys	Girls
Constant	0.999**	1.532**	0.582	1.148	0.083	0.353
	(0.42)	(0.67)	(0.43)	(0.81)	(0.70)	(0.34)
YOUNG	-0.620***	-0.743***	0.040	0.117	-0.553***	-0.649***
	(0.14)	(0.19)	(0.08)	(0.14)	(0.17)	(0.12)
UNREADY	-0.197	-0.607***	0.152	0.050	-0.071	-0.179
	(0.16)	(0.23)	(0.11)	(0.16)	(0.10)	(0.18)
YOUNG*UNREADY	-0.051	0.785**	-0.298	-0.416	0.447*	0.345*
	(0.18)	(0.33)	(0.20)	(0.29)	(0.24)	(0.20)
Verbal Score - Age 4	0.192***	0.263***	-0.062	-0.015	0.008	-0.027
	(0.06)	(0.08)	(0.06)	(0.08)	(0.04)	(0.07)
Prosocial Behavior - Age 4	-0.083	-0.069**	0.041	0.031	0.151***	0.032
	(0.07)	(0.03)	(0.05)	(0.08)	(0.06)	(0.05)
Externalizing Behavior - Age 4	-0.040	-0.037	0.054	0.114***	0.130	0.201***
	(0.04)	(0.08)	(0.05)	(0.04)	(0.08)	(0.08)
Verbal Score - School Entry	-1.011***	-1.084***				
	(0.09)	(0.13)				
Prosocial - School Entry			-0.676***	-0.689***		
			(0.04)	(0.09)		
Externalizing - School Entry					-0.581***	-0.627***
					(0.06)	(0.06)
Family Controls	✓	✓	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓	✓	✓
YOUNG [Girls=Boys]	0.27		0.62		0.47	
UNREADY [Girls=Boys]	0.19		0.59		0.44	
Y*UNREADY [Girls=Boys]	0.05		0.73		0.73	
N	359	306	374	324	374	324
R-Squared	0.35	0.38	0.41	0.36	0.30	0.36

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months. P-values are from testing equality of coefficients for girls and boys.

Table 2.12: Subsample Analysis - Effect size - by Gender

	Verbal		Prosocial		Externalizing	
	Boys	Girls	Boys	Girls	Boys	Girls
Old & Ready	0.999**	1.532**	0.582	1.148	0.083	0.353
	(0.42)	(0.67)	(0.43)	(0.81)	(0.70)	(0.34)
Old & Unready	0.802**	0.924	0.734*	1.198	0.012	0.174
	(0.35)	(0.63)	(0.41)	(0.78)	(0.71)	(0.36)
Young & Ready	0.379	0.788	0.622	1.265	-0.470	-0.297
	(0.42)	(0.57)	(0.43)	(0.83)	(0.73)	(0.36)
Young & Unready	0.131	0.966	0.476	0.900	-0.095	-0.130
	(0.33)	(0.59)	(0.44)	(0.74)	(0.67)	(0.30)
Ready[Old=Young]	0.00	0.00	0.61	0.39	0.00	0.00
Unready[Old=Young]	0.00	0.89	0.14	0.21	0.58	0.06
Old[Ready=Unready]	0.21	0.01	0.15	0.75	0.50	0.33
Young[Ready=Unready]	0.15	0.25	0.32	0.09	0.05	0.28
Dif-in-Dif	0.78	0.02	0.14	0.16	0.06	0.09
N	359	306	374	324	374	324

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. P-values come from testing equality of coefficients.

Appendix

B.1 Additional Tables

Table B.1: Verbal Ability Acquisition - First Stage Estimation

	YOUNG	Y*UNREAD
Constant	0.215	0.280**
	(0.15)	(0.12)
UNREADY	0.022	0.039*
	(0.03)	(0.02)
Exp YOUNG	0.878***	-0.008
	(0.02)	(0.01)
Exp YOUNG*UNREADY	-0.081**	0.813***
	(0.04)	(0.04)
Underidentification Test Stat	11.45	
P-value	0.00	
N	665	665
R-Squared	0.77	0.79

Controls are included in the estimation, not reported. Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Underidentification test is a LM test of whether the equation is identified.

Table B.2: Verbal Ability Acquisition

	First Grade				Second Grade			
	OLS	OLS/spec2	IV	IV/spec2	OLS	OLS/spec2	IV	IV/spec2
Constant	1.237***	1.034*	1.258***	1.060**	1.739**	-0.043	1.824***	0.026
	(0.38)	(0.54)	(0.37)	(0.49)	(0.66)	(0.66)	(0.61)	(0.62)
YOUNG	-0.682***	-0.676***	-0.669***	-0.639***	-0.112	-0.075	-0.199	-0.081
	(0.14)	(0.14)	(0.15)	(0.15)	(0.14)	(0.22)	(0.14)	(0.18)
UNREADY	-0.383***	-0.373***	-0.360***	-0.329***	0.114	0.363	-0.046	0.689
	(0.12)	(0.11)	(0.10)	(0.09)	(0.21)	(0.36)	(0.32)	(0.54)
YOUNG*UNREADY	0.343**	0.490***	0.288**	0.388***	-0.226	-0.427	-0.054	-0.764
	(0.14)	(0.15)	(0.13)	(0.15)	(0.27)	(0.45)	(0.36)	(0.60)
Gender - Girl	-0.124*	-0.073	-0.122**	-0.068	-0.069	0.011	-0.071	0.018
	(0.06)	(0.06)	(0.06)	(0.06)	(0.09)	(0.12)	(0.08)	(0.10)
Verbal Score - Age 4	0.224***	-0.053	0.224***	-0.055	0.210**	-0.044	0.213***	-0.042
	(0.06)	(0.10)	(0.06)	(0.10)	(0.09)	(0.18)	(0.08)	(0.17)
Prosocial Behavior - Age 4	-0.091**	-0.146***	-0.091**	-0.146***	-0.029	-0.053	-0.027	-0.058
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)	(0.04)	(0.06)
Externalizing Behavior - Age 4	-0.023	0.016	-0.024	0.015	-0.049	-0.018	-0.048	-0.023
	(0.05)	(0.06)	(0.04)	(0.06)	(0.04)	(0.05)	(0.04)	(0.04)
Verbal Score - School Entry	-1.022***		-1.025***					
	(0.10)		(0.10)					
Verbal Score - First Grade					-1.218***		-1.219***	
					(0.14)		(0.13)	
Family Controls	✓	✓	✓	✓	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓	✓	✓	✓	✓
P-value: Constant		0.465		0.431		0.012		0.011
P-value: YOUNG		0.938		0.748		0.871		0.558
P-value: UNREADY		0.896		0.817		0.428		0.178
P-value: YOUNG*UNREADY		0.053		0.132		0.547		0.215
N	665	665	665	665	247	247	247	247
R-Squared	0.34	0.11	0.34	0.11	0.56	0.15	0.56	0.15

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months. Specification 2 excludes initial level of ability measure.

Table B.3: Prosocial Behavior

	First Grade				Second Grade			
	OLS	OLS/spec2	IV	IV/spec2	OLS	OLS/spec2	IV	IV/spec2
Constant	0.786	0.939	0.808	0.988	0.383	0.801	0.224	0.780
	(0.53)	(0.68)	(0.51)	(0.65)	(0.63)	(0.87)	(0.54)	(0.74)
YOUNG	0.032	0.058	0.079	0.159	0.120	0.036	0.119	0.013
	(0.09)	(0.13)	(0.08)	(0.12)	(0.16)	(0.21)	(0.22)	(0.26)
UNREADY	0.063	0.184*	0.112	0.291***	-0.485	-0.460	-1.143	-0.672
	(0.09)	(0.09)	(0.09)	(0.10)	(0.60)	(0.64)	(0.98)	(0.83)
YOUNG*UNREADY	-0.238	-0.180	-0.352*	-0.428*	0.524	0.668	1.222	0.894
	(0.17)	(0.21)	(0.18)	(0.23)	(0.63)	(0.73)	(0.99)	(0.90)
Gender - Girl	0.140	0.164	0.146	0.177	0.127	-0.017	0.105	-0.023
	(0.09)	(0.12)	(0.09)	(0.12)	(0.12)	(0.16)	(0.11)	(0.14)
Verbal Score - Age 4	-0.038	-0.081	-0.040	-0.085	-0.140*	-0.177*	-0.148**	-0.179**
	(0.07)	(0.09)	(0.07)	(0.08)	(0.07)	(0.10)	(0.07)	(0.09)
Prosocial Behavior - Age 4	0.034	-0.109**	0.034	-0.111**	0.175*	0.062	0.182**	0.064
	(0.06)	(0.05)	(0.06)	(0.05)	(0.09)	(0.09)	(0.09)	(0.09)
Externalizing Behavior - Age 4	0.079**	-0.041	0.079**	-0.043	0.126	0.098	0.134**	0.100
	(0.04)	(0.04)	(0.03)	(0.04)	(0.07)	(0.08)	(0.07)	(0.07)
Prosocial - School Entry	-0.677***		-0.678***					
	(0.04)		(0.03)					
Prosocial - First Grade					-0.620***		-0.618***	
					(0.08)		(0.07)	
Family Controls	✓	✓	✓	✓	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓	✓	✓	✓	✓
P-value: Constant		0.116		0.599		0.334		0.290
P-value: YOUNG		0.509		0.225		0.565		0.598
P-value: UNREADY		0.217		0.006		0.908		0.272
P-value: YOUNG*UNREADY		0.361		0.400		0.607		0.447
N	698	698	698	698	261	261	261	261
R-Squared	0.37	0.06	0.37	0.06	0.36	0.11	0.35	0.11

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months. Specification 2 excludes initial level of ability measure.

Table B.4: Externalizing Behavior

	First Grade				Second Grade			
	OLS	OLS/spec2	IV	IV/spec2	OLS	OLS/spec2	IV	IV/spec2
Constant	0.199	0.537	0.153	0.487	1.203	1.501	1.699**	2.091*
	(0.40)	(0.44)	(0.37)	(0.41)	(0.77)	(1.14)	(0.84)	(1.13)
YOUNG	-0.596***	-0.722***	-0.591***	-0.696***	-0.715**	-0.622*	-1.160**	-1.108**
	(0.13)	(0.18)	(0.13)	(0.16)	(0.29)	(0.32)	(0.46)	(0.53)
UNREADY	-0.112	-0.218	-0.142	-0.234*	-1.101**	-1.234**	-1.470**	-1.461*
	(0.12)	(0.14)	(0.12)	(0.14)	(0.52)	(0.45)	(0.69)	(0.76)
YOUNG*UNREADY	0.340**	0.609***	0.416***	0.654***	1.021*	1.030*	1.432*	1.292
	(0.14)	(0.18)	(0.15)	(0.18)	(0.55)	(0.50)	(0.75)	(0.82)
Gender - Girl	0.128*	0.097	0.124*	0.095	0.118	0.041	0.120	0.048
	(0.07)	(0.08)	(0.07)	(0.08)	(0.11)	(0.12)	(0.11)	(0.11)
Verbal Score - Age 4	-0.019	0.064	-0.018	0.065	0.019	0.064	0.030	0.079
	(0.05)	(0.06)	(0.05)	(0.06)	(0.09)	(0.11)	(0.08)	(0.10)
Prosocial Behavior - Age 4	0.077**	-0.005	0.076**	-0.006	0.110	0.009	0.116*	0.012
	(0.04)	(0.04)	(0.03)	(0.04)	(0.07)	(0.08)	(0.07)	(0.08)
Externalizing Behavior - Age 4	0.165**	-0.103*	0.164***	-0.103**	0.317***	0.071	0.318***	0.067
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)
Externalizing - School Entry	-0.602***		-0.598***					
	(0.05)		(0.04)					
Externalizing - School Entry					-0.575***		-0.579***	
					(0.08)		(0.07)	
Family Controls	✓	✓	✓	✓	✓	✓	✓	✓
Child Characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Home Environment	✓	✓	✓	✓	✓	✓	✓	✓
P-value: Constant		0.231		0.252		0.566		0.505
P-value: YOUNG		0.073		0.140		0.462		0.678
P-value: UNREADY		0.028		0.063		0.516		0.936
P-value: YOUNG*UNREADY		0.004		0.001		0.962		0.563
N	698	698	698	698	262	262	262	262
R-Squared	0.31	0.11	0.31	0.11	0.34	0.14	0.34	0.13

Standard errors clustered at province level. * p<0.10, ** p<0.05, *** p<0.01. Family controls include urban residence, household size, number of kids, birth order, mother's years of completed education, father's years of completed education, mother's working status, pre-school enrollment at age 5, socio-economic status, mother's vocabulary knowledge and Home environment measures at age 4; language stimulation, academic stimulation, learning materials, punishment and responsivity. Child control includes verbal ability score at age 4, prosocial behavior score at age 4, externalizing behavior score at age 4, inhibitory control at age 4, measurement gap in months. Specification 2 excludes initial level of ability measure.

