Unlearning Traditionalism: The Long-Run Effects of Schools on Gender Attitudes

Javier Garcia-Brazales (CEMFI)*

First Version: January 2020. This Version: September 2021

Abstract

Can policy persistently change traditional gender attitudes? I leverage a unique opportunity to provide novel evidence on how exposure to females during pre-adolescence impacts views on gender roles, perceived relative gender ability, and gender-related behaviors up to nine year later. Observing a close-to-nationally-representative set of Vietnamese primary schoolers exogenously allocated to classes within schools I show that there are strong non-linear effects of a higher proportion of female peers on reduced gender traditionalism both for males and females. These attitudinal shifts translate into actual behavior. In line with contact theory, stronger cross-gender interactions — as identified through friendship nominations — lead males to increase their intensive and extensive margins of home production. Females' exposure to more same-gendered individuals reduces the pressure to conform to traditional life goals and raises their self-confidence, professional ambitions, and expected returns to education. This leads to significant increases in enrollment in university and in male-dominated majors. Building upon these results, I propose a simple and easily-implementable policy of class configuration that would significantly decrease overall traditionalism. This stands in stark contrast to the null impact that one year of education has on gender norms as estimated through an RDD exploiting laws of compulsory age of enrollment.

Keywords: Gender Norms, Attitudes Formation, Contact Theory, Long-term Peer Effects, Returns to Education.

JEL Codes: I24; I25; J16.

---

*I thank Manuel Arellano, Dmitry Arkhangelsky, Samuel Bentolila, Paula Bustos, Caterina Calsamiglia, Guillermo Caruana, Micole De Vera, Monica Martinez-Bravo, and multiple participants at Princeton's 2021 YES, Warwick's 9^{th} PhD Workshop, the 2021 EALE Annual Conference, Oxford's 2021 Education Symposium, and CEMFI's seminars for useful comments. The author gratefully acknowledges the funding received from the Fundación Ramón Areces. Address: Calle Casado del Alisal 5, 28014, Madrid, Spain. Mail: javier.garcia@cemfi.edu.es

The data used in this publication come from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India, Peru and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Department for International Development (DFID). The views expressed here are those of the author. They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders.

1 Introduction

An increasing body of literature has shown that traditional gender norms have important consequences on individual and family decisions and that this may translate into intra-household inequalities as well as into lower aggregate economic growth through talent misallocation (Hwang et al., 2019; Hsieh et al., 2019).

In order to correct for these frictions, the use of policy (e.g. gender quotas, affirmative action) has become commonplace.[1] While these measures have been identified as successful in various contexts (e.g. Bagde et al., 2016; Hospido et al., 2019), their effectiveness tends to be limited in space and time: eliminating formal barriers is unlikely to prevent returning to the same unsatisfactory situation if the core cause of the frictions — for instance, the way traditional gender views are formed in the first place — is not directly addressed (Bordalo et al., 2016; Bertrand et al., 2020).[2] Overcoming these limitations is particularly challenging given that, to date, there is almost no evidence on how to change norms through small scale policies (Bau, 2021; Field et al., 2021) and that views on gender norms, identity, and abilities are strongly transmitted across generations, are shaped from an early age, and remain rather traditional despite widespread cross-gender daily interactions (Fernández et al., 2004; Fernández and Fogli, 2006).

In this paper I explore the potential of primary schools to affect traditional gender roles in a persistent manner. This is an ex ante ideal context for thinking about effective policy interventions: young children's views are more malleable, their skills present dynamic complementarities, and children can be more easily reached given the virtual universality of compulsory primary education. In order to gauge the overall effect of schools, which is key to inform policy, I assemble a unique dataset that allows me to take a holistic approach in jointly addressing, *for the same sample*, the individual impact on gender attitudes of their three fundamental dimensions: classmates, teachers, and returns to education.

I combine Oxford University's Young Lives detailed longitudinal information tracking the the living conditions and the cognitive, attitudinal, and physical development of a cohort of Vietnamese students from age 1 to 19 with rich information on their Grade 5 peers, teachers, and schools obtained from a complementary, large-scale, and close to nationally-representative project featuring 92 school sites. This allows me to study the impact of early school conditions on a wide range of outcomes (gender attitudes, academic achievement, and home production, inter alia) and the mechanisms in place both in the short and in the long-term (up to 9 years

[1] See, for instance, Beaman et al. (2009).

[2] For more successful examples, see, for instance, Beaman et al. (2009).

later).

More specifically, I address the following, novel, set of questions: (i) does exposure to females (classmates and teachers) at an early age *persistently* impact views on traditional gender roles and perceived relative abilities between genders?; (ii) if the answer is affirmative, is *actual (and meaningful) behavior* carried out by males and females modified as a consequence and, if so, through which *channels*?; (iii) are attitudinal shifts accompanied by academic spillovers? If this is the case, how do the attitudinal and academic impacts compare in terms of *economic relevance and degree of persistence over time*? Finally, (iv) is there room for policy to persistently decrease aggregate traditionalism?

