Gender Stereotypes in the Classroom and Effects on Achievement *

Sule Alan, University of Essex

Seda Ertac, Koc University

Ipek Mumcu, University of Essex

January 2017

Abstract

We study the relationship between elementary school teachers' beliefs about gender roles and the educational outcomes of their students. To do this, we exploit the institutional features of a unique setting where conditional on district teachers are allocated to catchment areas, and conditional on catchment areas students are allocated to classrooms randomly. In this setting, a teacher might teach a pupil for as short as 1 year and as long as 4 years. Using rich data on student, family and teacher characteristics, we show that female students 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 on test scores. Our results imply that gender stereotypes in the classroom can contribute to gender achievement gaps early in childhood, and may have implications for gender gaps in occupational choice and labor markets.

JEL Categories:

Keywords: gender stereotypes; gender role beliefs; achievement; gender achievement gap

*Preliminary and Incomplete. Please do not quote without authors' permission. We would like to thank TUBITAK and the Turkish Academy of the Sciences (TUBA) for generous support. We are grateful to numerous research assistants who helped us collect the data. All errors are our own.

1 Introduction

Stereotypical beliefs about gender, which rigidly characterize differences between sexes and appropriate gender roles, are pervasive in most societies. These deep-seated beliefs reflect the society's expectations that individuals behave and make choices in accordance with their gender. Examples of gender stereotyped beliefs and how they are reflected in behaviors and attitudes abound. They can be expressed in parents choosing rigidly-gendered toys for their children or steering them (subtly or blatantly) toward certain educational and occupational choices that are deemed to be suitable for their gender. They can be expressed in social backlash against or disapproval of certain behaviors on the part of women, such as aggressiveness or ambition. To the extent that they are ingrained in a society and influence the actual choices and outcomes of individuals, these beliefs may in large part contribute to gender-achievement gaps as well as the underrepresentation of women in top executive positions and in leadership. As ample evidence suggests, such gender inequality, factually confirming and perpetuating traditional gender role beliefs, can be quite persistent (Bertrand (2011), Bertrand and Hallock (2001), Blau, Ferber and Winkler (2002), Fortin (2005)).

The formation of gender role beliefs and conforming behaviors and attitudes likely begin very early in childhood, and have a significant role in the development of cognitive and noncognitive skills. Through the intergenerational transmission of preferences and beliefs, families have the earliest, most direct impact on the development of children's cognitive and social development (Bisin and Verdier (2000)). The channels of such transmission could be outright didactic teaching (indoctrination) or a more indirect influence via parenting styles (including the language used with the kids) and providing role models. Examples for the former would be iterating what girls should and should not do, and discouraging male-typed behavior (girls are taught to be more obedient, kind and sharing), and examples for the latter would be a family where household chores are females' responsibility and males have more say in important decisions. Intergenerational transmission in the family may have important implications for children's attitudes, behavior and outcomes. 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 (2007), Johnston et al. (2014)).

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 more importantly, with another adult, the teacher. Teachers' views toward gender roles may affect students' behaviors and real outcomes through a variety of mechanisms.

First, a teacher's beliefs may influence students' own beliefs: the teacher may simply express his/her views in the classroom, and because he/she is a significant authority figure, students may be likely to adopt and internalize what the teacher says. A more direct mechanism is teachers' differential interactions with students of different genders. A teacher with strongly traditional gender role beliefs may think that acquiring academic skills is not as important for girls since they will be unlikely to put them into practice later in life. Such a teacher may reflect these beliefs in actual classroom practices through differences in giving feedback on performance, answering/dismissing questions, and even grading exams (Lavy (2008)). Similarly, traditional teachers may act according to their beliefs that math is boys' strong suit, and organize teaching practices accordingly. A very progressive teacher, on the other hand, may exert extra effort to engage girls in subjects that are typically considered male-dominant and try to break stereotypical attitudes in the classroom. The influence of teacher beliefs may affect achievement, especially longer term outcomes, by affecting the development of key noncognitive skills.[1] For example, a progressive teacher may boost girls' confidence (known to be lower than boys') and cultivate their interest/curiosity in male-dominated areas. Girl's performance (especially in math) can also be affected by the teacher's gender role beliefs through stereotype threat–if they feel that their teacher thinks math is boys' strong suit, they may experience anxiety with respect to conforming to the negative stereotype about girls, which indeed translates into lower performance (Spencer et al. (1999)).

The objective of this paper is to test whether teachers' gender role beliefs affect student outcomes. Using rich data from a large-scale field study involving approximately 4000 students and their teachers, we study the effect of teachers' gender role beliefs on their students' gender role beliefs and test scores in math and language. In order to identify the effects, we exploit the unique institutional features of our study site, Istanbul, Turkey. The educational system in Turkey provides us with three main advantages: First, all state elementary school students are allocated to their teacher in grade 1, randomly. Second, teachers are appointed to schools centrally by the Ministry based on the need for teachers and as such, they cannot self-select into catchment areas and schools. Finally, the general practice is such that students have the same teacher from grade 1 to 4, and this is disrupted only with rotations due to need and re-appointments. While the first and the second feature ensure that teachers do not self-select into schools or catchment areas and students are randomly assigned to a particular teacher within a school, the final feature allows us to study the moderating effect of being taught by

[1] It is well-known that non-cognitive skills in childhood are predictive of many important outcomes over the life cycle (Almlund et al. (2011))

the same teacher for a long time.

The data reported in this paper are collected as part of a large field study, with the specific goal of exploring the effects of teacher beliefs on student beliefs and outcomes. To measure gender role beliefs we use survey data collected from both teachers and students by physically visiting all classrooms. For each individual we construct a gender stereotype score using a battery of item-set questions based on a 4-point Likert scale. In addition, we have information on a rich set of covariates including student, family and teacher characteristics as well as measures of cognitive and non-cognitive skills for students. We show that teachers' stereotyped beliefs have a significant negative impact on girls' math and verbal test scores. This effect is bigger, the longer the time students are taught by the same teacher. In particular, a one standard deviation increase in teacher stereotypes leads to a 0.22 and 0.11 standard deviation decrease in math and turkish test scores, respectively, if the teacher has been teaching the student for at least 3 years, controlling for student IQ, gender role attitudes of family, SES and teacher characteristics such as gender and teaching effort. We estimate no significant effect on boys' test scores. We also document that that students' own gender role beliefs are significantly correlated with their teachers' beliefs, even after controlling for student, family and teacher characteristics. This correlation appears to be stronger for female students.

This is the first paper that shows that teachers' beliefs and attitudes are important in determining achievement outcomes and gender gaps in those outcomes, as well as shaping the beliefs and attitudes of students. There is a large literature that studies teacher characteristics such as teacher quality on educational attainment. In the context of gender differences in educational outcomes, the role of teacher gender has been an important focus, 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), Hoffman and Oreopoulos (2009), Carrell et al. (2010), Antecol et al. (2012)). The effect of teachers' beliefs and attitudes with respect to gender roles, however, has largely been overlooked. Our unique setting allows us to identify these effects by using the duration of teacher contact with students. The large dataset, since it was collected with the sole purpose of this paper, is detailed, rich and allows us to control for a host of important teacher, student and family characteristics that are not available in other data sets. The paper is the first to show that teachers' beliefs about gender affect the test scores of their students (controlling for teacher gender), and may play an important role in deepening or mitigating gender achievement gaps.