Chapter 3

Understanding the Link Between Aspirations and Noncognitive Skills in Childhood

3.1 Introduction

Raising educational and occupational aspirations of students has been at the core of policy debates for the last decade; see Cabinet Office (2011); Gutman and Akerman (2008). As achievement gaps appear to be widening around the world, these debates tend to focus on socioeconomically disadvantaged segments of the population. Children brought up in less favorable conditions perform worse in tests and obtain less education (Brooks-Gunn and Duncan, 1997; Dahl and Lochner, 2012; Heckman and Masterov, 2007). Since education has important social and economic benefits (Card, 1999; Grossman, 2006; Machin et al., 2011), encouraging students from disadvantaged backgrounds to adopt higher occupational and educational aspirations is thought to mitigate these attainment gaps and to induce better life outcomes(Cuthbert and Hatch, 2008).

Although understanding the nature of aspirations and its connection to real outcomes is crucial for designing targeted policy actions, it's a challenging task. Aspirations are usually described to be products of multidimensional factors reflecting both present and future perspectives, and as such, they are difficult to measure. They are considered to be an individual's ability to identify and set goals for the future, while being inspired in the present to work toward those goals (Quaglia and Cobb, 1996). Despite being compared to expectations, aspirations include a sense of idealism and desire for future outcomes. They begin to form early in child's life and they are shaped through different experiences and environments. As the child matures, his/her aspirations are modified through a better understanding of the world and his/her individual cognitive as well as noncognitive capabilities.

The main objective of this paper is to study the determinants of aspirations in childhood. Using unique data collected as part of a large field project, I explore the extent to which cognitive and noncognitive skills are associated with occupational and educational aspirations in elementary school students. The data contain rich information on third and forth graders (age range of 9-11) from a large number of elementary schools in Istanbul, Turkey. The main variables of interest in this study are the occupational and educational aspirations of students. To elicit the former, students were asked to write the occupation they would like to have when they grow up in an open-ended question. I group the answers into occupation categories the Standard Occupational Classification (2010 SOC) System by the Bureau of Labor Statistics. Students were also asked whether they want to attend higher education (university) which I use to define their educational aspiration. In addition to self-reported aspiration questions, teachers were asked to express their opinion about the most suitable occupation for a given pupil, using one of the 15 choices that are related to Standard Occupational Categories. These questions allow me to construct several measures of occupational and educational aspirations, both reported by children and their teachers.

The noncognitive skills I consider are those shown to be strong predictors of achievement in a multitude of domains in the literature. Specifically, I consider risk attitude, patience, competitiveness, self-confidence and challenge seeking behavior[1]. In my data, these skills were elicited in the classroom environment using strong incentives. Risk and time preferences were elicited using Gneezy and Potters (1997) investment game and Andreoni and Sprenger (2012) convex time budget task, respectively. Competitiveness and challenge seeking were also measured via incentivized tasks designed by Niederle and Vesterlund (2007) and Alan et al. (2018), respectively. I construct a self-confidence measure from student's answers about their own performance and their opponents in the real effort task for competition. In addition to these noncognitive skill measures, my data contain a rich set of cognitive ability indicators measured with Raven's Progressive Matrices (1938) and standardized math and verbal tests.

I first investigate the determinants of aspirations reported by students themselves. Then I conduct a similar investigation for teacher's aspirations for students. My results point to a significant gender difference in the type of noncognitive skills that predict educational and occupational aspirations of children. While competitiveness and challenge seeking appear as two key predictors for boys, patience and self-confidence emerge as the only critical noncognitive skills that predict aspirations for girls. The associations I estimate between these skills and student's self-reported aspirations are weaker than those reported by teachers. I find that teachers are more likely to have high aspirations for students who show inclination toward competitive and challenging tasks. A suggestive mediation analysis show that a nontrivial proportion (about 20%) of the effect of challenge seeking on aspirations estimated for boys is mediated through grades. These results suggest that teachers may have a differential valuation of individual traits and attitudes

[1] A large body of literature provides evidence on the impact of noncognitive skills on achievement. Skills; such as self-confidence, grit, competitiveness and challenge seeking, and preferences over risk and time are good predictors of academic performance and labor market outcomes; see Heckman et al. (2006); Duckworth and Seligman (2005); Koch et al. (2015); Alan et al. (2018); Alan and Ertac (2018a,b).

across genders when forming their beliefs.

In terms of novelty, the paper links two extant literatures using unique data on children. One of these literatures studies career aspirations as core determinants of educational outcomes(Eccles, 2009; Schoon et al., 2004), the other documents noncognitive skills as predictors of achievement over and above cognitive ability (Buser et al., 2014; Oreopoulos, 2007; Schulz and Thöni, 2016; Reuben et al., 2017). This study takes aspirations as important determinant of achievement and explores the associations between aspirations and noncognitive skills, controlling for cognitive ability and socio-economic indicators. It does so by utilizing the latest developments in experimental economics that measures noncognitive skills in an incentive compatible manner. The paper also considers self-reported aspirations as well as teacher reports by distinguishing aspirations and expectations for the latter. Finally, the paper naturally speaks to the vast literature on gender differences in preferences and attitudes. A strand of this literature documents that girls have lower aspirations for high paying, challenging STEM occupations (Chowdry et al., 2011; Goodman and Gregg, 2010; Schoon, 2006). Willitts et al. (2005) shows that girls with higher aspirations are more motivated to study further. My findings show that aspirations can be measured as reliable outcomes in the classroom environment early in childhood and inform educational interventions that aim to raise achievement and mitigate inefficient gender gaps.

The rest of the paper proceeds as follows. In Section 2, I give details of the data used in this research and describe how key measures are constructed. Section 3 describes the estimation procedure and presents a discussion of the results. The final section summarizes and concludes.

3.2 Data

My data come from the Alan and Ertac Noncognitive Skills Study, a large-scale field project launched in 2012. The project involves a large number of state elementary schools in urban districts of Istanbul and was conducted under the approval of the Turkish Ministry of Education and the institutional review board approval obtained from Koc University, Istanbul. The main objective of the project is to study the behaviors, attitudes and outcomes of elementary school students. Total of 41 schools with 106 teachers and 3800 students are recruited for the project. The actual data collection was done through several visits to classrooms during school's regular class hours by a group of experimenters and researchers. The data used in this paper is from the first wave, collected in 2013-2014[2].

A rich battery of tools was designed to collect data from students and teachers. They consist of teachers' evaluations of students, incentivized elicitation tasks for students, a cognitive ability test (Raven's progressive matrices) and student surveys that include inventories for measuring behavioral outcomes, demographic characteristics, future aspirations and expectations. Teachers were given questionnaires that include items to measure their beliefs and expectations about each student's behavior and later life outcomes. Around 28% of the information in the data is missing due to non-attendance by the students on the data collection day, while two of the surveys are administered only to 60% of the participating classrooms. My analysis is based on the teachers and students for whom I have complete set of information on key variables, forming a dataset with 36 schools and 58 teachers and 1180 students. Looking at the descriptive statistics on the key demographic variables presented in Table 3.2, sample of analysis is concentrated with more boys and more students from low or high socioeconomic background compared to full sample. Students are aged between 9-11 years old, with 60% studying at 3rd grade and the rest studying at 4th grade. Female students consti-

[2]Papers that use these data are Alan and Ertac (2015, 2018a,b); Alan et al. (2016, 2018).

tute half of the sample (by 49%). Only a quarter of these students has a working mother at home. This ratio is slightly lower than the female employment rate in Istanbul Province which was 26.7% in 2014(ILO, 2016). In the last few decades middle class families in Turkey tend to prefer private schools over under-resourced state schools for their children. Therefore, the project mainly reaches to student from a disadvantaged background.

3.2.1 Aspirations

The main variables of interest in this study are the student's later life educational and occupational aspirations. A substantial part of the aspiration literature focuses on the students' intentions for educational attainment, and on their drop-out decision. Although elementary school is an early stage in a student's life, analysis of this kind provides evidence on development trajectories of individual's future expectations and their academic motivation. Hence, by reporting the results for educational aspiration I aim to contribute to the existing literature and to compare the estimated effects with those of occupational aspirations. I identify students with positive attitudes towards higher education if they expressed their wish to go to a university when asked about their post-graduation plans. A small proportion of the sample, %9, explicitly states that they do not plan to continue with their studies after high school, which are coded as having low educational aspiration.

Students are asked to write the occupation they would like to have when they grow up in an open-ended question. The responses are rarely the products of children's fantasy but rather self-explanatory which allow me to classify them into a generic occupation category. Table 3.1 provides the list of the top 10 reported occupations with relative frequencies in the total sample and by gender. Most aspirations are career-oriented.The majority of students report that they desire to be a teacher (%21) or a doctor (%18) in the future. Gender subgrouping reveals that girls report more career oriented professions whereas the top choice

of boys is being a football player. I use the Standard Occupational Classification (2010 SOC) System by the Bureau of Labor Statistics to code each reported aspiration into a corresponding occupational category[3]. By doing so, I rescale the responses to facilitate the comparison of high and low aspirations. The 2010 SOC system includes 23 major categories that ranks the occupations from high prestige professions (e.g. management, legal, education and healthcare occupations), to a more intermediate and elementary professions such as technical, administrative and service based occupations[4]. Panel 1 in Figure 3.1 illustrates the distribution of student aspirations under SOC categories. I further convert the aspirations into a binary variable in which occupations that have SOC codes starting with 1 and 2 are identified as high aspirations and the rest as low aspirations. The descriptive statistics of the final variable for student aspirations reveal that 80 % of the sample hold high occupational aspirations where the frequency is considerably higher for girls (%91) compared to boys (%71).

I further categorize aspirations by several other classifications constructed on statistically measured characteristics from recent labour market analysis and demographic distribution of workers belong to the classified occupations. These classifications are mostly derived using subsidiary datasets in SOC system where information on skills and education level acquired by the working population of each occupational category as well as their underlying work interest and work values are provided. I utilize the Socioeconomic index constructed for the 2012 General Social Survey's occupational prestige study to identify which occupations belong to a high income group. The socioeconomic index was first introduced by Duncan (1961) as an attempt to measure socioeconomic status by using the occupational characteristics and it was improved by incorporating various aspects and measurement methods over the years. The version SEI10 that I use to categorize the occupations is estimated using US Census data on income, hours worked, ed-

[3]All SOC definitions are available online from the SOC section of the BLS Web site, https://www.bls.gov/soc/2010/.

[4]The list of SOC major categories is provided in the Appendix.

ucation, gender and employment [5][6]. I also use information provided in the guidelines for STEM (Science, Technology, Engineering, and Mathematics) occupations under the 2010 Standard Occupational Classification (SOC) system to categorize non-stem and stem related occupational aspirations. These occupations mostly consists of engineering, academic and medical professions and are more likely to be high prestige occupations. But there are high prestige managerial positions and legal service occupations that fall into non-stem category. Therefore, my intention is to use this classification to understand whether any student characteristics are predictive of preferences for STEM related occupations, and whether there is evidence available in my data for differential effects of gender as documented in the STEM literature.

O*NET Database is another comprehensive data source that has been specifically designed for understanding the changing nature of labor markets and workforce, in the context of the US, which is the only available dataset that provides detailed information in several domains. It consists of a model that applies occupational information across jobs, sectors, or industries as well as within occupation to identify the content of the work related to each category. These descriptors are organized into six major domains, namely worker characteristics, worker requirements, experience requirements, occupational requirements, workforce characteristics and occupation specific information, which enable researchers to focus on areas of information that specify the key attributes and characteristics of workers and occupations. I incorporate three aspects of the information available in O*NET to identify occupations with (i) social interest; that involve helping or providing service to others, (ii) technical skills; for developed capacities used to design, set-up, operate, and correct malfunctions involving application of machines or technolog-

[5]For detailed information on measurement methods, see Smith and Son (2014); Hout et al. (2015).