Apart from the unique richness of the dataset constructed, Vietnam[3] provides attractive contextual aspects required for the causal identification of the parameters of interest. First, while the literature on social interactions is subject to important identification issues such as self-selection and unobserved heterogeneity (Manski, 1993), I exploit plausibly exogenous group composition within schools. In particular, teachers reported the method followed to form classes. This allows me to work with the subset of sections where student assignment was "random" (around 80% of the original sample) or possibly exogenous (an additional 13%). Hence, conditional on accounting for students' selection into schools through school fixed effects, I am able to exploit quasi-experimental variation in group gender composition across classes within schools. This largely mitigates the pressing concerns in the literature about endogenous network formation. I provide multiple tests to validate the key identification assumption that the proportion of female peers across classes within schools is not driven by unobserved factors. For instance, I show that the proportion of female students in a class does not systematically correlate with a wide range of individual observable characteristics measured prior to class formation — including child cognition, parental attitudes towards and expectations for their children, and predicted gender attitudes computed from out-of-sample estimations of the relevance of key observables in determining gender views.

Second, in order to estimate the returns of one additional year of education on gender attitudes I employ a regression discontinuity design (RDD) benefiting from the exogenous variation in the age of compulsory enrollment into primary education by date of birth (Singh, 2020).

My main result is that a ten percentage points increase in exposure to female peers in Grade 5, the last year of primary education, reduces the degree of gender traditionalism by over a tenth

[3]This is a country where, despite a (slow) narrowing in gender gaps, women have consistently lagged behind men in many respects. For instance, in 2015, which is around the time when the data employed in my analysis was obtained, there being a list of 77 jobs legally forbidden to women (World Bank, 2015). More generally, Vietnam ranks around the median in the United Nation's Gender Empowerment Measure, a measure of opportunity that combines information on political participation and decision making, economic participation and decision making, and power over economic resources.

of a standard deviation five years later.[4] While the effects on overall gender attitudes are of similar size for male and female students when using a composite attitudinal measure, the wealth of dimensions elicited allows me to break down the impact across its different components and across genders, and hence to identify the attitudinal source of various forms of gender gaps such as labor force participation or home production (e.g. Jayachandran, 2015; Hwang et al., 2019). Among males, the largest changes are in terms of their views on acceptable female life purposes and on the extend to which females and males should be equally treated in their cross-gender interactions. This updating of views is larger when males are exposed to more able female classmates, and is channeled through increased peer-to-peer cross-gender interactions, as measured using friendship nominations. This is consistent with males' biases originating ex-ante, and then being revised positively as interactions with females reveal more accurate information about them. The fact that part of this argument relies on that the individuals in my sample — who are just 10 years of age but have been exposed to their peers since age 6 — already displayed gender-biased views at this early age is in line with findings by Bian et al. (2017) who show that 6 year old children in the USA already have stereotypical views of males being more academically able.

On the female side, girls exposed to more same-gendered peers became more prone to believe that women have similar abilities to males and that they should strive to have a successful professional career over one as a housewife. This highlights exposure to males as a key mechanism in increasing the pressure for females to conform to traditional roles and hence in perpetuating traditional gender norms. The fact that more interactions with other females allows them to move away from traditional views is in line with the overcoming of "stereotype threats" (Zölitz and Feld, forthcoming)[5], particularly in terms of what society imposes as a girls' adequate life goals. I find that this change towards desiring to develop professional careers that are more in line with those undertaken by males is accompanied, as one would expect, by higher academic aspirations, by a positive change in the perceived returns to schooling — this being another novel outcome that I explore[6] — and by a larger willingness to work in less traditionally female-dominated jobs (which is supported by the finding that these females eventually enroll more in male-dominated majors). This last point is at odds with recent results for Denmark in

[4]Given the novelty of my outcome, I compare the magnitude to the gender peer effects estimated for the commonly explored cognition in Gong et al. (2019), and I find it to be of about the same size. Dhar et al. (2018) also get similarly-sized treatment effects of a dummy indicating participation in their intervention of in-school discussion of gender norms.

[5]As females are grouped with more females, the intensity of negative stereotypes that they are exposed to from their peers is expected to decline.

[6]My work contributes to the literature on beliefs about the returns to education (e.g. Stinebrickner and Stinebrickner, 2012; Wiswall and Zafar, 2015), enrollment decisions and preferred field of study by placing the focus on the role of peers, in particular females, as a strong force for changing the expected returns to education through modifying cultural norms and the desired professional outcomes.

Brenøe and Zölitz (2020), who show that females with more female classmates are more likely to undertake girl-dominated majors. I find that these changes in gender attitudes and in the overall perceived value of education — whose relative importance I quantify following Gelbach (2016)'s decomposition — result in these females indeed being more likely to remain enrolled at school by age 19 (i.e. mostly in tertiary education).[7]

The effects on views on gender roles, which are long-lived, are in contrast with those on academic performance in mathematics and Vietnamese. Observing cognitive scores in standardized tests over five years — which have been shown to capture the *quality* of human capital much more accurately during the pre-tertiary education period than simple enrollment status (Singh, 2020), I find that more exposure to female peers improves cognitive abilities contemporaneously but the gains fade away over time. This is consistent with the existing literature, which finds inconclusive results for academic spillovers but more lasting ones for attitudes such as risky behavior or religiosity (Sacerdote, 2011).