Mitigating gender differences in economic behavior and outcomes is a major policy question. It is

well-known that women are less likely to select into STEM careers, and more likely to drop out. Gender gaps in educational attainment, occupational choice, labor market outcomes and leadership can cause significant inefficiencies, especially when able women select out of or otherwise cannot reach top positions. Policies such as affirmative action, quotas, educational programs targeting non-cognitive skills, mentoring and single-sex schooling have been proposed to mitigate some of these gaps (e.g. Booth and Nolen (2012), Niederle and Vesterlund (2011), Alan and Ertac (2016)). It has been documented that at least some of the observed differences are due to the cultural environment. In the context of the current paper, differences in educational performance in mathematics have been linked to differences in sociocultural measures of gender equity (Hyde and Merk (2009)). Our results highlight that the classroom environment, in particular the type of teacher, is an important part of the individual's social environment and already starts influencing girls' performance at the elementary school level. The results broadly suggest that implementing gender-equal classroom practices early on could prevent gender gaps in achievement that possibly cause multiplicative effects on academic persistence, occupational selection and labor market outcomes later in life.

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 Istanbul, therefore it primarily represents Turkey's lower socio-economic segment, with some variation with respect to socio-economic status. Most of our schools are located in relatively remote and deprived areas of Istanbul.

In studying the possible impact of teachers' beliefs on the actual outcomes of students, one faces two major selection issues. First, 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), rendering an association between teacher beliefs and student outcomes.

Second, even if students are randomly allocated into classrooms within a school, if teachers could choose the area and the school they wish to work in, a correlation between the teacher's beliefs and students' achievement outcomes may still arise. For example, a particularly progressive teacher with

respect to gender roles may want to work in a school that conforms with his/her views, which may mean working in higher income areas where students tend to have better outcomes. In the next two subsections, we detail some specific features of the Turkish education system in terms of allocation of students to teachers and the appointment of teachers to public schools, and argue that these features largely mitigate both channels of selection.

Apart from selection, another potential problem 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 Turkey. 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 in Section 2.2), we have substantial 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.

2.1 Allocation of Students to Teachers

Catchment areas are determined by the district education directorates and any student whose family resides within the catchment area is entitled to go to one of the designated schools. Catchment areas generally contain 1 to 3 schools within a district, and in our data, there is only one school in the majority of the catchment areas. District sizes substantially vary in Istanbul, with 83 schools in one of our districts and over 300 in another. School size (number of classrooms, teachers, administrators) depends on the population of the district.

After the registration of all first-graders in a given academic year, school administrators randomly allocate the students to teachers. Classroom sizes are not allowed to exceed 50, although a maximum of 30 is preferred. The common practice for the allocation of students to teachers (classes) is to conduct a draw in the presence of parents (conditional on balanced gender composition), with some schools applying last name alphabetical order or national id number. What is important for our analysis is that, contrary to the private system, there is very little room for parents to choose the teacher in the state system.[2] Once students are allocated to classrooms in grade 1, re-mixing in later grades is

[2]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

extremely rare, which means that students remain with the same classmates until they graduate from elementary school, unless the family moves, of course.

2.2 Allocation of Teachers to Schools

As of 1989, the minimum educational requirement to become a teacher was raised from a 2-year college degree to a 4-year university degree in education or any other branch. 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.[3] The appointment process is quite involved and often painful for the teachers in the public system. A new teacher has typically no say in which city, let alone district or school she will be appointed to. In 2015, amongst over 300,000 new teachers, only 40,000 were appointed. Every year an increasing number of teachers remain unappointed, waiting for the next round of appointments. Given this, it is possible to state that teachers have no bargaining power in choosing which area or school they will work in.

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 assigned points of 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 the 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.

The first appointment of a teacher is usually to a school in a remote region of the country, where service points can be accumulated faster. It is generally very difficult to be appointed to one's preferred city before 10 to 15 years, except for pure luck. Teachers cannot ask to be re-appointed before completing at least 3 years (over 5 years in actual practice) of service in their current school. After 3 years, they can ask to be re-appointed to another school, emphasizing the district of their choice if they are already in the city of their choice. This request is 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. In general, for an Istanbul teacher's re-appointment request to be honored

area residents. Sending the child to a school that requires transportation is costly, therefore extremely rare in this socio-economic group.

[3]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.

(in terms of her district choice), she/he must have about 20-25 years worth of service. Even with such a long service, it is extremely hard to be appointed to the generally desired (high SES) districts in Istanbul. Our interviews with teachers reveal that except for a few older teachers, none of them are currently working in the school (or catchment area) of their choice. In fact, about 80% of our teachers state that they are not working in the district of their choice, and this percentage is about 95% among the teachers who are younger than 50. This is not surprising, as our sample mainly covers relatively remote and deprived schools in Istanbul.

After his/her appointment, a teacher can be re-appointed to another school (generally within the same city) if i) there appears to be 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 to another district is honored (only after at least 3 years of service in the current school). A classroom may lose its teacher because of retirement and resignations but the most common reason is rotation due to excess supply and re-appointment to another school based on teacher request. Note that when a teacher is re-appointed to 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 also 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 (most do in our sample) do so mainly to work in the district of their choice, the centralized system largely rules out self-selection of teachers into catchment areas and schools conditional on district.[4] It is important to note that in Istanbul districts are typically very large and there is considerable variation in socio-economic status within each district as well as across districts. For example, a district of a given socio-economic status may house low as well as high socio-economic-status catchment areas. It is this 'within-district' variation that we exploit in order to overcome the self-selection of teachers into schools of a particular socio-economic segment. Providing a compelling quasi-experiment, random allocation of teachers to catchment areas conditonal on district renders identification possible in our study.

[4]We asked our teachers who stated that they are content with the district they work in, whether they would like to move to another area within the district. All said yes but they also told us that given the grim conditions for appointments, they do not see any chance of their request being honored so they do not bother. Note that the most important issue for most teachers is the hardship of commuting in Istanbul from one district to another. Once this problem is solved, they are likely to be content with their location.

3 Data

Our data are collected as part of a large-scale field project, which has been in effect since 2013. In collaboration with the Istanbul Directorate of Education, the project aims to study the behaviors, attitudes and outcomes of students in conjunction with the behaviors and attitudes of teachers. All student data are collected by the authors of the paper by physically visiting all classrooms multiple times.[5] For the experimental tasks as well as completing surveys, children are provided incentives in the form of small gifts that would be of interest to children of these ages.

The project officially covers 165 elementary school teachers and their over 5000 students in 3rd and 4th grades in public elementary schools located all across Istanbul. We collect information using a rich battery of tools, which include surveys, incentivized preference elicitation tasks and official grade records as well as mathematics and verbal tests that we prepared and conducted in the classroom. This endeavor amounted to having to visit each classroom multiple times to minimize disruption to daily teaching activities. Our analysis is based on the teachers and students for whom we have complete information on key variables, which lowers the number of teachers to 147 and the number of students to approximately 4000. Below, we provide a detailed account of the content of our data.[6]

3.1 Student and Family Characteristics

To account for the role of student and family characteristics in determining academic achievement, we collected 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, that 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 assessments 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 socio-economic status and the behaviors and attitudes of the parents. These include information such as the presence of a computer at home, gender roles attitudes in the family, mother's

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

[6] We should note that the reason for missing data is non-attendance on the days when experiments and surveys were conducted for children. There tends to be about a 20% non-attendance on each day due to sickness or other valid excuses.

education, mother's and father's involvement in the student's academic life, parenting styles etc.[7]

3.2 Outcome Measures

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 tests were prepared (and extensively piloted) by the authors of the paper by consulting a separate set of teachers, at a level tailored to each grade that is present in our data. Students' cognitive ability is measured via Raven's Progressive Matrices. We also have access to students' official math, Turkish and behavior grades, all given by their own teachers. Finally, we have experimentally elicited measures of economic preferences such as risk and time preferences.