[6]I categorized the high income occupations as any occupation with the SEI10 score higher than 72. Occupations with lower SEI10 scores are considered to be in a combined group of medium and low income occupations. I included answers for football players to the high income aspiration group, simply because professional football players are widely known for their high earnings.

ical systems, and (iii) university or postgraduate level of education required to perform the job. Each occupational category is based on the minor grouping of the SOC 2010 system. I took the average of the data values for corresponding skill, education or interest measures when there were more than one minor group related to the occupation reported by the student. Data value for technical skills, social interest were ranged between 1 to 7, where higher numbers correspond to higher importance of those skills. I identify technical and social occupations independently if the occupational category had a data value equal or bigger than 4. Moreover, the measure for level of education required consists of percentages of workers that has highest educational qualifications among 12 categories, ranging from no-degree, high school degree to higher education degree and postgraduate education degree. I identify occupations with level of minimum required education as university if the majority of the workforce has a university or a higher degree[7]. As a final classification, i categorized reported occupations if they were fame-related. As evident in Table 3.1, there is high frequency of reporting for some of the jobs, i.e such as football player, fashion designer, that are widely recognized by the public.

Similarly, teachers were asked to respond which occupation would be best suitable for each student in their classroom to elicit teacher's beliefs on student career choices. They were given 15 choices that are related to Standard Occupational Categories and a blank line to let teachers write down their answers if different than the options provided[8]. Only a small fraction of responses, 5%, were written in the other option and consisted of 32 type of occupations. Occupations that are similar to one of the provided options are coded with the corresponding SOC 2010 code which finalized the coding of teacher aspirations to a total of 30 occupations.

[7]For detailed information on measurement methods, see https://www.onetcenter.org.

[8]The 15 occupational choices consists of the following; teacher, production occupations, management occupations, business and financial occupations, administrative occupations, sales and retail services, doctor, arts and design, lawyer/legal services, scientist/researcher, engineer, politician/policy maker, professional sports person, policeman/military personnel/security forces, unemployed.

Moreover, I apply the same methodology to create binary variables to identify aspirations for high prestige, high income, STEM related, social interest oriented, technical skill dominated, fame-related, university degree required occupations as applied in student aspirations. Panel 2 in Figure 3.1 shows the distribution of teacher aspirations under SOC categories.

Although occupational categorization of teacher aspirations consisted of a higher level of aggregation compared to student aspirations, the aspiration classification methodology that i apply to identify high aspirations across these seven dimension allows me to compare student and teacher aspirations. Table 3.3 reports the correlation coefficients of aspirations between teachers and students as well as within each group. Measures within students and teachers are both highly correlated suggesting that there is important overlaps between occupational aspiration classifications. High prestige occupations are high income occupations and more likely to require university degree. On the other hand, high paying occupations are positively correlated with stem occupations and negatively correlated with social interest oriented occupations. Column 1 also reports significant correlations between teacher and student aspirations except for technical skill dominant jobs.

3.2.2 Noncognitive Skills

My data contain a rich set of variables that measures student's noncognitive skills and attitudes. These variables are constructed using reported decisions of the subjects in incentivized elicitation tasks that are rigorously tailored with the purpose of capturing intended skills. For all tasks, students are provided strong incentives in the form of small gifts that would be of interest to children of these ages and socioeconomic group[9]. These tasks are conducted during the classroom hours, where only the adult present was the experimenter. For some tasks such as eliciting time preferences of the subjects, a second visit is done a week later. Students

[9]These gifts are items of value to children, such as toys and stationary.

are ensured that there would be no uncertainty in receiving their gifts in case they choose the delayed reward option.

To measure risk attitude, students are given a risk preference elicitation task in which they make an individual decision under risk. The task is based on Gneezy and Potters (1997). Students are endowed with five tokens corresponding to gifts from a gift basket. They are confronted with a lottery where they were asked to allocate their tokens between the risky and riskless option. The tokens invested in the risky option is either tripled or lost, with a 50% chance. Instructions are simplified in order to reduce any possible misunderstanding. Students are told that the tokens invested in the risky option are put in a particular bowl and tokens that are not put in the bowl are safe. Uncertainty is resolved through a draw from an opaque urn that contains one yellow and one purple ball. If the yellow ball is drawn, the good outcome occurs. If the purple ball is drawn, the tokens placed in the risky bowl are lost. I use the number of tokens that are invested in the risky option as a measure of risky behavior. The variable takes the value between zero and five where higher values mean a more positive attitude towards risk. The average number of tokens invested in the risky option is 2.4. While 10% of the sample chooses to invest fully in the risky option, only 4% of the students puts all the endowed tokens in the safe option.

A measure of patience is derived from an intertemporal investment decision task. Students are asked to allocate five tokens between two options, today vs one week later, where the later option has an interest rate of r=0.25. The amount invested in the later option with higher return is a measure of patience. The variable included in the model specification ranges between 0-5 and the higher values corresponds to a higher patient behavior. Students show highly patient behavior as the average number of allocated tokens in the one week later option is 4.3. Moreover, a majority of the students (31%) chooses to fully invest in the later option whereas only 8% of the sample chooses to receive all five tokens today.

A measure of competitiveness is constructed using a real effort task, which involves adding two 2-digit numbers and one single-digit number in 2.5 minutes. For the purpose of eliciting willingness to compete, each student is matched with another student from a different school, who had done the same tasks before and whose performance was recorded. In the first period, students perform the addition task under a piece-rate incentive scheme where the payoff is determined as one token for each correct calculation. In the second period, students are asked to choose between the piece-rate and a tournament where they are matched with an opponent to compete for the reward. The payoff scheme for the piece-rate option remains same as in period 1. Under the tournament, students are rewarded with three tokens per correct answer if their performance exceeds that of their opponent. If they choose the tournament and perform lower than their opponent, they receive zero tokens. In case of a tie, they receive one token per correct answer. The choice submitted for period 2 is used to construct the binary variable for willingness to compete which equals to 1 for tournament and 0 otherwise. Students who choose the tournament constitute 47% of the sample. Before proceeding with the task in period 2, students are requested to report their beliefs on how many correct answers they and their opponent will have. I use the reported answers to construct a continuous measure of relative self-confidence by taking the difference between the expected number of answers about own performance and that of the opponent.

Students are presented with another real effort task which is designed to measure the challenge seeking behavior of the subjects. This is a novel task developed by Alan and Ertac with the purpose of measuring various facets of grit, a noncognitive skill shown to be crucial for academic achievement (see Alan et al. (2016)). The task consists of five rounds of a mathematical puzzle in which the goal is to find pairs of numbers that add up to 100. Students are considered to be successful if they meet the performance target which is to find three pairs within 1.5 minutes. One of the five rounds is selected at random and rewards are given based on

student performance in that respective round. The students are provided with two different booklets with 5 puzzles in each, a 4-gift game booklet and 1-gift game booklet. They are also distributed a choice sheet where they are given the option to choose between the 4-gift puzzle and the 1-gift puzzle for each round. These two games differ in terms of their payoff scheme and their difficulty levels. When the performance target is achieved under the 4-gift puzzle, the student receives four gifts. If the 1-gift puzzle is chosen in a round, the student receives one gift in case of success. If the performance target is not met, no gifts are given in both games. The 4-gift puzzle is more difficult compared to the 1-gift puzzle even though the success criteria are the same in both games. In particular, the grid is smaller in the 1-gift puzzle, hence the matching pairs are easier to identify. Before starting the each round, students mark their game of choice in their choice sheet. Once every student completes marking, they are instructed to open the relevant page of their booklet of choice at the same time. They are given 1.5 minutes to find as many pairs of numbers as they can. Once the 1.5 minutes are over, students are instructed to turn the page and wait for the experimenters to check their performance. They are given immediate feedback whether they succeeded or failed that round's game. This is repeated for all 5 rounds which enables students to switch back and forth between the two games as the rounds progress. The total number of rounds that a student chooses to solve the 4 gift puzzle is defined to as the measure of challenge seeking behavior. Higher values of the variable corresponds to a higher taste and willingness for taking the challenge and hence, it is considered to be an indicator of perseverance. The majority of students (32%) chooses the 4-gift puzzle in all rounds whereas only 7% of the sample consistently solves the 1-gift puzzle.

In addition to aforementioned variables, my rich data set includes standard socio-demographic information on students, reported by students themselves as well as reported by teachers for each student. These variables include information on siblingship size, mother's employment status and family's socioeconomic status.

A descriptive statistics of all the variables used in the analysis are provided in Table 3.4 for full sample and gender subgroup and Table 3.5 for subgroups by socioeconomic background of the students. There are significant mean differences for teacher and student occupational aspirations between boys and girls, whereas for socioeconomic subgroups these differences are evident between low vs medium and low vs high ses students. Further inspection of noncognitive skills reveals that skills except willingness to compete are significantly different between boys and girls while risk and challenge seeking behavior varies across socioeconomic groups.

3.3 Results

3.3.1 Empirical Model

In this section, I report the estimated associations between individual noncognitive skills and aspirations. I adopt a multiple definitions for categorizing the student aspirations when evaluating its determinants. The sample of students that are subjects of the analysis are at their early stage of academic course. I examine student attitudes towards higher education and classify their occupational aspirations through different dimensions. The main analysis includes occupational aspirations under a broad definition based on high prestige occupation categories. Then, I move on to analysis of a categorization of aspirations that is done by occupations' ratings in socioeconomic index and by its associations to STEM (science, technology, engineering, and mathematics) subjects.

I estimate the following model for the determinants of aspirations in elementary school students:

$$Y_i = cons + \mathbf{X_i}\alpha + \mathbf{SOC_i}\beta + \mathbf{A_i}\gamma + \varepsilon_{iks}$$

where the dependent variable Y_i is a dummy variable which equals to 1 if the reported aspirations for student i is high. Matrix $\mathbf{X_i}$ includes student characteristics such as age (in months), standardized cognitive ability measure (Raven IQ test) and $\mathbf{SOC_i}$ contains family characteristics and socioeconomic indicators. Finally, $\mathbf{A_i}$ includes the vector of noncognitive ability variables including risk attitude, patience, competitiveness, self-confidence and challenge seeking behavior. The noncognitive skills are given special attention when reporting the results as my main motivation in this paper is to understand how the individual differences in these skills explain the aspiration formation. Although teacher characteristics, especially gender role beliefs, are important predictors of student expectations and performance as well as teacher's aspirations about the students, data on teacher characteristics, measures of their beliefs, and teaching styles are not available to incorporate. I estimate a logit regression model where standard errors are clustered at classroom level.

3.3.2 Student Aspirations

I begin by analyzing the impact of noncognitive skills on the student reported aspirations. Table 3.6 and 3.7 reports the marginal effects from logit regressions for students' educational aspirations and occupational aspirations. The binary dependent variables indicate whether the student wants to enroll in a higher education institute after graduating high school, or wants to work in a high prestige occupation. All tables have the same structure for reporting the estimation results. Column 1 presents the marginal effects of students' cognitive ability and demographic characteristics; including student gender, age, family structure and socio-economic status indicators. Column 2 to 7 add the measures of noncognitive attitudes including risk attitude, patience, willingness to compete, relative self-confidence and challenge seeking behavior. Finally, column 7 presents the result of the model with a complete set of variables.

Results in Table 3.6 indicate that the students' cognitive abilities and gender are persistent predictors for high educational aspirations. Girls have higher positive attitudes towards higher education, by 7 percentage points, when compared to boys. Further investigation on the noncognitive skills in column 2-6 reveals that, after controlling for individual characteristics, two important skills have significant predictive powers. Students who are competitive have positive attitudes towards higher education. Moreover, challenge seeking behavior is also significant, this variable ranges between 0-5 and higher values corresponds to higher challenge seeking behavior, and a unit increase leads to an increase in the probability of having high aspirations by 1.6 percentage points. These effects are robust to inclusion of all the variables in column 7.