Given that I also uncover the presence of nonlinearities[8], together with the fact that potential side-effects would be, if any, of second-order importance based on my ample set of outcomes explored, I propose an operational gender-mixing policy — reshuffling *existing* students in a school so that sections are more likely to reach the 55% and 45% thresholds of the proportion of females in a class (the cutoffs in my analysis of non-linearities), while ensuring that at all sections have at least 40% of females (which I take as a benchmark for gender parity). Kolmogorov-Smirnov tests indicate that the resulting distribution of traditionalism is significantly less regressive than the one arising under strict parity. This finding is particularly remarkable given that, relying on an RDD exploiting variation in education arising from differences in compulsory enrollment time based on date of birth, I show that there are no returns to education in terms of the degree of traditionalism of the views on gender roles. More generally, given that less traditional males regard females as more able to undertake jobs with large responsibilities, this could significantly reduce the misallocation of talent to professional careers and be efficiency enhancing (Fortin, 2015; Hsieh et al., 2019; Chiplunkar and Goldberg, 2021).[9] The uncovered shifts in attitudes and socio-emotional skills therefore bring in an addi-

[7]Increased female enrollment under more female-inclusive environments is in line with the finding in Kazianga et al. (2013) who showed this to be the case when building more girl-friendly schools (for instance, with separate toilets) in Burkina Faso.

[8]Under a linear-in-means framework, shifting peer composition can only lead to social gains under nonlinearities. Otherwise, moving individuals around has offseting effects for the two groups involved (Carrell et al., 2013).

[9]Jayachandran (2015) reviews the root causes of gender inequalities in developing countries and concludes that, while low economic growth is an key driver, specific cultural characteristics such as patrilocality also play a role. She also proposes a series of policy approaches to reduce gender bias, either through direct intervention in household-related aspects (e.g. giving financial incentives for parents to invest in girls), or by exposing females to role models that can change their gender views. My work highlights an specific approach within the latter category in that simple exposure to peers can be effective in changing attitudes, and this actually not only impacts females but also males (albeit through different channels).

tional dimension to the cost-benefit analysis of such potential policy that would otherwise be obviated.

My paper advances various strands of the literature. First, while the vast work on academic peer effects arising from exposure to females has focused on cognitive spillovers and, in a handful of cases, on outcomes expected to be driven by gender norms — notably the likelihood of studying a male-dominated major at university (Anelli and Peri, 2019)[10], the pathways through which these outcomes are impacted and, more generally, the effects on non-academic outcomes, have been notably understudied (Lavy and Schlosser, 2011; Gong et al., 2018). I provide the first direct evidence of the causal impact on a novel socio-emotional dimension (gender attitudes) measured through a battery of twelve questions eliciting views on a wide range of aspects such as the perceived leadership abilities of females relative to males or how important it should be that females adhere to traditional life goals (e.g. females as housewives). This is crucial to understand the mechanisms in place behind the impact on any gender-related outcome and hence to inform policy, and I do so at an stage in the children's development — pre-adolescence — that has been recently emphasized as crucial for the development of socio-emotional skills (Choudhury et al., 2006; Ashraf et al., 2020).

Moreover, in as much as the study of the long-run effects of peers is scarce and limited to academic and labor market outcomes in developed countries (Gould et al., 2009; Black et al., 2013; Carrell et al., 2018; Bietenbeck, 2020), my work contributes novel evidence on the persistent effects of peers and group integration on attitude formation, particularly in the much less understood context of developing countries, where gender roles are a fundamental source of frictions (Quinn and Woodruff, 2019; Jayachandran, 2020).

Second, while most research on contact theory has focused on the short-run impact of exposure to minorities (Paluck et al., 2019)[11], my findings emphasize that shifts in attitudes are also present in the long-run, and may arise from a group with which there is already large exposure since birth (females). This last point is also made in Dahl et al. (forthcoming), and is particularly relevant in my context where all classes are co-ed. A key difference with respect to that paper is that, while they focus on cross-gender interactions during a short period of time, my context of interest relies on intense exposure that is maintained over several academic years. As the authors themselves posit, this is likely to be the explanation for why their effects fade away

[10]They find that males exposed to more females during secondary education are not more likely to choose college majors traditionally favored by females nor to have their labor market outcomes impacted.

[11]This paper provides a comprehensive review of existing work on inter-group contact theory from a social psychology perspective. The general finding is that there is a strong positive correlation between inter-group interactions and increased understanding/acceptance, although most work lacks a causal interpretation. The two studies that observe the outcome of interest for the longest period are Mousa (2020) and Dahl et al. (forthcoming), up to 6 months later.

rapidly over time and for why they expect studies exploiting other situations featuring a longer time of exposure to find more persistent impacts — as I do.[12] My work is therefore the first one to identify *persistent* effects in gender norms, which is critical for policy to yield gains across generations (Daruich, 2018). Moreover, since I observe other relevant outcomes (e.g. school enrollment) nine years after class assignment, I am able to study the *real-world implications* of the stated changes in attitudes. Importantly, I do so for novel outcomes in the academic contact theory literature that are closely linked to traditional gender views, such as home production (Hwang et al., 2019; Hyun, 2020).[13]

Third, my paper connects with the literature on attitudes formation. While the literature on gender norms has mostly versed about their presence, transmission, and impact on outcomes, much less evidence exists on *how to change* norms (Field et al., 2021), and even less that speaks to how to do it for members of *both* genders (Dhar et al., 2018). More generally, evidence on the potential for policy to reshape culture is scarce and limited to quantifying the impacts of *large* shocks, such a switch of political regime (e.g. Alesina and Fuchs-Schündeln, 2007), which are prone to affect multiple dimensions of culture. Bau (2021), who recently showed that a relatively small policy — a pension reform — decreases the incentives for maintaining traditional kinship relations, is a notable exception.[14] Another important impediment for the effectiveness of existing policy intervention is the lack of explicit and detailed propositions — as opposed to broader prescriptions such as "increasing education" or "empowering women".