3.3 Teacher Characteristics

The primary purpose of the paper is to show the effect of the teacher's gender role beliefs on students' outcomes. However, we acknowledge that these beliefs are likely to be correlated with some underlying teacher characteristics, which are important for student achievement. Without adequately controlling for teacher quality, even in the absence of selection, it is hard to give the association between beliefs and achievement outcomes causal interpretation. While there is consensus that teacher quality matters a lot for achievement over and above student characteristics (cognitive and noncognitive skills) and family background, it has proven to be very difficult to measure it. This is possibly because it is multidimensional, often involving unobservables such as teaching styles, effort and care. Acknowledging this difficulty, we make a serious attempt to 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 of our teachers. Teachers' styles of teaching the 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)). Using item-set questions directed to teachers, we construct 4 distinct teaching style variables. We call these constructive vs. traditional, growth vs. fixed mindset, warm vs. distanced, and extrinsic vs intrinsic motivator. The full inventory we use to construct each style score is given in the appendix.

[7]We did not attempt to collect these information directly from parents, as our previous experience show that the response rate of the parents are 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.

Second, we collect information on teachers' extra-curricular activities that focus on teaching improvement and student achievement. 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 socio-economic segment we cover. We take the reported frequency of these volunteer activities as measures of teacher effort.

3.4 Measuring Gender Role Beliefs

Our variable of interest is a measure of the teachers' gender role beliefs. In order to explore the mechanisms behind the effect of these beliefs on student performance, we also measure students' beliefs. For both students and teachers, we construct a "gender stereotype score", using a battery of item-set questions based on a 4-point Likert scale. Items given to teachers and students are the same and given in the appendix. Some examples are "It is much more important for boys to go to university than girls", "Men are better at math than women", which are to be answered using the scale of "I strongly agree", "I agree", "I disagree" and "I strongly disagree". Figure 1 shows the distribution of the stereotype score of children and teachers, with larger numbers representing more traditional views. In both pictures we see substantial variation in gender role beliefs, with men (male students and male teachers) reporting more traditional views. We should note that we took great care to ensure that the teacher was not present when the students filled in the questionnaires, to avoid potential demand effects.

3.5 Descriptive Statistics

We first give a brief description of our student and teacher data. Table 1 presents the descriptive statistics of the variables we have for students, separately for boys and girls. A number of interesting

gender differences are worth mentioning here. First, in terms of test scores, while there appears to be no gender math gap in the unconditional statistics, there is a significant gender gap in verbal test scores in favor of girls. In addition, we observe that girls score significantly better in a fluid IQ test (Raven's progressive matrices). Official grades also show a significant gender gap in verbal ability and disciplinary conduct in school in favor of girls.

Table 2 presents the descriptive statistics of the variables we have for teachers. Most of our teachers are female (76%) and the age range is very wide (23 to 63), with a mean age of 39. A large proportion of our teachers have a 4-year university degree and the majority of them have education as their main study branch. While we do not see a significant gender difference with respect to our teaching style and teacher effort measures across gender, male teachers appear to hold more traditional views with respect to gender roles (p-value=0.02).

Our primary purpose is identifying the effect of teacher gender role beliefs on student achievement outcomes (test scores), and we are aware that these beliefs are likely to be a proxy for a host of individual characteristics of teachers. The results in Table 3 attest to this. About 50% of the variation in the teacher's gender stereotype score can be explained by their characteristics such as qualifications, teaching styles and effort. Interestingly, gender and teaching experience are not significant predictors of teachers' gender role beliefs. The most prominent predictor of these beliefs is teaching style: a joint test of all style measures is decisively rejected with a p-value of 0.0001. We now turn to estimating the effect of teacher gender role beliefs on the actual and behavioral outcomes of students.

4 Results

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

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

where y_{ikc} is an outcome measure (test scores or behavioral outcomes) for student i, who is being taught by teacher k in catchment area c. The variable $Exposure$ captures the number of years student i has been taught by teacher k. The variable GRB_{kc} is the continuous 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 outcomes with respect to the length of

exposure to the teacher. Matrix $\mathbf{X_1}$ contains student characteristics such as age, cognitive ability and noncognitive skills, matrix $\mathbf{X_2}$ contains family characteristics and socio-economic indicators, and $\mathbf{X_3}$ contains teacher characteristics such as gender, experience, education, teaching styles and effort. We also control for catchment area fixed effects (δ).

To allow for non-linearities and facilitate an easier interpretation, we divide the exposure variable into 3 groups: Children who have been taught by the participating teacher for at most 1 year are labeled as "short-term exposed", those who have been taught more than 1 year and at most 3 years are labeled as "medium-term exposed" and those who have been taught for more than 3 years (at most 4 years) are labelled as "long-term exposed".[8] As mentioned before, we have substantial variation in exposure due mainly to teacher relocations and to a lesser extent, family relocations. About 69% of our 3rd graders are labelled as medium-exposed, and about 54% (24%) of our fourth graders are labelled as long term (medium term). Overall, about 27% of the students were exposed to the same teacher for one or two semesters only (short-term exposed). Note that only the fourth grade students can be taught by the same teacher for more than 3 years in our sample, therefore our results regarding long-term exposure relates to the fourth graders.

Given the random allocation of students to teachers, the random allocation of teachers to schools conditional on district, and the fact that we allow for catchment area fixed effects, we are inclined to give the coefficient estimates α, which are the estimates of interest, causal interpretation. However, despite our efforts of collecting information on teachers, we are cautious about the possibility that teacher gender role beliefs may still be capturing some unmeasured teacher quality. We will discuss this in conjuction with our heterogeneity results in the next subsection, which we believe, largely mitigate this concern.

4.1 Gender Role Beliefs of Teachers and Student Achievement

Teachers' gender role beliefs may affect student achievement directly and/or indirectly. An example for the former channel is that a teacher with strong traditional gender role beliefs may think that acquiring academic skills is not as important for girls, since they will be unlikely to put them in practice later in life. Such a teacher may reflect these beliefs in actual classroom practices through providing feedback, answering questions and even grading exams. Similarly, a very progressive teacher may exert extra effort to engage girls in subjects that are typically considered male-dominant and try to break

[8] Because of the small sample size with respect to teachers in 3-year exposure we are not able to divide "medium exposure" further.

stereotyped attitudes in the classroom. Teachers' beliefs may also indirectly affect students' outcomes through their effects on students' non-cognitive skills such as self-confidence and perseverance, which have been shown to be associated with achievement. Before discussing such channels further, we first explore whether the teacher's beliefs indeed have an impact on student outcomes.

Table 4 presents the regression results where the dependent variable is standardized math scores. In all specifications (columns 1 to 6), we use catchment area fixed effects and cluster standard errors at the classroom (teacher) level.[9] Column 2 presents the results controlling for children's characteristics, such as age, cognitive ability, risk attitude and patience. Column 3 adds to that family characteristics such as socio-economic status indicators. Finally, columns 4 to 6 add teacher gender, teacher qualifications, teaching styles and teacher effort. We see here that the gender role beliefs of the teacher have significant impact on math test scores, but not unless the student is exposed to the teacher for a rather long time. The negative (positive) impact of a traditional (progressive) teacher is striking when the student has been taught for more than 3 years by the same teacher. While we cannot reject the equality of the effects of medium- and long-term exposure to a particular teacher, the point estimates clearly suggest a monotonic relationship with respect to the time spent with the teacher. Considering the last column, where all available student, family and teacher characteristics are controlled for, a one standard deviation increase in teachers' stereotyped beliefs leads to about a 0.14 standard deviations decline in test scores if the student has been exposed to the same teacher for more than 3 years. On the positive side, these results also suggest that spending the entire primary school years (4 years) with the same progressive teacher is expected to raise achievement considerably.[10].