Turning to the occupational aspirations in Table3.7, it is evident that some of the student characteristics are strong predictors of high aspirations. Findings from previous studies suggest that the family income affects educational attainment and later life outcomes which further explains the persistence of disadvantage across generations. Similar findings are present in the estimation results in Table 3.7. Being a student from a low SES family reduces the likelihood of having high aspirations by 7 percentage points. However, I find no significant effect of belonging to a high SES family. Note however that in my sample mainly contains lower income households in Istanbul so that the high SES indicator is only relative in my context. Mother's employment status and the number of siblings in the family which I use as proxies for socio-economic status likewise have significant associations with occupational aspirations.

It is difficult to identify the mechanisms through which these associations are realized, especially in my estimation strategy. Interpretation of the estimated results for family structure variables is not the main objective of this study. However, the estimated effect can be a combination of direct and indirect effects such as role modeling, information provision etc., and is likely to capture other confounding

effects such as maternal education. It is vital to control for family structure variables while estimating the model. Other student characteristics such as cognitive ability and student gender display important associations for having high aspirations. Note that the estimated effect of gender is notably higher compared to that of in educational aspirations, and robust to inclusion of noncognitive skills. Girls, when compared to boys, are more likely to report high occupational aspirations by almost 19 percentage points. It is also consistent with the findings in the literature which suggest that girls usually tend to hold higher positive educational attitudes and aspirations (Rampino and Taylor, 2013).

Estimates of the noncognitive skills in columns 2-7 of Table3.7 reveal important patterns. I find that willingness to compete and challenge seeking behavior are important and significant predictors of high aspirations, whereas skills such as risk attitude, patience and self confidence are not significant. The effects sizes for willingness to compete and challenge seeking behavior is very similar to the estimates obtained from analysis of educational aspirations. Students who are competitive are 4 percentage points more likely to have high aspirations. Challenge seeking behavior is positively associated with high aspirations. Column 7, in Table3.7, presents the estimation results of the full model. Challenge seeking behavior remains statistically significant when all set of variables are jointly included. Although the the size of the effect of competitiveness is slightly lower than that of column 4, the effect is no longer significant.

3.3.3 Aspirations Based on Alternative Definitions

I present analysis of student aspirations based on 6 different dimensions by combining information of student's reported ambitions for their future careers and the characteristics of these occupations; high income occupations, STEM-related, social interest oriented, technical skill dominated, fame-related, and minimum university degree required occupations.

Table 3.8 reports marginal effects from a logit regression of these 6 outcome variables for the full model specification. Unlike the educational and prestige based occupational aspirations, a reversed gender effect is present for high income, technical and fame-related occupational aspirations. Boys are more likely to hold high income occupational aspirations, almost the same effect size as in the prestige based outcome analysis, but less likely to have aspirations for STEM related occupations. One potential explanation is that boys might be drawn to monetary rewards more than girls, or beliefs of high income earners being men in the society are transmitted explicitly through cultural references or simply by the family which further affects their expectations and desires for future outcomes. The education requirement of the professions and the prestige of the occupation categories simply matter less than the financial returns of these occupations. It could also be due to their reported aspirations reflecting their preference over other characteristics of the occupation that are highly correlated with socioeconomic status. It is highly probable that boys tend to report aspirations for occupations that have a high concentration of male workers and high earners. If this was the case, high income aspirations could simply be incorporating current dynamics of social values, implicit biases present in the labor market.

I also observe a weaker link compared to the prestige analysis between family income and socioeconomic status. Students from low SES families have negative significant associations with occupational aspirations. Interpreting the estimated coefficients is extremely limited within this framework due to lack of exogenous variation to establish causal links, but the results present suggestive evidence for aspiration traps. If low income students have early realization of their potential credit constraint or similarly lack of information on the potential returns of education on earnings, and evaluate their set of choices based on the available resources, then this might perpetuate the transmission of adverse income effect across generations[10]. Further inspection of the impact of noncognitive skills on aspirations

[10]Studies in developing country context explores lack of information on returns of education

reveals that willingness to compete is the only strong predictor for aspirations for high income occupations, whereas relative self confidence has a significant association with STEM related occupational aspirations. It should also be noted that challenge seeking behavior has a similar size compared to the prestige analysis, but nonsignificant effect. However, it is negatively correlated with social occupations. Analysis of the remaining classifications, technical, fame and university required occupations, provides no evidence between non-cognitive skills and aspirations.

3.3.4 Teacher Reported Aspirations

Teachers are an important evaluators in student related analysis. They maintain a constant stream of information regarding student characteristics and behaviors, and often update their opinions after receiving information. They also have tools to measure certain aspects of student performance and can directly and indirectly influence students through their actions, provision of assistance, materials and feedback. Hence, including analysis based on teacher reports provides important insights to understand the dynamics of student performances and decisions. I apply the same empirical model as for student reported aspirations. All the explanatory variables, individual characteristics and family characteristics are the same, including the skills which are collected from students using experimental tasks and psychometric tests. Essentially the aim is to understand which direct measures of student characteristics and skills explain the aspirations reported by teachers. The teacher reported aspirations could be a product of several factors; reflecting what the student communicates to the teacher and what the teacher thinks student is capable of achieving. Hence, I do not aim to disentangle the effect of the teacher judgement of student capabilities, but rather report the effect of skills in predicting teacher's aspirations for their students.

on earnings find that when people are provided with information about this link, it has a positive effect on aspirations and educational investment decisions (**??**).

I use the same format as in student aspirations when reporting the results in Table 3.9. Teachers are more likely to have high aspirations for female students than their male counterparts by 19 percentage points. As expected, they also have high aspirations for students from high socio-economic backgrounds. Being a student from a high SES family, compared to the baseline group of medium SES, increases the probability of high aspirations by 11 percentage points, while low SES reduces it by 15 percentage points. Noncognitive attitudes are also statistically significant in explaining high aspirations. Teachers associate competitiveness, self-confidence and challenge seeking behavior with high prestige occupational aspirations. Table 3.10 reports the result of the estimates which use remaining occupational definitions of aspirations as outcomes. Similar to the student reported aspiration results (see Table 3.7), the gender effect is positive and stronger associations with cognitive measures are documented. The effect for challenge seeking behavior is present for high income, stem, social occupations and occupations with minimum university degree required but the competitiveness is only statistically significant for high income aspirations and university required aspirations, with an impact around 5 percentage points. While relative confidence is negatively correlated with social occupations, it is positively correlated with university required occupations.

3.3.5 Gender Heterogeneity

Large and persistent gender differences in academic performance and educational attainment are well documented (Duckworth and Seligman, 2006). One potential channel of these gender gaps is differences in noncognitive skills and gender-specific attitudes towards and aspirations for education (Buchmann et al., 2008). I now turn to the question whether my estimation results are similar across genders. Tables 3.11 and 3.13 report subsample analysis for student and teacher reported aspirations by gender. The predictor of educational aspirations and occupational aspirations based on prestige and income differ by gender. None of the noncogni-

tive attitudes contribute in having high aspirations for girls whereas, the estimated effect of competitiveness and challenge seeking behavior is driven by boys. These differential effects are more prominent in teacher reported aspirations. The effect of cognitive ability has a stronger association with teacher's aspirations for boys (Table 3.13), more than double in size of that for girls. The probability of a teacher having high aspirations for boys increases by 17 percentage points when cognitive skills increase by one standard deviation, whereas it increases by 8 percentage points for girls. Boys also tend to have greater associations with regard to competitiveness. Students who are willing to compete have higher aspirations by 15 percentage points on prestige classification and 10 percentage points for high income classification. Although positive, the effect is not significant for STEM subject related aspirations. Challenge seeking behavior is the only significant predictor for both genders with similar effect size in the analysis of STEM classification.

There is a clear pattern evident in the results for understanding which skills are important for boys' aspirations. Willingness to compete and challenge seeking behavior persistently appear to have significant associations with aspirations, regardless of the definitions used. However, only teacher reported aspirations for high prestige occupations produces significant effects for girls, where patience, self confidence and challenge seeking behavior are positively associated with having high aspirations. Teachers might attribute patience and self-confidence to female students when evaluating their capacity of achieving the future goals. Their differential evaluations play an important role in shaping student's beliefs and performances. Alan et al. (2018) document that teachers' gender stereotyping beliefs decrease girl's achievement scores through changes in student's own gender beliefs. Their aspirations for student's future outcomes may translate into future gender gaps in occupational choices. It's important to note that a lack of significant effects in other dimensions limits the discussion on the mechanisms in which girls aspirations are influenced. Yet, the results suggest that aspirations are

related to fundamentally different noncognitive skills for boys and girls. Since different occupations require different skills such that a leader, manager or politician I investigate the potential mechanism behind these results in the next section.

3.4 Discussion

One mechanism for how noncognitive skills influence student aspirations could potentially be through grades. Grades are a product of a host of important attributes, including cognitive skills, learning efforts, various parental and teacher inputs, and noncognitive skills. If students who are more willing to compete or who show higher challenge seeking behavior also perform better in school subjects, put more effort, or receive positive feedback from their teachers, then they might further update their aspirations. They might also believe they have higher chances to attain their desired career choices. They could display persistent effort and be more ambitious about their goals which would translate into their future aspiration.

While being cautious due to the fact that my estimates are not causal, I perform a statistical mediation analysis to explore whether the effect of two important skills, competitiveness and challenge seeking, on student aspirations is explained by the differences in teacher given grades on core subjects. I construct an average of grades for four subjects; Maths, Turkish Language, Science and Social Sciences in the academic term from teacher reports. In order to establish the link between skills and aspirations through grades, I run a causal mediation analysis developed by Imai et al. (2010) that utilizes a potential outcome estimation method. In previous sections, the potential outcomes were only a function of the non-cognitive skills, whereas in a causal mediation analysis the potential outcomes depend on the mediator, in my case the reported grades, as well as the non-cognitive skills. By evaluating the noncognitive skills as treatment variables, I investigate how grades

mediate the effect of the treatment on aspirations. In the potential outcome framework, the effect of treatment, non-cognitive skills, on the outcome measure is called the total effect, equation 3.1, and can be decomposed into the causal mediation, equation 3.2, and direct effects, equation 3.3. Causal mediation effect is the difference between the potential outcome that would result under treatment status x, and the potential outcome that would occur if the treatment status is the same and yet the mediator takes a value that would result under the other treatment status.

$$E[Y_i(1) - Y_i(0)] = E[Y_i(1, M(1)) - Y(0, M(0))] \tag{3.1}$$

$$E[Y_i(x, M(1)) - Y_i(x, M(0))] \tag{3.2}$$

$$E[Y_i(1, M(x)) - Y_i(0, M(x))] \tag{3.3}$$

Willingness to compete is defined to be a binary variable which facilitates an easy interpretation for the treatment effect. However, the challenge seeking behavior is constructed in a continuous way that ranges between 0 and 5. I choose to keep continuous measure of the latter variable in mediation analysis to maintain the nature of its interpretation as in the previous analysis. Therefore, I apply a method suggested in Imai et al. (2010), where the levels of 1 and 4 is assigned to be control and treatment values for mediator model. Grades are simulated taking into account these introduced levels but the continuous variable of challenge seeking is included in the aspiration model. Although treatment is not randomly assigned to students, I treat the aforementioned variables in an exogenous perspective for the sake of understanding the mechanism and purely take the evidence at its face value for suggestive interpretation. Note that additional strong assumptions

are required to identify the Average causal mediation effect (ACME) and average direct effect (ADE).