With these limitations in mind, I exploit exogenous variation within real-world environments across a close-to-nationally-representative sample to evaluate whether mechanisms for females' reduced performance/aspirations in minority contexts previously identified in laboratory settings hold under observational data, something not certain a priori (Levitt and List, 2007). The root cause of increased female enrollment into tertiary education that I uncover — a modernization of gender roles — constitutes a complementary explanation for the enrollment impact of female peers that has traditionally focused on self-confidence as a mechanism (e.g.

[12] Other notable differences is that they focus on a male-dominated, very specific context (the army), on adults (instead of children), on males' shifts in attitudes (and not females'), and explore a different set of long-term outcomes and mechanisms.

[13] My study complements two recent pieces of work exploring how being around racial minority individuals (black students in the US) and poor classmates in India can foster acceptance towards these groups in academic settings (Carrell et al., 2019; Rao, 2019) by focusing on direct measures of gender attitudes as well as gender-related outcomes and by carefully dissecting the mechanisms in place. One novel piece of research looking into the effects of experimentally-manipulated teams outside the academic context is Lowe (forthcoming) for cross-caste interactions in sports teams.

[14] Dhar et al. (2018) explore the impact of a school-based intervention consisting of engaging secondary students in India to in-class discussion about gender (in)equality. While their main results on the impact on a self-reported composite measure of gender norms are in line with mine (including the economic magnitude), their work differs from mine in several aspects. Most notably, they evaluate an intervention aiming at changing norms, while I simply rely on variation in cross-gender interactions arising naturally. Furthermore, my study is able to study longer-term attitudes as well as how they translate into actual behavior — and the mechanisms in place. This leads me to suggest a different (although complementary) form of policy interventions. Last, I put the emphasis on schools as a whole by also looking at the role of teachers and the returns to education.

Reuben et al., 2017; Cools et al., 2019). My policy proposal based on the existing nonlinearities is a specific, cost-effective, and implementable one that provides a novel means of favoring intergenerational mobility and that contributes to the raising literature on female empowerment (e.g. Bandiera et al., 2020; Field et al., 2021).

Fourth, my work offers joint causal evidence of the impact of various treatments on socio-emotional and academic outcomes *for the same set of students*, something extremely uncommon in the education literature (Oreopoulos and Salvanes, 2011). It therefore provides important insights that short-term and narrower-scoped studies cannot address, including the direct comparison of the effectiveness of various treatments and the highlighting of the importance of going beyond the exploration of purely academic outcomes. The latter is a relevant result in the non-pecuniary returns to schooling that has implications both for our understanding of sorting as a source of inequalities and for thinking of public policies for class composition that are less likely to underestimate their true social returns.

The rest of the paper is organized as follows. Section 2 introduces in detail the data employed. Section 3 outlines the empirical approach and provides evidence on the validity of the identification assumptions. Section 4 reports my main results on views on gender norms. Section 5 provides evidence on the mechanisms behind the main results. Section 7 probes the robustness of my results to different specifications. Section 8 concludes.

2 Data

2.1 Young Lives

Young Lives (YL) is a longitudinal study led by Oxford University that collects high-quality data on the living conditions, environment, and decisions of two cohorts of children (the younger one was aged 6-18 months at the first round, in 2002) from childhood to early adulthood (the fifth, and last, round was collected in 2016) across four developing countries.

In Vietnam, children from five provinces belonging to four (Northwest, Red River Delta, South Central Coast, and Mekong River Delta) out of the country's eight regions were sampled. In order to reflect the cultural and geographic diversity of the country, twenty sites (four per province, with an oversampling of poor communes) were randomly selected based on governmental rankings of all communes in the province in terms of poverty indicators (quality of infrastructure and incidence of poverty and child malnutrition). Households within each commune who had children aged either 1 or 8 in 2001 (the first year of the YL survey) were then randomly sampled. For more details, refer to Young Lives (2014) and Section C.1.

In total, 2,000 children were sampled for the young cohort. I construct a panel containing, information, among others, on their physical health (e.g. height, weight), cognitive abilities (as measured by age-specific tests), socio-emotional skills (e.g. academic aspirations, self-confidence), time use on a normal weekday, as well as rich household information (e.g. composition, earnings, investments).