Similar, albeit weaker results are obtained using verbal test scores. As can be seen in Table 5, a one standard deviation increase in teachers' stereotyped beliefs leads to about a 0.09 standard deviations decline in test scores if the student has been exposed to the same teacher for more than 3 years. Note, however, that the estimated effect does not reach statistical significance once teacher effort is controlled for.

We next turn to the question of whether the teacher's beliefs affect girls' and boys' outcomes differently. 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

[9] All our analysis can be carried out with school fixed effects however we have 1 school with only one classroom and such an analysis amounts to losing this school. As described above catchment areas are very narrowly defined and schools in these area are very similar to each other. Nevertheless, we did repeat all our analyses using school fixed effects instead and all our results hold. Results are available upon request.

[10] The table also shows that having the same teacher for a long time has a strongly positive effect on test sores.

aspect of teacher quality. If, although we control for teacher effort, beliefs still proxied teacher quality, we would estimate similar effects on both genders. However, as presented in Table 6, this is not the case. Here, we re-estimate our full specification (the last column of the previous tables) separately for boys and girls. What is clearly seen in this table is that the estimated effects we find for both math and verbal scores are mainly driven by girls. The impact of teachers' gender role beliefs on girls' math test scores is of considerable size, particularly when the girls have been taught by the same teacher for more than 3 years. A one standard deviation increase in teachers' gender stereotyped beliefs lowers girls' test scores in mathematics by about 0.22 standard deviations. There seems to be no statistically significant effect on boys and while the equality of medium exposure-gender stereotype score interaction coefficients across genders is not rejected (p-value=0.73), the equality is rejected for long-term exposure interactions (p-value=0.05). A similar finding holds true for the verbal scores (columns 3 and 4). Again, the impact of the teacher's stereotyped beliefs on girls' verbal test scores is of considerable size (0.11 standard deviations) and statistically significant at the 10% level.

These results suggest that traditional gender role beliefs on the part of the teacher has a detrimental effect on girls' performance in both mathematics and verbal tests, especially when they have been with the teacher for more than 3 years. No such effect is present for boys.

To get a deeper understanding of the differential effects we estimate, we relax our assumption of a linear parametric model. 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. To do this, we modify our empirical model as follows:

$$y_{ikc} = cons + \mathbf{X_{1,ikc}}\beta + \mathbf{X_{2,ikc}}\gamma + \mathbf{X_{3,kc}}\theta + \delta_c + f(GRB_{kc}) + \epsilon_{ikc}$$

where 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 large numbers of GRB indicate more traditional (stereotyped) beliefs. Figure 2 depicts the results for math test scores. Our findings in the linear models re-emerge in visual clarity in these pictures. Looking at long-term exposure results, one can see the decreasing relationship between gender stereotyped beliefs of teachers and girls' math test scores, while the relationship is slightly increasing but generally flat for boys. While the relationship looks similar for medium-term exposure for girls, the negative relationship we see clearly in the longer term is not as apparent here. For short-term exposure, the relationship is virtually flat

for both boys and girls.

Similarly, as can be seen in Figure 3, the negative functional relationship between the teacher's beliefs and girls' verbal test scores is visible, while, with some exception at the corner (most progressive teachers), the relationship is flat for boys. Consistently with the results presented in column 4 of Table 6, this picture confirms that traditional teachers negatively affect girls' verbal test scores.

Given the strong differential effects on test scores we find across gender, it is natural to ask how teachers' beliefs contribute to the much-cited math and verbal ability gaps across gender. The literature extensively documents a persistent gap in math performance in favor of boys and in in verbal test performance in favor of girls. These gaps are in fact present in our data. Using our full specification, we estimate a 0.18 standard deviations gender gap in math and a -0.06 standard deviations gender gap in verbal scores, both statistically significant (p-value=0.0001 for math and p-value=0.059 for verbal).[11].

For illustrative purposes, as the results for medium exposure are imprecise, we estimate gender gaps for the two types of teachers for two extreme exposure lengths. Figure 4 presents estimated gender gaps for math and verbal scores for teachers whose gender stereotyping score is smaller than the median score (labelled as "progressive") and for those bigger than the median score (labelled as "traditional").[12].The figure provides suggestive evidence that due to their gender-differential effects on math and verbal scores, teacher beliefs have an impact on gender achievement gaps. The top panel of the figure depicts the gender gaps for students who have been taught by the same teachers for a short period and the bottom panel for a long period. It is clear in the top panel that in the common model where a teacher teaches a class for one year only, we observe no gender gap in math scores regardless of the strength of the teacher's beliefs but a strong gender gap (in favor of girls) in verbal scores is apparent in classrooms taught by more traditional teachers.

When the students have been taught by a teacher with strong traditional beliefs for more than 3 years, the gender gap in math is large and significant (in favor of boys). Although we cannot reject the equality of the presented coefficients, the estimated mean gap in classrooms with progressive teachers is about 0.08 standard deviations (p-value=0.375) and with traditional teachers it is about 0.23 standard deviations (p-value=0.025). For verbal scores, it appears that girls' advantage is eliminated by both types of teachers, likely via different reasons. As shown in our non-parametric results very progressive

[11] Recall that we do not have an unconditional gender gap in math scores. See Table 1

[12] Despite the fact that we have a large number of children in our sample, our sample of teachers is not large. This prevents us from investigating the gender gaps at narrower quantiles of teacher beliefs.

teachers increase boys' verbal performance, mitigating the gender gap in verbal scores, which was in favor of girls (see the top right panel of Figure 3). As also shown by the non-parametric results, girls' verbal test scores suffer under traditional teachers, eliminating girls' advantage over boys (see the top left panel of Figure 3).

4.2 Transmission of Gender Role Beliefs of Teachers to Students

Our data suggest that teachers' beliefs regarding gender roles matter for test scores. A natural question now is to what extent teachers' gender role beliefs are adopted by their students. Figure 5 presents the raw correlation between the beliefs of teachers and the beliefs of girls and boys, separately. The figure implies a positive correlation between the mean gender stereotype score of the classroom and the teacher's score for both genders. The overall correlation coefficient is 0.19 with p-value=0.006. The correlation seems to be much stronger for girls than that of boys, with a coefficient of 0.20 (p-value=0.005) for girls and 0.12 (p-value=0.085) for boys.

We now turn to investigate whether the correlation between children's and teachers' beliefs survives once we control for student and teacher characteristics. Table 7 presents regression results where we regress children's gender-stereotype score on that of their teacher's. As we clearly see in the table the correlation of gender role beliefs between students and teachers remains strong even after controlling for student and family characteristics as well as teacher characteristics. Taking our full specification (column 6), a one standard deviation increase in teachers' gender stereotype score is associated with about a 0.12 standard deviations increase in students' score. The immediate question is whether this correlation is moderated by the time spent with the teacher. Put differently, is it the case that longer exposure to the same teacher is required for transmission of gender role beliefs from teachers to students? The answer seems to be no. The correlation between teachers and students appears to come about fairly quickly, and additional years do not seem to add to this. In fact the coefficients on the interaction terms are all negative, suggesting the relationship between students and teachers become weaker as time passes. Note, however, that the coefficient estimates of the interaction terms are generally not statistically significant.

It is worth mentioning a number of interesting results that emerge from these regressions[13]. It appears that students' gender role beliefs can be well predicted by their cognitive and noncognitive skills. For example, a one standard deviation increase in the child's cognitive ability, measured by

[13]Full results showing all coefficient estimates are not reported in this table for the sake of space but are available upon request

Raven's Matrices, is associated with about 0.14 standard deviations decrease in reported stereotyped beliefs. In terms of family background, not surprisingly, we find that students who belong to relatively higher socio-economic status have about 0.16 standard deviations lower stereotyping score. Regarding the effect of teacher characteristics, we find that students who are taught by more qualified teachers, teachers who apply more constructive teaching methods and hold more optimistic beliefs regarding the role of effort in success hold more progressive beliefs. Teacher gender, interestingly, has no significant effect.