Estimation of the average mediation effect and direct effect consists of multiple steps and two models; the mediator model of students' grades, and the outcome model of aspirations, where all the student characteristics and background variables are considered as covariates, and competitiveness and challenge seeking are separately included as treatment variables, presented in equation 3.4. In addition to these variables, the outcome model of aspirations incorporates the mediator variable, grades, as in equation 3.5. I first generate two sets of predicted mediator values for students who are treated, simply meaning to be willing to compete or challenge seeking, and who are not. These predicted mediator values are used to impute potential outcomes of aspirations for various treatment and non-treatment level mediators while conditioning on the skills of interest. The average causal mediator effect (ACME) is then generated by taking the difference between aspirations levels for the two estimated mediating grade values across observations.

$$M_i = cons + \alpha_2 T_i + A_i^j \beta_{\mathbf{2}} + SOC_i \xi_{\mathbf{2}} + X_i \gamma_{\mathbf{2}} + \varepsilon_{i2} \tag{3.4}$$

$$Y_i = cons + \alpha_3 T_i + \vartheta M_i + A_i^j \beta_{\mathbf{3}} + SOC_i \xi_{\mathbf{3}} + X_i \gamma_{\mathbf{3}} + \varepsilon_{i3} \tag{3.5}$$

Tables 3.15 and 3.16 report the results for mediation analysis, the average causal mediation effect (ACME), average direct effect (ADE) and total effect for both student and teacher aspirations for each gender. There are no significant effects of competitiveness for boys on student aspirations (Table 3.15, Panel 1), but some evidence is present for teacher aspirations (Table 3.15, Panel 2). Boys who are willing to compete have higher aspirations for prestige and high income occupa-

tions and these effects are not mediated by the changes in their grades, as ACME is not significant and proportionally very small to the total effect. Turning to the second half of the Panel 1, the results give a hint that grades mediate some of the significant effects of challenge seeking behavior for boys. The total effect is significant for prestige and income level occupational aspirations and roughly 18-22 percentage of the effect comes from grades being affected by challenge seeking. Although the total effect is not significant for STEM aspirations, the percentage mediated by the grades is similar to the other aspiration outcomes. The main analysis in the previous section does not provide any empirical support for competitiveness and challenge seeking as determinants of aspirations in girls. In line with this, mediation analysis fails to establish significant total or mediated effect, reported in Table 3.16, leaving no interpretation for teacher reported grades as a mechanism for aspiration development in girls.

3.5 Conclusion

In this paper I investigate the determinants of aspirations in elementary school students using a unique data set that allows me to include measures for individual characteristics, socio-economic indicators and noncognitive attitudes. My first observation is that girls have higher occupational and educational aspirations compared to boys. Larger proportions of girls report wanting to study further in university after graduating high school and they aspire to work in high level occupations later on. Furthermore, teachers hold higher levels of occupational aspirations for female students and expect them to work in high level occupations. Although the estimated effect of gender varies for different dependent variables, being a girl consistently associated with positive and high aspirations.

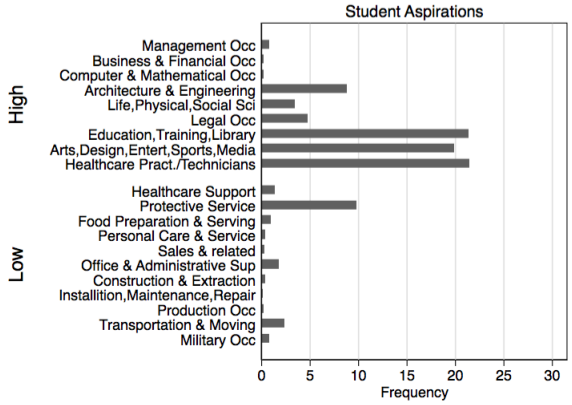
The estimation results suggest that some of the noncognitive attitudes are highly correlated with aspirations. I find that risk preferences influences neither student

nor teacher aspirations, but attitudes towards competition and challenge do. Although the relationship between willingness to compete and student aspiration is weak, teachers have consistently higher aspirations and expectations for students with positive attitudes to competition. It is even more striking that this effect primarily comes from boys. While only patience and relative self-confidence seems to influence teacher's beliefs about female students' future career outcomes, competitive behavior is highly associated for male students. This is particularly important as gender differences in expectations about future earnings while in college are strongly related to confidence and competitiveness. This paper indicates that understanding the mechanisms in early life aspirations might provide insights for emerging gaps in later life outcomes.

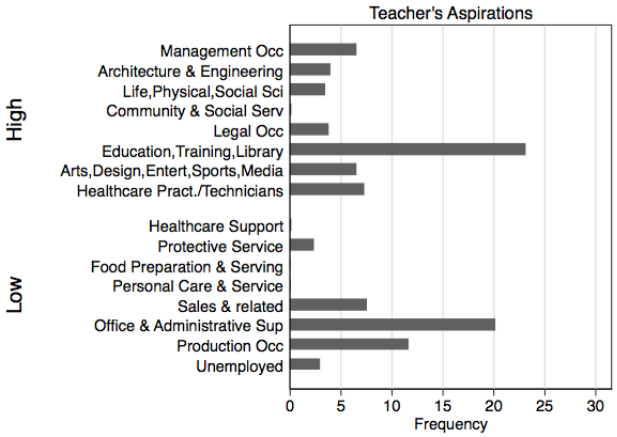
Figures

Figure 3.1: Distribution of Aspirations

Panel 1



Panel 2



Tables

Table 3.1: Top 10 Occupations reported by Students - Total Sample & by Gender

	Total	%	N	Boys	%	N	Girls	%	N
1.	Teacher	21.0	248	Football player	23.3	140	Teacher	37.6	219
2.	Doctor	18.5	218	Police	15.5	93	Doctor	24.4	142
3.	Football player	11.9	141	Doctor	12.7	76	Lawyer	4.3	25
4.	Police	9.6	113	Engineer	8.3	50	Architect	3.6	21
5.	Engineer	5.0	59	Teacher	4.8	29	Police	3.4	20
6.	Architect	3.7	44	Architect	3.8	23	Fashion Designer	3.3	19
7.	Lawyer	3.1	36	Pilot	3.5	21	Nurse	2.7	16
8.	Pilot	1.9	23	Scientists	3.3	20	Surgeon	2.9	15
9.	Scientist	1.9	22	Astronaut	2.0	12	Musician	1.7	10
10.	Fashion Designer	1.7	20	Lawyer	1.8	11	Judge	1.5	9

Table 3.2: Comparison of Full vs Sample of Analysis

	Full Sample	**Non-missing Sample**	**P-value(diff)**
Gender = Male	0.515(0.500)	0.492(0.500)	0.039
Age in Months	119.535(4.988)	119.411(4.592)	0.274
Sibship size	2.686(1.181)	2.644(1.162)	0.124
Working Mother	0.235(0.424)	0.244(0.430)	0.364
Low SES	0.321(0.467)	0.290(0.454)	0.004
Medium SES	0.459(0.498)	0.450(0.498)	0.461
High SES	0.221(0.415)	0.260(0.439)	0.000
N	2074	1004	

Standard Errors are reported in parentheses.

Table 3.3: Occupational Aspirations Classifications, Pearson's correlation coefficients

	Student vs Teacher	Student		Teacher	
		Prestige	Income	Prestige	Income
Prestige	0.199***				
Income	0.164***	0.511***		0.476***	
STEM	0.093***	0.265***	0.611***	0.358***	0.763***
Social	0.192***	-0.187***	-0.374***	0.136***	-0.170***
Technical	0.062*	-0.043	0.138***	-0.198***	0.163***
Fame-related	0.181***	0.220***	0.194***	0.237***	0.188***
University required	0.262***	0.609***	0.135***	0.938***	0.348***

* p<0.10, ** p<0.05, *** p<0.01.
See text for definition of occupations for Student and Teacher aspirations (binary variable).

Table 3.4: Summary Statistics, by Gender

	Full Sample	**Girls**	**Boys**	**P-value Boys=Girls**
Gender = Male	0.492(0.500)			
Raven Score	0.178(0.892)	0.197(0.898)	0.158(0.886)	0.488
Age in Months	119.411(4.592)	119.331(4.269)	119.494(4.906)	0.575
Sibship size	2.644(1.162)	2.635(1.153)	2.654(1.173)	0.800
Working Mother	0.244(0.430)	0.231(0.422)	0.257(0.437)	0.343
Reported Grades	0.293(0.836)	0.361(0.790)	0.223(0.876)	0.010
Low SES	0.290(0.454)	0.282(0.451)	0.298(0.458)	0.596
Medium SES	0.450(0.498)	0.451(0.498)	0.449(0.498)	0.960
High SES	0.260(0.439)	0.267(0.443)	0.253(0.435)	0.623
Noncognitive Skills:				
Risk Attitudes	2.455(1.285)	2.345(1.256)	2.569(1.306)	0.006
Patience	4.392(1.572)	4.500(1.454)	4.281(1.679)	0.027
Willing to Compete	0.488(0.500)	0.476(0.500)	0.500(0.501)	0.456
Relative Self-conf	-0.003(1.012)	-0.069(1.030)	0.065(0.989)	0.037
Challenge Seeking	2.986(1.716)	2.800(1.716)	3.178(1.696)	0.000
Student Aspirations:				
Educational- HE	0.907(0.291)	0.938(0.240)	0.874(0.332)	0.000
Occupational - Prestige	0.810(0.393)	0.900(0.300)	0.717(0.451)	0.000
Occupational - Income	0.547(0.498)	0.456(0.499)	0.642(0.480)	0.000
Occupational - STEM	0.357(0.479)	0.385(0.487)	0.328(0.470)	0.059
Occupational - Social	0.574(0.495)	0.733(0.443)	0.409(0.492)	0.000
Occupational - Technical	0.104(0.305)	0.031(0.174)	0.178(0.383)	0.000
Occupational - Fame	0.211(0.408)	0.120(0.325)	0.306(0.461)	0.000
Occupational - University	0.692(0.462)	0.900(0.301)	0.478(0.500)	0.000
Teacher Aspirations:				
Occupational - Prestige	0.555(0.497)	0.651(0.477)	0.455(0.499)	0.000
Occupational - Income	0.220(0.415)	0.176(0.382)	0.265(0.442)	0.001
Occupational - STEM	0.146(0.353)	0.114(0.318)	0.178(0.383)	0.004
Occupational - Social	0.501(0.500)	0.675(0.469)	0.331(0.471)	0.000
Occupational - Technical	0.190(0.392)	0.083(0.276)	0.294(0.456)	0.000
Occupational - Fame	0.070(0.255)	0.050(0.218)	0.089(0.285)	0.015
Occupational - University	0.539(0.499)	0.679(0.467)	0.402(0.491)	0.000

Standard Errors are reported in parentheses. Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task. Student and Teacher aspirations (binary variable) are classified according to definition of occupations, see text for further details.

Table 3.5: Summary Statistics, by Socioeconomic Status

	Low SES	Medium SES	High SES	P-value		
	(1)	(2)	(3)	(1=2)	(2=3)	(1=3)
Gender = Male	0.505(0.501)	0.491(0.500)	0.479(0.501)	0.710	0.754	0.539
Raven Score	-0.151(0.983)	0.221(0.827)	0.471(0.770)	0.000	0.000	0.000
Age in Months	119.945(5.601)	119.314(4.335)	118.985(3.641)	0.085	0.301	0.019
Sibship size	3.062(1.239)	2.485(1.064)	2.456(1.121)	0.000	0.735	0.000
Working Mother	0.155(0.362)	0.274(0.447)	0.291(0.455)	0.000	0.630	0.000
Reported Grades	-0.161(0.864)	0.380(0.781)	0.659(0.649)	0.000	0.000	0.000
Noncognitive Skills:						
Risk Attitudes	2.409(1.308)	2.398(1.285)	2.605(1.253)	0.912	0.037	0.073
Patience	4.158(1.676)	4.423(1.543)	4.602(1.468)	0.028	0.129	0.001
Willing to Compete	0.460(0.499)	0.485(0.500)	0.525(0.500)	0.523	0.299	0.131
Relative Self-conf	-0.049(1.003)	-0.011(1.042)	0.062(0.969)	0.618	0.360	0.188
Challenge Seeking	2.708(1.732)	3.040(1.687)	3.203(1.714)	0.010	0.216	0.001
Student Aspirations:						
Educational- HE	0.852(0.356)	0.922(0.269)	0.942(0.234)	0.003	0.312	0.001
Occupational - Prestige	0.715(0.452)	0.841(0.366)	0.862(0.345)	0.000	0.444	0.000
Occupational - Income	0.460(0.499)	0.585(0.493)	0.579(0.495)	0.001	0.859	0.006
Occupational - STEM	0.296(0.457)	0.379(0.486)	0.387(0.488)	0.019	0.836	0.023
Occupational - Social	0.653(0.477)	0.560(0.497)	0.510(0.501)	0.011	0.196	0.001
Occupational - Technical	0.069(0.253)	0.106(0.308)	0.138(0.345)	0.084	0.206	0.007
Occupational - Fame	0.172(0.378)	0.221(0.416)	0.238(0.426)	0.102	0.617	0.055
Occupational - University	0.608(0.489)	0.705(0.457)	0.762(0.426)	0.006	0.098	0.000
Teacher Aspirations:						
Occupational - Prestige	0.347(0.477)	0.584(0.493)	0.736(0.442)	0.000	0.000	0.000
Occupational - Income	0.117(0.322)	0.230(0.421)	0.318(0.467)	0.000	0.010	0.000
Occupational - STEM	0.069(0.254)	0.153(0.360)	0.218(0.414)	0.001	0.027	0.000
Occupational - Social	0.482(0.501)	0.517(0.500)	0.494(0.501)	0.361	0.562	0.774
Occupational - Technical	0.263(0.441)	0.183(0.387)	0.124(0.330)	0.011	0.039	0.000
Occupational - Fame	0.044(0.205)	0.068(0.252)	0.100(0.301)	0.185	0.124	0.011
Occupational - University	0.343(0.476)	0.573(0.495)	0.687(0.465)	0.000	0.003	0.000

Standard Errors are reported in parentheses. Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task. Student and Teacher aspirations (binary variable) are classified according to definition of occupations, see text for further details.