YL additionally undertook a School Survey in the 2011-2012 academic year.[15] Its objective was to provide an in-depth study of the academic environment (peers, teachers, and schools) of a subset of YL children. For this, all schools located in any of the twenty sites with at least one YL student enrolled in Grade 5 (age 10) were sampled.[16] For each Grade 5 class, up to twenty randomly selected peers[17] were also surveyed both at the beginning and at the end of the 2011-2012 academic year. Grade 5 is an important point in a child's academic career, as it is the last year of primary education (Figure 1 depicts the Vietnamese education system and how the timing of the surveys relates to it). This should be emphasized in my context because all schools in the sample provide teaching up to Grade 5, a point in time when children transition to different schools to undertake lower secondary education. This means that the class configurations that I observe are dismantled after one year.

The original 2011 School Survey dataset contains 3,284 students distributed across 176 classes in 92 school sites[18]. Among these students, 1,138 belonged to the longitudinal component of the YL survey meaning that, *unlike their peers*, they continued being tracked after the end of the academic year.

I obtain information on YL children and their peers along the following dimensions: (i) sociodemographics: standard questions on the gender (which allows me to compute the leave-out-mean of female peers in the class), age, parental education, ethnicity, and household size were asked. I also compute a wealth and an academic resource indices based on the ownership of various relevant assets (e.g. motorbike, calculator); (ii) cognitive scores: YL designed, distributed, and graded mathematics and Vietnamese tests aiming at evaluating the knowledge of official Grade 5 curricula both at the beginning and at the end of the academic year; (iii) friendship nominations: a unique feature of the school survey is that, at the end of the academic year, every child answered the following two questions with respect to *each and every*

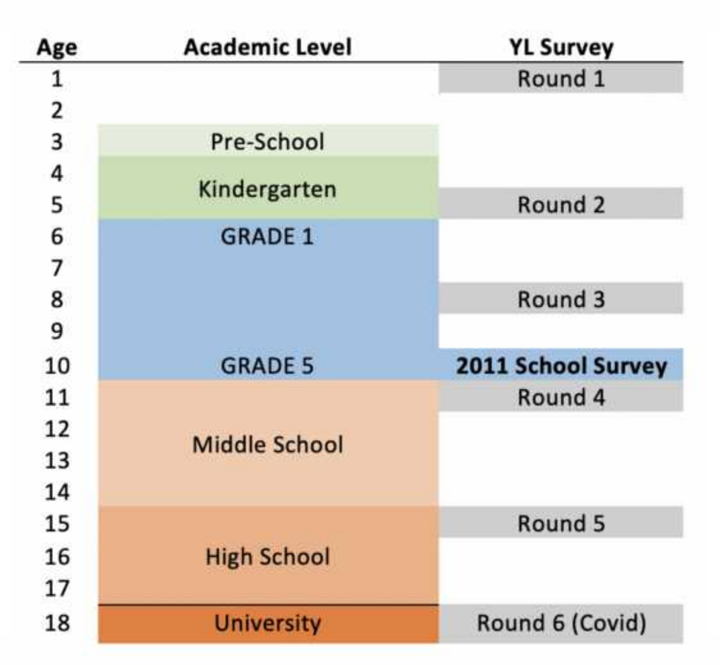
[15] A second one took place in 2016-2017.

[16] The main reason for not being part of the School Survey was that the YL child was born after 31 December, hence enrolling one academic year later.

[17] For instance, if a school had three Grade 5 classes, up to 60 students were sampled. Prior to randomly drawing these peers, all Grade 5 YL children were sampled first to ensure their participation (recall that YL are themselves a random sample of the children born in their cohort).

[18] It is relatively common in rural Vietnam for schools to build satellite locations to cater students living in less accessible areas; these 92 sites correspond to 56 different schools.

Figure 1: Vietnamese Academic System and YL's Timing



Notes. The mapping between age and YL surveys is exact. The mapping between age and academic level is an approximation based on standard school progression. Gray cells indicate that the survey corresponds to the regular longitudinal tracking of the YL young cohort. Round 6 was a phone-based survey undertaken in mid-2020 to gather information on the pre- and post-pandemic living conditions of the YL participants.

surveyed classmate: (1) how would you describe your friendship with this classmate?, and (2) how much do you do things with this classmate outside of school?; (iv) non-cognitive information: school-related personality traits such as self-reported levels of effort and perception of own-ability were elicited, and (v) teacher characteristics: among others, rich information on gender, education, wealth, experience, and locus-of-control (ability to influence children outcomes) were collected. More explicit information is provided in Section C.2.

I link school and peer characteristics for the YL students with their longitudinal information available from the Young Lives young cohort panel in order to construct a unique dataset in terms of the richness of information available and the long time span covered. Additionally, it is crucial for my empirical analysis to provide a more detailed description of both my main outcome of interest (gender attitudes) and the exogenous variation that I exploit in order to identify the causal impact of peers on such outcomes.

Views on Gender Norms. I obtain long-term views on gender norms from a battery of twelve questions elicited in the fifth wave (age 15) of the YL longitudinal survey (the only wave where they are elicited). As such, it is available for all YL students, but not for their classmates in the School Survey. This is, however, not problematic for my causal inference of interest, as

YL students were selected to constitute a random and representative sample of the population.