We already know that the raw correlation is stronger between teachers and girls in our data. It would be informative to see whether this correlation survives our rich controls. Table 7 presents the results separately for girls and boys. It appears that the correlation of beliefs mainly holds true for girls and it appears in the short term: a one standard deviation increase in teacher's stereotyped beliefs are associated with about 0.15 standard deviation increase in girls' beliefs, and the correlation is significant at 5% level. Although we cannot reject the equality of coefficients across gender (p-value=0.25), the correlation between teachers and boys does not reach statistical significance. It appears that female students internalize their teacher's gender role beliefs fairly quickly. The gender difference we estimate here may be due to the fact that girls tend to be more conforming and obedient at this age[14].

5 Conclusion

We exploit a quasi-experimental setting 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 female students 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 sole purpose of answering the research question we pose in this paper, allows us to control for a host of important teacher, student and family characteristics that are directly relevant. The results show that controlling for the teacher's own gender and other characteristics, a teacher's beliefs about gender roles affect the test scores of their students. These results show that the views of the elementary school teacher may play an important role in mitigating

[14]We do have information on teacher rated obedience and disciplinary conduct measure for all students and find that teachers rate girls significantly more obedient and better behaved than boys.

or widening gender-achievement gaps. Given the importance of the childhood period for long-term choices and outcomes, the results suggest that what type of teacher a child is assigned to in elementary school may have long-lasting consequences. Our results broadly suggest that implementing gender-equal classroom practices early on or training teachers to raise awareness of such biases and their effects could prevent inefficient gender gaps in achievement, occupational selection and labor market outcomes.

References

[1] Alan, S. and Ertac, S. (2014), "Good Things Come to Those Who (Are Taught How to) Wait: Results from a Randomized Educational Intervention on Time Preference", HCEO Working Paper.

[2] Alan, S., Boneva T. and Ertac S. (2016), "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit", HCEO Working Paper.

[3] Almlund, M., A.L. Duckworth, J.J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." Handbook of the Economics of Education, 1–181.

[4] Antecol, H., Eren, O., & Ozbeklik, S. (2012). The effect of teacher gender on student achievement in primary school: evidence from a randomized experiment.

[5] Bettinger, Eric P., and Bridget Terry Long. (2005). "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." American Economic Review , 95(2): 152–157.

[6] Bertrand, M. (2011), "New perspectives on gender." Handbook of labor economics, 4, pp. 1543-1590.

[7] Bertrand, M., & Hallock, K. F. (2001). "The gender gap in top corporate jobs". Industrial & Labor Relations Review, 55(1), 3-21.

[8] Bisin, A., and Verdier, T. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences". Journal of Economic Theory , 97: 298-319.

[9] Blau, F. D., Ferber, M.A., & Winkler, A.E. (2002), The Economics of Women, Men and Work (Englewood Cliffs, NJ: Prentice Hall).

[10] Booth, A. L., & Nolen, P. (2012). Gender differences in risk behaviour: Does nurture matter?. The Economic Journal, 122(558), F56-F78.

[11] Carell, Scott, M. Page, and J. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." Quarterly Journal of Economics , 125(3).

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

[13] Farré, L., & Vella, F. (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. Economica, 80(318), 219-247.

[14] 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.

[15] Hoffmann, F., & Oreopoulos, P. (2009). A professor like me the influence of instructor gender on college achievement. Journal of Human Resources, 44(2), 479-494.

[16] Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. Proceedings of the National Academy of Sciences, 106(22), 8801-8807.

[17] Johnston, D. W., Schurer, S., & Shields, M. A. (2014). Maternal gender role attitudes, human capital investment, and labour supply of sons and daughters. Oxford Economic Papers, 66(3), 631-659.

[18] Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. Journal of public Economics, 92(10), 2083-2105.

[19] Niederle, M., & Vesterlund, L. (2011). Gender and competition. Annu. Rev. Econ., 3(1), 601-630.

[20] Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. Journal of experimental social psychology, 35(1), 4-28.

Tables

	Student				
Variable	Boys	Girls	tstat	min	max
Maths Score	0.01(1.01)	-0.01(0.99)	0.72	-1.56	2.17
Turkish Verbal Score	-0.09(1.01)	0.09(0.98)	6.36	-2.20	1.67
Grade	3.44(0.51)	3.46(0.51)	1.09	3	4
Age (in months)	109.80(8.12)	109.33(7.71)	2.16	66	183
Raven Score	-0.06(1.03)	0.06(0.96)	4.30	-2.85	2.60
School Performance					
Turkish	4.05(1.13)	4.03(0.97)	7.81	1	5
Maths	3.98(1.23)	4.30(1.17)	1.50	1	5
Behavioral	4.30(0.91)	4.58(0.72)	11.31	1	5
Behavior and Attitudes					
Gender S-typing Score	0.15(0.99)	-0.15(0.98)	11.06	-2.47	2.09
Academic Self-Confidence	2.51(0.54)	2.46(0.54)	2.65	1	3
Growth Mindset	0.00(1.03)	-0.00(0.96)	0.12	-3.30	2.12
High Occupational Aspiration	0.71	0.85		0	1
Risky Behavior	0.05(1.00)	-0.05(1.00)	3.35	-1.65	1.70
Impatience	0.03(1.02)	-0.03(0.97)	1.80	-1.34	1.66
Family					
Family Gender Styping	2.33(1.14)	2.28(1.11)	1.58	1	4
SES	1.85(0.74)	1.86(0.75)	0.44	1	3
Low income (Column %)	36.0	36.3			
middle income (Column %)	43.3	41.8			
High income (Column %)	20.7	21.9			
Working Mother	0.29(0.46)	0.27(0.44)	2.04	0	1
Computer	0.76(0.43)	0.74(0.44)	1.67	0	1
Parents are reading to the child	0.29(0.45)	0.28(0.45)	0.95	0	1

Standard errors in parentheses

Table 1: summary stat - by student gender and teacher gender

	Teacher				
Variable	Male Teacher	Female Teacher	Tstat	Min	Max
Gender (%)	23.6	76.4			
Age	41.82(11.79)	38.82(9.09)	1.67	23	63
Years of Teaching	18.05(9.83)	14.48(8.25)	2.52	1	44
# of terms as the class teacher	5.15(2.21)	5.24(2.21)	0.21	1	8
Tenure	1.00	0.89	2.20	0	1
Degree (column %)					
2-year Community College	15.4	4.8			
University	76.9	88.1			
Graduate school	7.7	7.1			
Major (column %)					
Classroom Teacher	84.6	65.6			
Linguistics	0	7.2			
Science & Engineering	7.7	11.2			
Social Sciences	2.6	5.6			
Other	5.1	10.4			
Teacher Effort					
Number of Volunteer Activities	12.69(9.09)	9.93(8.01)	1.82	0	30
Number of Extra Curricular Programs	3.76(5.04)	3.83(4.86)	0.08	0	40
Teaching Styles					
Growth Mindset	14.18(2.77)	14.64(2.29)	1.02	7	20
Extrinsic Motivator	10.37(1.95)	10.64(1.48)	0.92	5	16
Constructive/non-traditional teaching	18.53(2.89)	19.39(2.74)	1.67	11	24
Warm Approach	11.68(2.43)	11.99(2.47)	0.67	5	16
Teacher Gender Role Beliefs	0.24(1.05)	-0.12(0.97)	2.01	-1.91	2.51