Table 3.6: Student's Aspirations to Higher Education

Risk attitudes		-0.008					-0.010
		(0.01)					(0.01)
Patience			0.003				0.002
			(0.01)				(0.01)
Willingness to compete				0.044**			0.037**
				(0.02)			(0.02)
Relative Self-confidence					0.009		0.004
					(0.01)		(0.01)
Challenge seeking behavior						0.015***	0.013**
						(0.01)	(0.01)
S Gender = Male	-0.064***	-0.062***	-0.064***	-0.064***	-0.065***	-0.069***	-0.067***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Raven Score	0.035***	0.034***	0.034***	0.033***	0.035***	0.028***	0.026**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age in months	0.002	0.002	0.002	0.002	0.002	0.002	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.005	-0.005	-0.005	-0.005	-0.005	-0.006	-0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working Mother	0.008	0.008	0.008	0.007	0.009	0.009	0.008
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Low SES	-0.039	-0.040	-0.039	-0.039	-0.039	-0.037	-0.038
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
High SES	0.018	0.019	0.017	0.016	0.017	0.017	0.018
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	988	988	988	988	988	988	988

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variable: Binary variable, is equal to 1 if the student wants to study further after high school.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task.

Table 3.7: Student's Occupational Aspirations - Prestige

Risk attitudes		0.009					0.007
		(0.01)					(0.01)
Patience			-0.001				-0.002
			(0.01)				(0.01)
Willingness to compete				0.039*			0.026
				(0.02)			(0.02)
Relative Self-confidence					0.013		0.007
					(0.01)		(0.01)
Challenge seeking behavior						0.016**	0.014*
						(0.01)	(0.01)
S Gender = Male	-0.182***	-0.184***	-0.182***	-0.183***	-0.184***	-0.188***	-0.190***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Raven Score	0.039***	0.040***	0.039***	0.037***	0.038***	0.031**	0.032**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age in months	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.022**	-0.022**	-0.022**	-0.022**	-0.022**	-0.023**	-0.023**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working Mother	0.051*	0.051*	0.051*	0.050*	0.052*	0.051*	0.052*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Low SES	-0.076***	-0.075***	-0.076***	-0.075***	-0.076***	-0.072***	-0.072***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
High SES	0.014	0.012	0.015	0.013	0.013	0.014	0.011
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
N	1004	1004	1004	1004	1004	1004	1004

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variable: Binary variable for the top two major category in Standard Occupational Classifications. This category includes Management Occupations, Business and Financial Occupations, Computer and Mathematical Occupations, Architecture and Engineering, Life, Physical, and Social Sciences, Community and Social Services, Legal Occupations, Education, Training, and Library, Arts, Design, Entertainment, Sports and Media, Healthcare Practitioners and Technicians.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task.

Table 3.8: Student Occupational Aspirations

	Income	STEM	Social	Technical	Fame	University+
Risk attitudes	-0.007	-0.005	-0.004	0.001	0.005	0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Patience	0.004	0.004	0.009	-0.006	-0.001	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Willingness to compete	0.067**	0.011	-0.028	-0.018	0.036	-0.037
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Relative Self-confidence	0.021	0.026*	0.013	-0.005	-0.007	0.008
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Challenge seeking behavior	0.014	0.009	-0.016**	0.005	0.006	0.009
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
S Gender = Male	0.176***	-0.060*	-0.295***	0.134***	0.181***	-0.384***
	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)
Raven Score	0.037	0.039**	-0.055***	0.038***	0.004	0.052***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Age in months	0.001	-0.002	-0.004	0.001	0.004	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.023	0.016	0.044***	-0.003	-0.045***	0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working Mother	0.037	-0.003	-0.067*	-0.003	0.047	0.005
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Low SES	-0.086**	-0.073*	0.043	-0.048*	-0.022	-0.066**
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
High SES	-0.021	-0.006	-0.037	0.044**	0.016	0.038
	(0.04)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)
N	1003	1003	1004	1004	1004	1003

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variable: Income, a binary variable for occupations that are 0.72 or higher in SEI2010, Socioeconomic Index measured by Duncan's (1961). STEM, a binary variable for occupations that are classified as stem or stem-related in Standard Occupational Classification. Social, a binary variable if data value for social skills in O*NET system is 4 or higher. Technical, a binary variable if data value for technical skills in O*NET system is 4 or higher. University+, binary variable if occupation has the majority of university and postgraduate degree worker ratio in O*NET system.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task.

Table 3.9: Teacher Aspirations for Students - Prestige

Risk attitudes		-0.009					-0.018*
		(0.01)					(0.01)
Patience			0.013				0.011
			(0.01)				(0.01)
Willingness to compete				0.096***			0.068**
				(0.03)			(0.03)
Relative Self-confidence					0.055***		0.046***
					(0.01)		(0.01)
Challenge seeking behavior						0.027***	0.018*
						(0.01)	(0.01)
S Gender = Male	-0.181***	-0.179***	-0.178***	-0.184***	-0.188***	-0.193***	-0.191***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Raven Score	0.151***	0.151***	0.146***	0.144***	0.148***	0.135***	0.127***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Age in months	0.005	0.005	0.005	0.005	0.005	0.004	0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.036**	-0.035**	-0.034**	-0.036***	-0.035**	-0.037***	-0.033**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working Mother	-0.071**	-0.072**	-0.068**	-0.070**	-0.071**	-0.072**	-0.070**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Low SES	-0.158***	-0.159***	-0.157***	-0.157***	-0.159***	-0.154***	-0.154***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
High SES	0.114***	0.117***	0.114***	0.112***	0.111***	0.114***	0.113***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
N	1004	1004	1004	1004	1004	1004	1004

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variable: Binary variable for the top two major category in Standard Occupational Classifications. This category includes Management Occupations, Business and Financial Occupations, Computer and Mathematical Occupations, Architecture and Engineering, Life, Physical, and Social Sciences, Community and Social Services, Legal Occupations, Education, Training, and Library, Arts, Design, Entertainment, Sports and Media, Healthcare Practitioners and Technicians.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task.

Table 3.10: Teacher Aspirations for Students

	Income	STEM	Social	Technical	Fame	University+
Risk attitudes	0.008	0.003	-0.007	-0.007	-0.002	-0.019*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Patience	0.008	0.004	0.006	-0.003	-0.003	0.009
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Willingness to compete	0.046*	0.023	0.004	0.002	0.006	0.058*
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Relative Self-confidence	0.013	-0.006	-0.031**	-0.006	0.012	0.039***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Challenge seeking behavior	0.018**	0.021***	0.016*	-0.001	-0.005	0.023***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
S Gender = Male	0.084***	0.064***	-0.312***	0.211***	0.042***	-0.255***
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Raven Score	0.085***	0.072***	0.034	-0.045***	0.008	0.127***
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Age in months	0.005	0.002	-0.004	-0.001	0.002	0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.014	-0.023**	-0.028**	0.017	0.000	-0.029**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working Mother	0.001	-0.018	-0.056	0.007	0.018	-0.081***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)
Low SES	-0.092**	-0.068*	0.005	0.034	-0.029	-0.154***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)
High SES	0.049*	0.033	-0.038	-0.052	0.028	0.073**
	(0.03)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)
N	1004	1003	976	976	976	976

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variable: Income, a binary variable for occupations that are 0.72 or higher in SEI2010, Socioeconomic Index measured by Duncan's (1961). STEM, a binary variable for occupations that are classified as stem or stem-related in Standard Occupational Classification. Social, a binary variable if data value for social skills in O*NET system is 4 or higher. Technical, a binary variable if data value for technical skills in O*NET system is 4 or higher. University+, binary variable if occupation has the majority of university and postgraduate degree worker ratio in O*NET system.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challenge seeking behavior is constructed from a real effort task.

Table 3.11: Subgroup Analysis by Gender - Occupational aspirations by Student

	Higher Education		Prestige		Income		STEM	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Risk attitudes	-0.013*	-0.006	0.014	0.002	0.000	-0.013	0.005	-0.016
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Patience	0.006	-0.000	-0.004	0.001	0.024*	-0.004	0.011	0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Willingness to compete	0.027	0.056*	-0.002	0.056	0.053	0.076*	0.039	-0.019
	(0.02)	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)	(0.06)	(0.03)
Relative Self-confidence	0.013	-0.007	0.009	0.001	0.035	0.010	0.032	0.020
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Challenge seeking behavior	0.002	0.022**	-0.001	0.029**	-0.005	0.030**	-0.006	0.022
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Raven Score	0.003	0.045***	0.013	0.052**	0.021	0.047	0.018	0.058**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Age in months	0.003	0.001	0.004	0.000	0.007	-0.003	0.003	-0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.021***	0.009	-0.008	-0.041***	-0.002	-0.044**	0.011	0.013
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Working Mother	0.008	0.006	0.077	0.034	-0.007	0.085*	-0.033	0.028
	(0.03)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Low SES	-0.012	-0.054	-0.062**	-0.078	-0.078	-0.079	-0.057	-0.083
	(0.02)	(0.04)	(0.03)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)
High SES	0.023	0.016	-0.005	0.026	-0.025	-0.003	-0.060	0.050
	(0.03)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
N	504	484	510	494	509	494	509	494

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01

Dependent Variables: Higher Education, Binary variable, is equal to 1 if the student wants to study further after high school. Prestige, Binary variable for the top two major category in Standard Occupational Classifications. This category includes Management Occupations, Business and Financial Occupations, Computer and Mathematical Occupations, Architecture and Engineering, Life, Physical, and Social Sciences, Community and Social Services, Legal Occupations, Education, Training, and Library, Arts, Design, Entertainment, Sports and Media, Healthcare Practitioners and Technicians. Income, a binary variable for occupations that are 0.72 or higher in SEI2010, Socioeconomic Index measured by Duncan's (1961). STEM, a binary variable for occupations that are classified as stem or stem-related in Standard Occupational Classification. Social, a binary variable if data value for social skills in O*NET system is 4 or higher. Technical, a binary variable if data value for technical skills in O*NET system is 4 or higher. University+, binary variable if occupation has the majority of university and postgraduate degree worker ratio in O*NET system.

Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challange seeking behavior is constructed from a real effort task.