The children were asked to describe their degree of agreement with gender-related statements in a four-point scale (strongly disagree, disagree, agree, strongly agree). I classify them into four categories: (1) “life purpose”: measures how much females should adhere to traditional goals in their lives as being good housewives or schooling being more important for males; (2) “abilities”: captures whether it is believed that females are less capable than males in cognitive and leadership aspects; (3) “cross-gender interactions”: measures the degree of agreement with females being treated unequally in cross-gender interactions (e.g. not being allowed to play rough sports), and (4) “intra-household decisions” captures the agreement with males having a stronger bargaining position in the household.[19]

To gauge and overall effect, to increase statistical power, and to facilitate interpretation, in my main analysis I compute the average level of agreement towards traditional norms — higher values indicate a more traditional view on the position of females, so first I flip the ordering of the responses when relevant.

Class Assignment. A crucial aspect of the School Survey is that it provides information, as reported by the class teacher, on the way that students were allocated to the section. The responses were as follows: (i) randomly (77.73% of the students); (ii) there was only one Grade 5 class (6.68%); (iii) by location of residence (6.62%); (iv) by general ability (4.75%); (v) by ability in mathematics (2.76%), (vi) by age (1.47%), (vii) by ethnicity (0%); (viii) by ability in Vietnamese (0%), and (ix) other (0%). These figures show a very attractive characteristic of my context of interest: almost 80% of the students were exogenously allocated to classes. This is in line with anecdotal qualitative evidence consistently pointing at the fact that, while there is no country-wide regulation imposing that class allocation is made random, the vast majority of schools during primary education do so, at least during the first grades (Behr, 2005) — as children reach Grade 5 some schools (or selected classes within schools) may start tracking high-ability students in preparation for middle-school (which shows up in 4.75+2.75% of the cases in my sample).[20] I provide detailed evidence on the success of the quasi-experimental allocation of students to classes in Section 3.2.

Sample Selection. As mentioned, my main outcome of interest, long-run attitudes towards gender norms, is only observed in the fifth wave of the YL longitudinal survey. This means that in my baseline analysis I need to restrict my sample to those children in the SS that belonged

[19]The use of the extent to agreement to similar statements is widespread in gender literature. See, for instance, Bertrand et al. (2015) and Goussé et al. (2017).

[20]One may be concerned that school headmasters could manipulate class composition if a good balance on observable characteristics was not achieved in the first place. As long as this reorganization is random, this will simply reduce the within-school variation in characteristics that I observe, but would not bias my results.

to the YL survey (1,138). Additionally, my preferred estimation strategy relies on exogenous assignment of students to classes. Excluding those individuals assigned by ability or by age (which may reflect ability for repeater and grade skippers)[21] and those with small peer groups (classes with less than nine pupils) leaves me with 937 student-level observations distributed across 74 school sites (i.e. including satellite locations[22]) and 152 classes. While this is my main estimating sample, the number of observations is slightly reduced in certain specifications with additional controls due to missing covariates.

2.2 Descriptive Statistics

Table 1 provides descriptive statistics for standard sociodemographic characteristics, for the raw scores in the mathematics and Vietnamese tests in the first round of the school survey, and for my main explanatory variable, the proportion of female peers in the class. Briefly, the sample is evenly split in terms of gender.[23] The average age is 10.29, as expected given that it pertains to Grade 5. The proportion of female peers has an average of 0.47 and a standard deviation of 0.1. Plotting the histogram of the distribution of the proportion of females in a class (not of female peers) in Figure B.1 shows that the highest density in the proportion of female peers is around 0.5 or slightly less (recall that there are more males than females in the society) while extreme proportions (e.g. below 0.3 or above 0.6) are very infrequent[24], as consistent with exogenous allocation of students to classes[25].

In Table 2 I turn to the descriptives of the attitudinal questions elicited at age 15. To ease the reporting, I compress responses into a binary variable taking the value of 1 when the child agreed or strongly agreed with the respective traditional view indicated by the row variable — the separate distributions for male and female children using the four possible responses are provided in Online Appendix Figures B.3 and B.4.[26]

[21] Feld and Zölitz (2017) show how non-random selection based on ability largely biases the estimation of peer effects upwards.

[22] These are blocks of classrooms located at a distance from the main school — 50 out of the original 56 main schools in my sample — in order to facilitate schooling access to remote areas (Rolleston et al., 2013). Indeed, 96% of the main school principals report that the goal of the satellite sites is to enable children in remote areas to attend school. Therefore, given their geographical separation from the main school in my main specifications I include school site fixed effects (i.e. I treat satellite sites as separate from the main school site). The main results do not change if I use school fixed effects instead.

[23] This is consistent with the fact that the sex ratios at birth in 2001, the year of birth of my cohort of interest, were very close to the natural rate of 105 boys per 100 girls (Guilmoto, 2009) and that primary education attendance is compulsory in Vietnam — males and females do not differ in their enrollment rates (Nguyen, 2012).

[24] There is still enough *within-school* variation in the proportion of female peers. In Figure B.2 I plot the distribution of the difference in the maximum and minimum proportion of female peers within schools. One should note, however, that large within-school differences are very infrequent, as consistent with exogenous assignment of students to classes. Finally, it is worth remarking that co-ed classes are universal in the sample.

[25] The same Monte Carlo simulations of student class allocation as performed for Figure 3 confirm the positive skewness of the distribution of the proportion of female students per class (available upon request).