Standard errors in parentheses

Table 2: summary stat - by gender of the teacher / gender bias teachers

Table 3: Predictors of Teacher's Gender Role Beliefs

	(1)	(2)	(3)	(4)
Male	0.191	0.213	0.074	0.087
	(0.16)	(0.16)	(0.15)	(0.15)
University Degree	-0.130	-0.209	-0.216	-0.214
	(0.27)	(0.27)	(0.26)	(0.26)
Graduate Degree	-0.488	-0.576	-0.535	-0.524
	(0.37)	(0.37)	(0.35)	(0.34)
Years of Experience	-0.011	-0.015*	-0.003	-0.002
	(0.01)	(0.01)	(0.01)	(0.01)
Tenured	0.531*	0.831**	0.630**	0.581**
	(0.27)	(0.33)	(0.25)	(0.26)
Number of Terms in the School	0.004	0.000	0.004	0.016
	(0.03)	(0.03)	(0.03)	(0.03)
Education Degree		0.552***	0.572***	0.655***
		(0.21)	(0.18)	(0.19)
Linguistics		-0.197	-0.281	-0.267
		(0.29)	(0.28)	(0.28)
Natural Sciences		0.482	0.283	0.304
		(0.30)	(0.18)	(0.20)
Social Sciences		0.168	0.025	0.030
		(0.29)	(0.24)	(0.24)
Growth Mindset			-0.123***	-0.120***
			(0.03)	(0.03)
Extrinsic Motivator			0.063*	0.051
			(0.04)	(0.04)
Constructive Teaching			-0.034	-0.026
			(0.02)	(0.02)
Warm Approach			-0.070***	-0.069***
			(0.02)	(0.02)
Number of Extra_C Programs				0.001
				(0.01)
Number of Volunteer Activities				-0.023**
				(0.01)
Observations	147	147	147	147

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Teacher Gender Role Beliefs and Student Math Score

Teacher G-Styping	0.007	0.036	0.056	0.051	0.004	0.001
	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Medium Term	0.118*	0.092*	0.096*	0.104*	0.096*	0.075
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Long-Term	0.278***	0.230***	0.219***	0.224***	0.207***	0.186***
	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Medium Term*Teacher G-Styping	-0.067	-0.081	-0.113*	-0.110*	-0.103	-0.090
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Long Term*Teacher G-Styping	-0.126	-0.153**	-0.176**	-0.152**	-0.170**	-0.138*
	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)
Catchment Fixed Effect	✓	✓	✓	✓	✓	✓
Student Characteristics		✓	✓	✓	✓	✓
Family Characteristics			✓	✓	✓	✓
Teacher Characteristics				✓	✓	✓
Teaching Styles					✓	✓
Teacher Effort						✓
N	3897	3897	3897	3897	3897	3897

Student characteristics: student gender, age in months, Raven IQ score, self reported confidence, self-reported perseverance, teacher reported behavior score. Family characteristics: student reported gender roles at home, student reported parental involvement in academic life, mother's employment status, teacher reported socio-economic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warmth vs distanced, extrinsic vs intrinsic motivator, traditional vs constructive 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 5: Teacher Gender Role Beliefs and Student Verbal Score

Teacher G-Styping	-0.010	0.016	0.033	0.005	0.003	-0.006
	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
Medium Term	0.077	0.055	0.057	0.036	0.026	0.008
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Long-Term	0.124*	0.088	0.081	0.053	0.034	0.020
	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
Medium Term*Teacher G-Styping	-0.052	-0.060	-0.087	-0.064	-0.066	-0.047
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Long Term*Teacher G-Styping	-0.099	-0.109	-0.128*	-0.115	-0.112	-0.085
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Catchment Fixed Effect	✓	✓	✓	✓	✓	✓
Student Characteristics		✓	✓	✓	✓	✓
Family Characteristics			✓	✓	✓	✓
Teacher Characteristics				✓	✓	✓
Teaching Styles					✓	✓
Teacher Effort						✓
N	3903	3903	3903	3903	3903	3903

Student characteristics: student gender, age in months, Raven IQ score, self reported confidence, self-reported perseverance, teacher reported behavior score. Family characteristics: student reported gender roles at home, student reported parental involvement in academic life, mother's employment status, teacher reported socio-economic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warmth vs distanced, extrinsic vs intrinsic motivator, traditional vs constructive 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 6: Heterogenous Effects of Teacher Gender Role Beliefs on Outcomes

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.019	-0.005	0.072	-0.079
	(0.08)	(0.06)	(0.05)	(0.06)
Medium Term	0.024	0.104	-0.001	0.018
	(0.07)	(0.07)	(0.06)	(0.07)
Long-Term	0.146*	0.245***	0.012	0.027
	(0.08)	(0.07)	(0.07)	(0.07)
Medium Term*Teacher G-Styping	-0.107	-0.086	-0.111*	0.018
	(0.08)	(0.08)	(0.06)	(0.08)
Long Term*Teacher G-Styping	-0.240**	-0.053	-0.159*	-0.017
	(0.10)	(0.09)	(0.09)	(0.08)
Catchment Fixed Effect	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
N	1907	1990	1910	1993

Student characteristics: student gender, age in months, Raven IQ score, self reported confidence, self-reported perseverance, teacher reported behavior score. Family characteristics: student reported gender roles at home, student reported parental involvement in academic life, mother's employment status, teacher reported socio-economic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warmth vs distanced, extrinsic vs intrinsic motivator, traditional vs constructive 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 7: Teacher Gender Role Beliefs and Student Gender Role Beliefs

Teacher G-Styping	0.124**	0.125***	0.113**	0.115**	0.119**	0.117**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Medium Term	-0.027	-0.028	-0.031	-0.020	-0.019	-0.014
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Long-Term	-0.128*	-0.087	-0.081	-0.061	-0.045	-0.038
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Medium Term*Teacher G-Styping	-0.077	-0.094*	-0.076	-0.070	-0.066	-0.066
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Long Term*Teacher G-Styping	-0.070	-0.089	-0.075	-0.071	-0.065	-0.073
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Catchment Fixed Effect	✓	✓	✓	✓	✓	✓
Student Characteristics		✓	✓	✓	✓	✓
Family Characteristics			✓	✓	✓	✓
Teacher Characteristics				✓	✓	✓
Teaching Styles					✓	✓
Teacher Effort						✓
N	3908	3908	3908	3908	3908	3908

Student characteristics: student gender, age in months, Raven IQ score, self reported confidence, self-reported perseverance, teacher reported behavior score. Family characteristics: student reported gender roles at home, student reported parental involvement in academic life, mother's employment status, teacher reported socio-economic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warmth vs distanced, extrinsic vs intrinsic motivator, traditional vs constructive 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 8: Teacher Gender Role Beliefs and Student Gender Role Beliefs: Gender Heterogeneity

	Girls		Boys	
Teacher G-Styping	0.083**	0.148**	0.056*	0.089
	(0.04)	(0.07)	(0.03)	(0.06)
Medium Term		-0.025		-0.003
		(0.06)		(0.06)
Long-Term		-0.086		0.005
		(0.08)		(0.08)
Medium Term*Teacher G-Styping		-0.074		-0.067
		(0.08)		(0.08)
Long Term*Teacher G-Styping		-0.028		-0.121
		(0.08)		(0.08)
Catchment Fixed Effect	✓	✓	✓	✓
Student Characteristics		✓		✓
Family Characteristics		✓		✓
Teacher Characteristics		✓		✓
Teaching Styles		✓		✓
Teacher Effort		✓		✓
N	1913	1913	1995	1995

Student characteristics: student gender, age in months, Raven IQ score, self reported confidence, self-reported perseverance, teacher reported behavior score. Family characteristics: student reported gender roles at home, student reported parental involvement in academic life, mother's employment status, teacher reported socio-economic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warmth vs distanced, extrinsic vs intrinsic motivator, traditional vs constructive 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.