Table 3.12: Subgroup Analysis by Gender - Occupational aspirations by Student

	Social		Technical		Fame		University+	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Risk attitudes	-0.021	0.011	0.017*	-0.014	0.007	0.003	0.015	0.004
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
Patience	-0.002	0.016	0.000	-0.011	0.001	-0.003	-0.002	0.013
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Willingness to compete	-0.011	-0.048	0.005	-0.045	-0.025	0.098**	-0.000	-0.068
	(0.04)	(0.04)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)
Relative Self-confidence	-0.002	0.030	0.014	-0.022	-0.010	-0.004	0.013	-0.001
	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)
Challenge seeking behavior	-0.008	-0.022**	0.002	0.008	-0.001	0.013	-0.005	0.024*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Raven Score	-0.038	-0.065**	0.030**	0.042	0.034**	-0.020	0.013	0.092***
	(0.02)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)
Age in months	-0.009*	-0.002	0.003	0.000	0.005	0.003	0.002	-0.000
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	0.061**	0.035**	-0.020**	0.008	-0.031*	-0.056***	-0.007	0.018
	(0.03)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Working Mother	-0.104***	-0.016	-0.011	-0.006	0.055*	0.029	0.059	-0.039
	(0.04)	(0.06)	(0.03)	(0.05)	(0.03)	(0.06)	(0.05)	(0.05)
Low SES	0.022	0.061	-0.045	-0.060	-0.072**	0.002	-0.057**	-0.066
	(0.06)	(0.05)	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.06)
High SES	-0.025	-0.047	0.001	0.090**	0.038	-0.028	0.012	0.060
	(0.05)	(0.06)	(0.02)	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)
N	510	494	510	494	510	494	509	494

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01

Dependent Variables: Higher Education, Binary variable, is equal to 1 if the student wants to study further after high school. Prestige, Binary variable for the top two major category in Standard Occupational Classifications. This category includes Management Occupations, Business and Financial Occupations, Computer and Mathematical Occupations, Architecture and Engineering, Life, Physical, and Social Sciences, Community and Social Services, Legal Occupations, Education, Training, and Library, Arts, Design, Entertainment, Sports and Media, Healthcare Practitioners and Technicians. Income, a binary variable for occupations that are 0.72 or higher in SEI2010, Socioeconomic Index measured by Duncan's (1961). STEM, a binary variable for occupations that are classified as stem or stem-related in Standard Occupational Classification. Social, a binary variable if data value for social skills in O*NET system is 4 or higher. Technical, a binary variable if data value for technical skills in O*NET system is 4 or higher. University+, binary variable if occupation has the majority of university and postgraduate degree worker ratio in O*NET system.

Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challange seeking behavior is constructed from a real effort task.

Table 3.13: Subgroup Analysis by Gender - Occupational aspirations by Teacher

	Prestige		Income		STEM	
	Girls	Boys	Girls	Boys	Girls	Boys
Risk attitudes	-0.026	-0.003	0.012	0.008	0.004	0.005
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Patience	0.034**	-0.004	0.008	0.008	0.008	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Willingness to compete	-0.012	0.146***	-0.016	0.106***	-0.002	0.049
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Relative Self-confidence	0.061***	0.030	0.020	0.010	0.006	-0.017
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Challenge seeking behavior	0.021*	0.010	0.022	0.014	0.019*	0.024**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Raven Score	0.087***	0.175***	0.057**	0.114***	0.039**	0.106***
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Age in months	0.007	0.003	0.006	0.005	0.002	0.002
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Siblingship size (0-5)	-0.034*	-0.032*	-0.024	-0.004	-0.027	-0.020
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Working Mother	-0.116***	-0.029	-0.035	0.031	-0.038	-0.003
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Low SES	-0.116**	-0.195***	-0.083	-0.103*	-0.034	-0.100**
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
High SES	0.120**	0.110**	0.057	0.039	0.047	0.020
	(0.06)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)
N	510	494	510	494	510	493

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variables: Prestige, Binary variable for the top two major category in Standard Occupational Classifications. This category includes Management Occupations, Business and Financial Occupations, Computer and Mathematical Occupations, Architecture and Engineering, Life, Physical, and Social Sciences, Community and Social Services, Legal Occupations, Education, Training, and Library, Arts, Design, Entertainment, Sports and Media, Healthcare Practitioners and Technicians. Income, a binary variable for occupations that are 0.72 or higher in SEI2010, Socioeconomic Index measured by Duncan's (1961). STEM, a binary variable for occupations that are classified as stem or stem-related in Standard Occupational Classification. Social, a binary variable if data value for social skills in O*NET system is 4 or higher. Technical, a binary variable if data value for technical skills in O*NET system is 4 or higher. University+, binary variable if occupation has the majority of university and postgraduate degree worker ratio in O*NET system.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challange seeking behavior is constructed from a real effort task.

Table 3.14: Subgroup Analysis by Gender - Occupational aspirations by Teacher

	Social		Technical		Fame		University+	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Risk attitudes	-0.014	-0.005	0.006	-0.019	-0.000	-0.004	-0.032	-0.000
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Patience	0.017	0.001	-0.003	-0.003	-0.001	-0.006	0.039**	-0.010
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Willingness to compete	-0.012	0.018	-0.026	0.028	-0.024	0.034	-0.007	0.119***
	(0.05)	(0.05)	(0.03)	(0.04)	(0.02)	(0.03)	(0.05)	(0.04)
Relative Self-confidence	-0.038	-0.029	0.018	-0.032	0.004	0.020	0.055***	0.020
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)
Challenge seeking behavior	0.009	0.022	-0.007	0.006	0.008	-0.017*	0.021*	0.018*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Raven Score	0.044	0.024	-0.021	-0.065***	-0.009	0.025*	0.072***	0.183***
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)
Age in months	-0.005	-0.003	0.000	-0.001	0.002	0.002	0.010**	-0.000
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Siblingship size (0-5)	0.006	-0.059***	0.021*	0.010	0.002	0.001	-0.032*	-0.027
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Working Mother	-0.031	-0.082	0.024	-0.021	0.025	0.010	-0.116***	-0.044
	(0.05)	(0.06)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)	(0.04)
Low SES	0.039	-0.022	-0.004	0.074*	-0.052	-0.024	-0.091*	-0.219***
	(0.06)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
High SES	-0.025	-0.041	-0.045	-0.053	0.036*	0.020	0.088	0.059
	(0.05)	(0.06)	(0.04)	(0.05)	(0.02)	(0.03)	(0.06)	(0.04)
N	483	493	483	493	483	493	483	493

Standard errors clustered at the teacher(classroom) level. * p<0.10, ** p<0.05, *** p<0.01
Dependent Variables: Prestige, Binary variable for the top two major category in Standard Occupational Classifications. This category includes Management Occupations, Business and Financial Occupations, Computer and Mathematical Occupations, Architecture and Engineering, Life, Physical, and Social Sciences, Community and Social Services, Legal Occupations, Education, Training, and Library, Arts, Design, Entertainment, Sports and Media, Healthcare Practitioners and Technicians. Income, a binary variable for occupations that are 0.72 or higher in SEI2010, Socioeconomic Index measured by Duncan's (1961). STEM, a binary variable for occupations that are classified as stem or stem-related in Standard Occupational Classification. Social, a binary variable if data value for social skills in O*NET system is 4 or higher. Technical, a binary variable if data value for technical skills in O*NET system is 4 or higher. University+, binary variable if occupation has the majority of university and postgraduate degree worker ratio in O*NET system.
Student characteristics: Student gender, Raven IQ score, student's age, siblingship size, mother's employment status, student's risk attitude, student's patience, willingness to compete, relative self-confidence, challenge seeking behavior. Willingness to compete and relative self-confidence are constructed from an incentivized competition game. Risk attitude is the amount invested in the risky option in a lottery (ranged from 0-5). Patience is constructed from an intertemporal allocation task (ranged from 0-5). Challange seeking behavior is constructed from a real effort task.

Table 3.15: Reported Grades as a channel for Aspirations: Mediation Analysis for Boys

PANEL 1: Student Reports

	Competition			Challenge Seeking		
	Prestige	SES	STEM	Prestige	SES	STEM
ACME	0.003	0.003	0.003	0.020**	0.017**	0.018**
	(0.54)	(0.54)	(0.56)	(0.00)	(0.02)	(0.04)
ADE	0.047	0.064	-0.029	0.070*	0.073*	0.050
	(0.26)	(0.17)	(0.48)	(0.09)	(0.08)	(0.22)
TOTAL	0.050	0.067	-0.026	0.090**	0.089**	0.068
	(0.23)	(0.16)	(0.51)	(0.04)	(0.02)	(0.12)
Percentage Mediated (%)	6.0%	4.5%	-11.5%	22.2%**	19.1%**	26.5%

PANEL 2: Teacher Reports

	Competition			Challenge Seeking		
	Prestige	SES	STEM	Prestige	SES	STEM
ACME	0.010	0.008	0.007	0.054***	0.044***	0.039***
	(0.47)	(0.56)	(0.59)	(0.00)	(0.00)	(0.00)
ADE	0.115***	0.085**	0.033	-0.023	0.000	0.038
	(0.00)	(0.03)	(0.28)	(0.52)	(0.99)	(0.22)
TOTAL	0.125***	0.093**	0.040	0.031	0.044	0.077**
	(0.00)	(0.03)	(0.24)	(0.42)	(0.27)	(0.02)
Percentage Mediated (%)	8.0%	8.6%	17.5%	174.2%	100.0%	50.6%**

ACME: Average causal mediation effect, ADE: Average direct effect. G.R.: Gender role. Estimates (standard deviation effects) and robust standard errors, in parentheses, are obtained via Imai et al. (2010). The estimation sample is restricted to boys only. Number of simulations is 500. * p<0.10, ** p<0.05, *** p<0.01

Table 3.16: Reported Grades as a channel for Aspirations: Mediation Analysis for Girls

PANEL 1: Student Reports						
	Competition			Challenge Seeking		
	Prestige	SES	STEM	Prestige	SES	STEM
ACME	0.000	0.000	-0.001	0.005	0.015**	0.01*
	(0.96)	(0.94)	(0.86)	(0.12)	(0.02)	(0.07)
ADE	-0.006	0.041	0.025	-0.010	-0.028	-0.021
	(0.86)	(0.38)	(0.6)	(0.65)	(0.54)	(0.62)
TOTAL	-0.006	0.040	0.024	-0.005	-0.012	-0.011
	(0.86)	(0.38)	(0.62)	(0.82)	(0.79)	(0.76)
Percentage Mediated (%)	0.0%	0.0%	-4.2%	-100.0%	-125.0%	-90.9%

PANEL 2: Teacher Reports						
	Competition			Challenge Seeking		
	Prestige	SES	STEM	Prestige	SES	STEM
ACME	-0.003	-0.002	-0.001	0.048***	0.03***	0.023**
	(0.91)	(0.91)	(0.97)	(0.00)	(0.00)	(0.01)
ADE	-0.007	0.005	0.004	0.001	0.024	0.022
	(0.89)	(0.88)	(0.89)	(1.00)	(0.44)	(0.42)
TOTAL	-0.009	0.004	0.004	0.049	0.053*	0.045
	(0.84)	(0.88)	(0.91)	(0.24)	(0.09)	(0.12)
Percentage Mediated (%)	33.3%	-50.0%	-25.0%	98.0%	56.6%*	51.1%

ACME: Average causal mediation effect, ADE: Average direct effect. G.R.: Gender role. Estimates (standard deviation effects) and robust standard errors, in parentheses, are obtained via Imai et al. (2010). The estimation sample is restricted to girls only. Number of simulations is 500. * p<0.10, ** p<0.05, *** p<0.01

Appendix

C.1 Aspiration Categories of Occupations

11-0000 Management Occupations

13-0000 Business and Financial Operations Occupations

15-0000 Computer and Mathematical Occupations

17-0000 Architecture and Engineering Occupations

19-0000 Life, Physical, and Social Science Occupations

21-0000 Community and Social Services Occupations

23-0000 Legal Occupations

25-0000 Education, Training, and Library Occupations

27-0000 Arts, Design, Entertainment, Sports, and Media Occupations

29-0000 Healthcare Practitioners and Technical Occupations

31-0000 Healthcare Support Occupations

33-0000 Protective Service Occupations

35-0000 Food Preparation and Serving Related Occupations

37-0000 Building and Grounds Cleaning and Maintenance Occupations

39-0000 Personal Care and Service Occupations

41-0000 Sales and Related Occupations

43-0000 Office and Administrative Support Occupations

45-0000 Farming, Fishing, and Forestry Occupations

47-0000 Construction and Extraction Occupations

49-0000 Installation, Maintenance, and Repair Occupations

51-0000 Production Occupations

53-0000 Transportation and Material Moving Occupations

55-0000 Military Specific Occupations

Bibliography

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the chicago public high schools. *Journal of labor Economics*, 25(1):95–135.