[26] Additionally, Figure B.5 uses World Values Survey (WVS) information from its fifth round (collected between 2005 and 2009) from two comparable statements to those in YL: "university is more important for a boy than for a girl" and "men make better

Table 1: Descriptive Statistics of Selected Variables in the School Survey (Age 10)

	(1) Mean	(2) Standard Deviation	(3) Min	(4) Max	(5) Count
Female	0.49	0.50	0	1	932
Age (years)	10.29	0.25	9.92	10.83	922
Father Can Read	0.96	0.20	0	1	915
Mother Can Read	0.95	0.21	0	1	929
Ethnic Minority (non-Kinh)	0.08	0.03	0	1	932
Home Educational Resources Index	0.72	0.25	0	1	937
Wealth Index	0.51	0.15	0	1	937
No Books at Home	0.19	0.39	0	1	929
1-5 Books at Home	0.25	0.43	0	1	929
6-10 Books at Home	0.15	0.35	0	1	929
More than 10 Books at Home	0.42	0.49	0	1	929
No Health Problem	0.71	0.45	0	1	937
Mathematics Raw Score First Test	18.01	5.57	1	30	930
Vietnamese Raw Score First Test	19.88	5.30	2	30	933
Proportion Female Peers	0.47	0.10	0.21	0.86	925

Notes. Descriptive statistics computed from the estimating sample for the long-run effects on attitudes (74 school sites and 152 classes). All variables are indicators, unless stated otherwise.

Overall, gender norms are rather regressive among the youth. For instance, 62% support the idea that females should aim at being good housewives to the detriment of their professional careers and 42% believe that females lack leadership abilities relative to males. At the same time, in other dimensions there is significantly less support for traditional norms. For example, only 8% of the respondents consider that males and females differ in intelligence.

In Table A.3 I report the same moments of the data for the *old cohort* of the Young Lives data. These children were around 21 years of age at the time of the 2016 interview. One can appreciate that the figures are very similar to those in Table 1, which provides external validity to the attitudes reported by the young cohort.

3 Empirical Approach

3.1 Regression Framework

My main empirical strategy is based on the widely used linear-in-means specification to estimate Equation 1:

$$y_{ics} = \alpha + \beta_1 PF_{-ics} + \beta_2 Female_{ics} + \beta_3 X_{ics} + \beta_4 \overline{X}_{-ics} + \lambda_s + \epsilon_{ics}, \tag{1}$$

where y refers to an outcome of interest (primarily views on gender norms at the YL's 2016 round), and ics denotes person i[27] in class c in school s.

PF measures the proportion (i.e. support is the closed interval 0-1) of female students in the class, excluding the person of reference.[28] β_1, our coefficient of interest, estimates the average effect of sustained exposure to a higher proportion of female classmates (starting from Grade 1 and finishes when children progress to lower secondary school for Grade 6 — i.e. the year after I observe them — since children invariably transition to different schools, often times significantly far from their primary education centers).[29] β_1 therefore captures an "exogenous"

business executives than women do". This allows me to: (i) provide external validity to the attitudinal questions elicited in YL, as the averages for Vietnam from YL and from WVS are very similar — i.e. the level of agreement with the university question is 0.17 in YL and 0.2 in WVS while the level of agreement with leadership positions in 0.42 in both surveys; and (ii) place Vietnam in the international context: it is an average/somewhat-less-traditional Southeast country (similar to China or Japan and significantly less traditional than India and Malaysia). Relative to countries from other continents (as indicated by the color code of the bars), Vietnam tends to be slightly less traditional than African country but much more traditional than America or European (unreported) ones.

[27] Adding a "-" sign in front is short-hand notation for indicating that a variable is computed as the leave-out-mean among the peers of individual i, located in class c at school s.

[28] Classmates are a natural reference group: in Vietnam different sections study in separate classrooms, they remain together for all subjects, and maintain the same teachers throughout the academic year. In Grade 5 the majority of students in my sample stay around 5 hours per day at school (of which close to 4 are devoted to teaching). These are termed "half-day schedules", and are typically complemented with private tuition (either taught by the main teacher at school or one-on-one — Nguyen et al. (2021)). A smaller fraction of students undertake the full day track (about 6 hours of instruction). For more details, refer to Dang and Glewwe (2018).

[29] The headmasters from all 74 school sites in my sample report that the school exclusively teaches grades 1 to 5. It is worth noting

Table 2: Descriptive Statistics of Attitudes towards Gender Norms (YL Round 5, Age 15)

	(1)	(2)	(3)	(4)	(5)
Agree/Strongly Agree with Traditional...	Mean	Standard Deviation	Min	Max	Count
Life Purpose					
(a) Women should focus on being good wives	0.62	0.48	0	1	936
(b) Women should have less freedom	0.08	0.28	0	1	937
(c) Men should be prioritized for college	0.17	0.38	0	1	935
(d) School performance more important for males	0.20	0.40	0	1	936
Abilities					
(e) Men are better leaders	0.42	0.49	0	1	937
(f) Women are not as smart as men	0.08	0.27	0	1	936
Cross-Gender Interactions					
(g) Women should not ask for a date	0.22	0.42	0	1	926
(h) Women should not pay for a date	0.56	0.50	0	1	931
(i) Women should not play rough sports	0.09	0.28	0	1	937
(j) Women should not swear	0.58	0.49	0	1	936
Intra-household Decisions					
(k) Fathers should have more authority in decisions	0.45	0.50	0	1	935
(l) Chores should not be split if both spouses work	0.05	0.22	0	1	937
Mean Score (Full Sample)	2.18	0.32	1.17	3.17	937
Males	2.26	0.31	1.33	3.17	481
Females	2.09	0.31	1.17	2.83	456

Notes. All variables except Mean Score are indicators taking the value one if the child agrees or strongly agrees with a traditional view on gender norms for each dimension of interest. The exact questions were reported in the "Views on Gender Norms (longitudinal YL)" subsection. Mean Score is computed as the average score (on a 1-4 scale) across the twelve dimensions at the individual level.

peer effect, as it arises from background characteristics of the students, and not from malleable dimensions such as their achievement or their behavior (i.e. "endogenous effect").