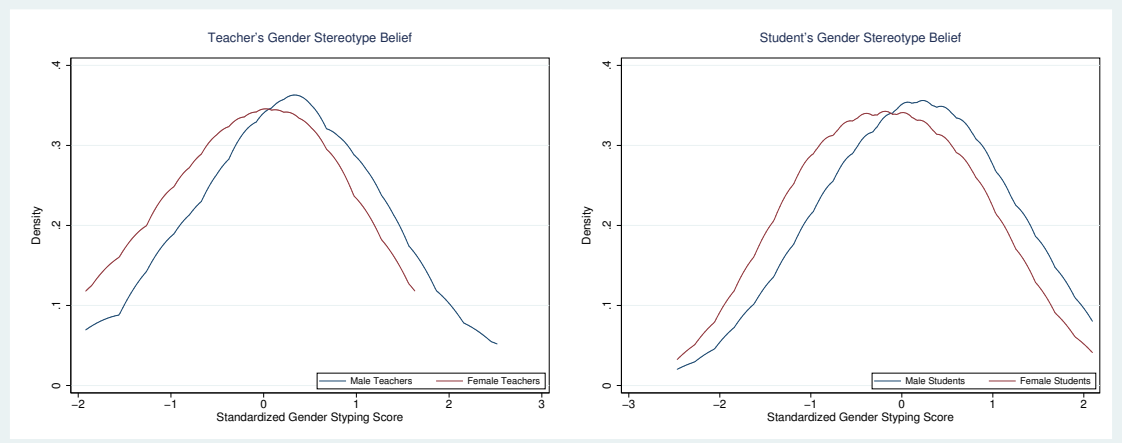


Figure 1: Distribution of Gender Role Beliefs

Figure 2: Teacher Gender Stereotyping and Math Test Scores: Non-parametric

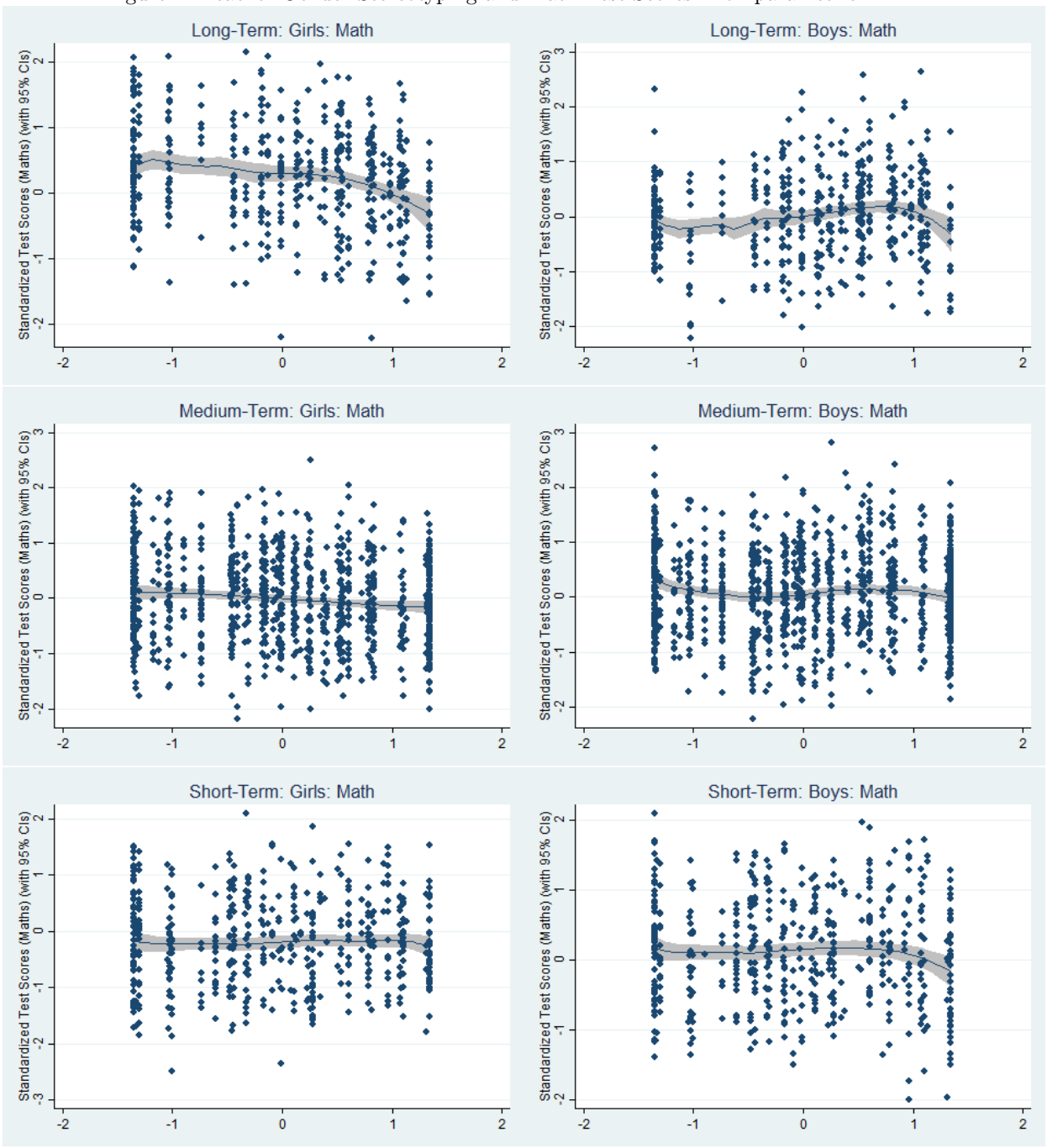


Figure 3: Teacher Gender Stereotyping and Verbal Test Scores: Non-parametric

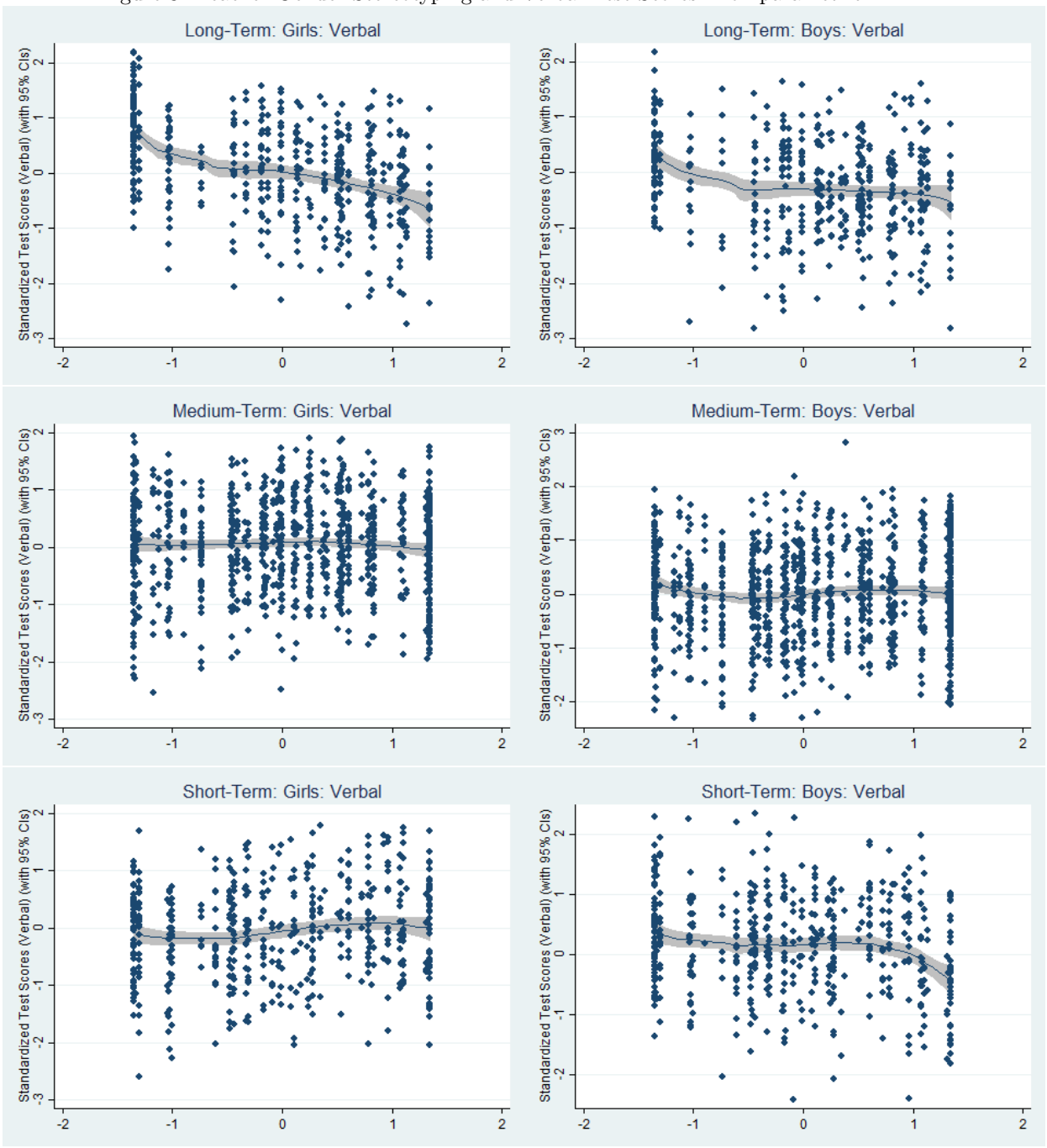


Figure 4: Gender Gap in Math and Verbal Ability

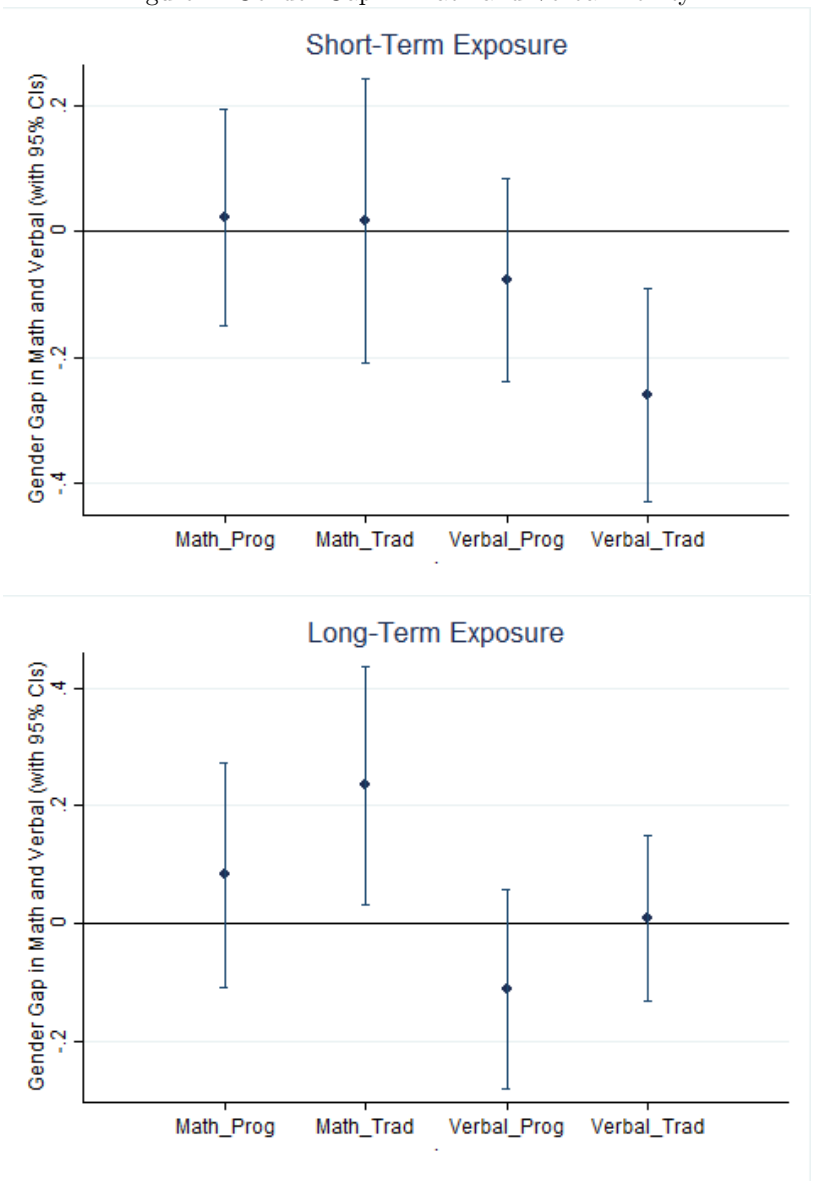
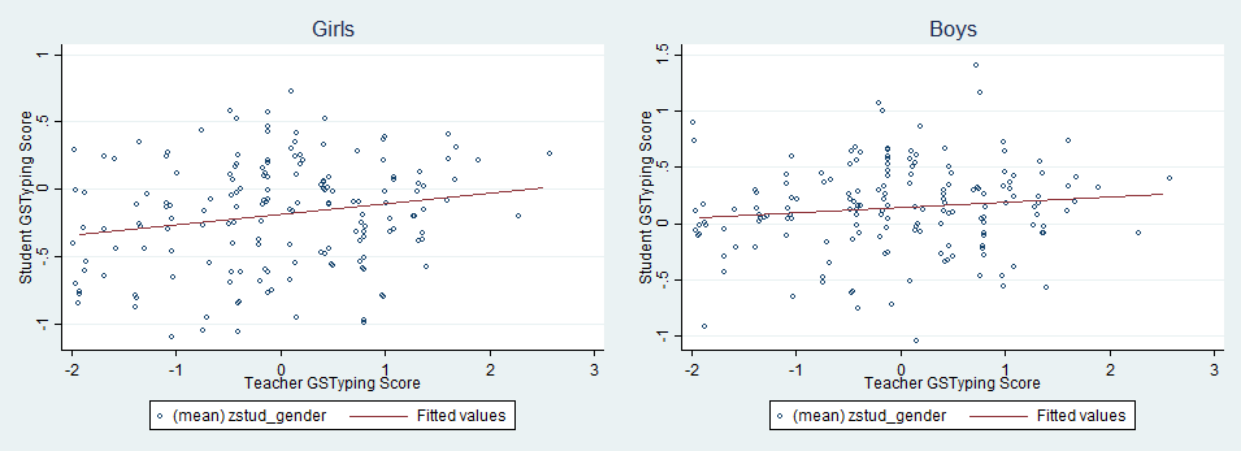


Figure 5: Correlation between teachers beliefs and student's beliefs



Appendix

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 compare to women, hence they are better leaders.

Questions Used for Constructing Teaching Styles

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

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 class.
6. Punishment is necessary to attain a disciplined and ordered classroom environment.
7. I often reward students (applauding, giving stars etc.) to get the outcomes and behaviors that I aim for.
8. I often punish students (grounding them on the breaks, making them sit alone etc.) to get 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.
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.

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.