Alan, S., Boneva, T., and Ertac, S. (2016). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit (ssrn scholarly paper no. id 2761390). rochester, ny: Social science research network. *Rochester, NY: Social Science Research Network*.

Alan, S. and Ertac, S. (2015). Patience, self-control and the demand for commitment: Evidence from a large-scale field experiment. *Journal of Economic Behavior & Organization*, 115:111–122.

Alan, S. and Ertac, S. (2018a). Fostering patience in the classroom: Results from a randomized educational intervention. *Journal of Political Economy*.

Alan, S. and Ertac, S. (2018b). Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment. *Journal of the European Economic Association*. forthcoming.

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender stereotypes in the classroom and effects on achievement. *Review of Economics and Statistics*.

Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, volume 4, pages 1–181. Elsevier.

Andersson, B.-E. (1992). Effects of day-care on cognitive and socioemotional competence of thirteen-year-old swedish schoolchildren. *Child development*, 63(1):20–36.

Andreassen, C. and Fletcher, P. (2007). Early childhood longitudinal study, birth cohort (ECLS–B) psychometric report for the 2-year data collection (NCES 2007–084). *Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education*, 84.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–56.

Antecol, H., Eren, O., and Ozbeklik, S. (2014). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics*, 33(1):63–89.

Baker, M., Gruber, J., and Milligan, K. (2005). Universal childcare, maternal labor supply, and family well-being. Technical report, National Bureau of Economic Research.

Baydar, N., Küntay, A., Gökşen, F., Yağmurlu, B., and Cemalcılar, Z. (2010).

Bedard, K. and Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, pages 1437–1472.

Belsky, J. (1986). Infant day care: A cause for concern? *Zero to three*.

Belsky, J. and Rovine, M. J. (1988). Nonmaternal care in the first year of life and the security of infant-parent attachment. *Child development*, pages 157–167.

Bertrand, M. (2011). New perspectives on gender. In *Handbook of labor economics*, volume 4, pages 1543–1590. Elsevier.

Bertrand, M. and Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, 55(1):3–21.

Berument, S. and Guven, A. (2010). Turkish expressive and receptive language test: Receptive vocabulary sub-scale (TIFALDI-RT). *Turkish Psychological Society*.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review*, 95(2):152–157.

Bietenbeck, J. (2014). Teaching practices and cognitive skills. *Labour Economics*, 30:143–153.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory*, 97(2):298–319.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics*, 93(2):455–467.

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development*, 78(1):246–263.

Blau, F. D., Ferber, M. A., and Winkler, A. E. (2013). *The economics of women, men and work*. Pearson Higher Ed.

Borghans, L., Golsteyn, B. H., Heckman, J. J., and Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, 113(47):13354–13359.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2):123–142.

Brooks-Gunn, J. and Duncan, G. J. (1997). The effects of poverty on children. *The future of children*, pages 55–71.

Buchmann, C., DiPrete, T. A., and McDaniel, A. (2008). Gender inequalities in education. *Annu. Rev. Sociol*, 34:319–337.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Cabinet Office (2011). Opening doors, breaking barriers: A strategy for social mobility.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, volume 3, pages 1801–1863. Elsevier.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Carrell, S. E. and West, J. E. (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–79.

Chowdry, H., Crawford, C., and Goodman, A. (2011). The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16. *Longitudinal and Life Course Studies*, 2(1):59–76.

Crawford, C., Dearden, L., and Meghir, C. (2010). When you are born matters: the impact of date of birth on educational outcomes in england. Technical report, IFS working papers.

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

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

Cuthbert, C. and Hatch, R. (2008). Aspiration and attainment amongst young people in deprived communities: analysis and discussion paper. *DCSF, Cabinet Office Social Exclusion Task Force, Short studies*.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, 102(5):1927–56.

Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, 25(1):43–62.

Datar, A. and Gottfried, M. A. (2014). School entry age and children's social-behavioral skills evidence from a national longitudinal study of us kindergartners. *Educational evaluation and policy analysis*, page 0162373714547268.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human resources*, 42(3):528–554.

Deming, D. and Dynarski, S. (2008). The lengthening of childhood. Technical report, National Bureau of Economic Research.

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School choice, school quality, and postsecondary attainment. *American Economic Review*, 104(3):991–1013.

Dobkin, C. and Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review*, 29(1):40–54.

Domino, G. (1971). Interactive effects of achievement orientation and teaching style on academic achievement. *Journal of Educational Psychology*, 62(5):427.

Duckworth, A. L. and Seligman, M. E. (2005). Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological science*, 16(12):939–944.

Duckworth, A. L. and Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, 98(1):198.

Duncan, O. D. (1961). A socioeconomic index for all occupations. *Class: Critical Concepts*, 1:388–426.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. Random House Digital, Inc.

Dweck, C. S., Davidson, W., Nelson, S., and Enna, B. (1978). Sex differences in learned helplessness: Ii. the contingencies of evaluative feedback in the classroom and iii. an experimental analysis. *Developmental psychology*, 14(3):268.

Eccles, J. (2009). Who am i and what am i going to do with my life? personal and collective identities as motivators of action. *Educational Psychologist*, 44(2):78–89.

Elder, T. E. and Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers. *Journal of Human Resources*, 44(3):641–683.

Eyberg, S. M. and Robinson, E. A. (1983). Conduct problem behavior: standardization of a behavioral rating. *Journal of Clinical Child & Adolescent Psychology*, 12(3):347–354.

Farre, L. and Vella, F. (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica*, 80(318):219–247.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across oecd countries. *oxford review of Economic Policy*, 21(3):416–438.

Fredriksson, P. and Öckert, B. (2006). Is early learning really more productive? the effect of school starting age on school and labor market performance. Technical report, Working Paper, IFAU-Institute for Labour Market Policy Evaluation.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from new york city public schools. *Journal of Labor Economics*, 31(2):373–407.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Goodman, A. and Gregg, P. (2010). *Poorer children's educational attainment: How important are attitudes and behaviour?* Joseph Rowntree Foundation York.

Grossman, M. (2006). Education and nonmarket outcomes. *Handbook of the Economics of Education*, 1:577–633.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880):1164–1165.

Gutman, L. and Akerman, R. (2008). *Determinants of aspirations [wider benefits of learning research report no. 27]*. Centre for Research on the Wider Benefits of Learning, Institute of Education, University of London.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education review*, 30(3):466–479.

Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, 2:1051–1078.

Harris, D. N. and Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7-8):798–812.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

Heckman, J. J. and Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3):446–493.

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

Hidalgo-Cabrillana, A. and Lopez-Mayan, C. (2015). Teaching styles and achievement: Student and teacher perspectives. Technical report, Barcelona Graduate School of Economics.

Hill, A. J. and Jones, D. B. (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review*, 64:1–12.

Hoffmann, F. and Oreopoulos, P. (2009). A professor like me the influence of instructor gender on college achievement. *Journal of Human Resources*, 44(2):479–494.

Hogan, A. E., Scott, K. G., and Bauer, C. R. (1992). The adaptive social behavior inventory (asbi): A new assessment of social competence in high-risk three-year-olds. *Journal of Psychoeducational Assessment*, 10(3):230–239.

Hout, M., Smith, T. W., and Marsden, P. V. (2015). Prestige and socioeconomic scores for the 2010 census codes. *Methodological Report MR124, Chicago, NORC. http://gss. norc. org/get-documentation/methodological-reports.*

ILO (2016). Analysis of istanbul labour market from a gender equality perspective. international labour organization.

Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, pages 51–71.

Jackson, C. K. et al. (2016). What do test scores miss? the importance of teacher effects on non-test score outcomes. Technical report, National Bureau of Economic Research, Inc.

Jeon, M. and Rabe-Hesketh, S. (2012). Profile-likelihood approach for estimating generalized linear mixed models with factor structures. *Journal of Educational and Behavioral Statistics*, 37(4):518–542.

Joensen, J. S. and Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1):171–198.

Johnston, D. W., Schurer, S., and Shields, M. A. (2013). Maternal gender role attitudes, human capital investment, and labour supply of sons and daughters. *Oxford Economic Papers*, 66(3):631–659.

Koch, A., Nafziger, J., and Nielsen, H. S. (2015). Behavioral economics of education. *Journal of Economic Behavior & Organization*, 115:3–17.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? evidence from a natural experiment. *Journal of public Economics*, 92(10-11):2083–2105.

Lavy, V. and Sand, E. (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases. Technical report, National Bureau of Economic Research.

Lubotsky, D. and Kaestner, R. (2016). Doskills beget skills'? evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood. *Economics of Education Review*, 53:194–206.

Machin, S., Marie, O., and Vujić, S. (2011). The crime reducing effect of education. *The Economic Journal*, 121(552):463–484.

McEwan, P. J. and Shapiro, J. S. (2008). The benefits of delayed primary school

enrollment discontinuity estimates using exact birth dates. *Journal of Human Resources*, 43(1):1–29.

McMillan, J. H., Myran, S., and Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4):203–213.

Mühlenweg, A., Blomeyer, D., Stichnoth, H., and Laucht, M. (2012). Effects of age at school entry (ase) on the development of non-cognitive skills: Evidence from psychometric data. *Economics of Education Review*, 31(3):68–76.

Network, N. E. C. C. R. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, 74(4):976–1005.

NICHD, N. E. C. C. R. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, 74(4):976–1005.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.

OECD (2013). Talis user guide for the international database.

OECD (2015). *OECD Family Database*. OECD, Paris.

Olivetti, C., Patacchini, E., Zenou, Y., et al. (2016). Mothers, peers and gender identity. *Boston College, Department of Economics, Working Paper*, 904.

Oreopoulos, P. (2007). Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of public Economics*, 91(11-12):2213–2229.

Paglin, M. and Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1, Part 1):123–144.

Puhani, P. A. and Weber, A. M. (2008). *Does the early bird catch the worm?* Springer.

Quaglia, R. J. and Cobb, C. D. (1996). Toward a theory of student aspirations. *Journal of research in rural education*, 12(3):127–132.

Rampino, T. and Taylor, M. P. (2013). Gender differences in educational aspirations and attitudes. Technical report, ISER Working Paper Series.

Raven, J. C. et al. (1938). *Raven's progressive matrices*. Western Psychological Services Los Angeles, CA.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Robinson, J. P. and Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, 48(2):268–302.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2):247–252.

Sadker, M. and Sadker, D. (2010). *Failing at fairness: How America's schools cheat girls*. Simon and Schuster.

Schoon, I. (2006). *Risk and resilience: Adaptations in changing times*. Cambridge University Press.

Schoon, I., Parsons, S., and Sacker, A. (2004). Socioeconomic adversity, educational resilience, and subsequent levels of adult adaptation. *Journal of Adolescent Research*, 19(4):383–404.

Schulz, J. F. and Thöni, C. (2016). Overconfidence and career choice. *PloS one*, 11(1):e0145126.

Schwerdt, G. and Wuppermann, A. C. (2011). Is traditional teaching really all that bad? a within-student between-subject approach. *Economics of Education Review*, 30(2):365–379.

Smith, J. (2009). Can regression discontinuity help answer an age-old question in education? the effect of age on elementary and secondary school achievement. *The BE Journal of Economic Analysis & Policy*, 9(1).

Smith, T. W. and Son, J. (2014). *Measuring Occupational Prestige on the 2012 General Social Survey*. NORC at the University of Chicago.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, 35(1):4–28.

Stipek, D. (2002). *At what Age Should Children Enter Kindergarten?: A Question for Policy Makers and Parents*. Society for Research in Child Development.

Stipek, D. and Byler, P. (2001). Academic achievement and social behaviors associated with age of entry into kindergarten. *Journal of Applied Developmental Psychology*, 22(2):175–189.

Taylor, B. A., Dearing, E., and McCartney, K. (2004). Incomes and outcomes in early childhood. *Journal of Human Resources*, 39(4):980–1007.

Terrier, C. (2015). Giving a little help to girls? evidence on grade discrimination and its effect on students' achievement. Technical report, Centre for Economic Performance, LSE.

Willitts, M., Anderson, T., Tait, C., and Williams, G. (2005). *Children in Britain: findings from the 2003 Families and Children Study (FACS)*. Number 249. Corporate Document Services.

Yatchew, A. (2003). *Semiparametric regression for the applied econometrician.* Cambridge University Press.