Although it is well-known that the identification of exogenous peer effects is complicated by several aspects inherent to social network formation, my context of interest is particularly well-suited to deal with these. A first difficulty is that direct comparisons of students across schools is likely to yield biased estimates due to selection into schools — these are Manski (1993)'s "correlated effects". If this was the case, unobserved determinants of a student's gender views would likely also be correlated with her classmates' average characteristics, including the proportion of female peers. For instance, families with more traditional views might sort into locations/schools where other families share such views.[30] The standard way of dealing with this problem is through the inclusion of unit-level fixed effects at a higher level of aggregation than the one at which peer effects are measured. In my case, since I define peers at the classroom level, I include school site fixed effects (λ_s).[31]

The addition of school fixed effects cannot deal, however, with potential sorting of students into classes. For example, it could be the case that schools assign students to classes based on certain student characteristics that correlate with unobserved determinants of our outcome of interest. Another case would be if parents who pay particular care to their child's academic performance exerted some pressure to the school headmaster to place their children with high-ability students. While the existing literature has often exploited plausibly exogenous changes in peers' background characteristics across cohorts within schools in the absence of quasi-experimental group formation, I benefit from direct information on the way sections were formed, which allows me to focus on the ample-majority of schools that exogenously assign students to classes. My main findings will prove robust to multiple demanding checks, including randomization based inference (Imbens and Wooldridge, 2009), coefficient stability to selection on unobservables (Oster, 2019), and removal of influencial observations (Broderick et al., 2020).

that selection out of the school throughout primary education based on peers' composition is very unlikely: using retrospective information on the yearly schooling of the 2,000 children in YL's young cohort, I find that only 4% of YL children change schools at some point between grades 1 and 5. My main results are virtually unchanged when excluding from the estimation children who do transition across schools between grades 1 and 5.

[30] In practice, school choice is not a very sophisticated process. Using information on YL's round 3, I find that 88% of parents claim that the single most important reason for school choice is proximity to the location of residence, while 8% state that it is the good quality of teaching. On the schools' side, 85% of the 52 schools' headmasters in the 2011 School Survey claim that all students who apply get accepted (for those schools where not all students can be accepted, the single main criterion for selection was the area of residence). Moreover, 36.54% of the report that there is one other school in the same commune, 46.15% that there is one other school in the same commune, with the remaining having two (7.69%) or three (9.62%)

[31] The inclusion of these fixed effects also accounts for the fact that, even under annual random assignment of students across classes within schools, there is still persistence in the proportion of females across all classes in a given cohort. This persistence arises because there is little grade repetition and mobility across schools so the proportion of female peers for a given cohort remains similar as these students progress across grades.

Moreover, one central concern in long-run studies is selection out of the sample, or selection out of the assigned experimental group as years progress, as it is the case in, for instance, project STAR (Krueger, 1999). Young Lives made an impressive effort of participant tracking (attrition rate is only 2.5% over 15 years), which minimizes such concerns (I formally show this to be the case in Section 7).[32]

One important aspect to discuss is that, because the School Survey purposefully sampled 20 randomly-selected students per class, there is some measurement error in my peers variables. In particular, although I can observe the actual size of the class, the total number of males and females was not recorded, and I can only compute the proportion of females within the twenty students observed. If the resulting measurement error — which is present only for classes with more than 20 students[33] — is classical this will bias the estimates towards zero. Under random assignment to classes within schools and if students are missing at random, which should both hold given that the design of the school survey (i.e. random inclusion of 20 students per class), Sojourner (2013) shows that it is possible to account for the partial unobservability of peers by controlling for school fixed effects and their interactions with the proportion of total classmates that are observed (which I can compute because I know the size of each class — but not the total number of females in the class).

With these aspects in mind, identification in my empirical approach is achieved from variation in the percentage of female classmates within schools across classes.[34] The stability of the results will be verified by introducing additional covariates (both for the individual and the peers — i.e. in X and in $\overline{X}$) such as wealth or parental education. These inclusions, precisely stated as they become relevant, are not strictly necessary to ensure unbiasedness of my coefficient of interest, but they help in increasing precision. ϵ is the error term. I cluster the errors at the class level (the level of randomization) in order to account for possible correlation of the outcomes among classmates.[35]

[32]The second traditional identification issue in the peer effects literature is that of "reflection". Intuitively, this problem arises when it is hard to disentangle the effects of peers on an individual, and those from that student on her peers. In my particular exercise, in which the focus is on exogenous peer effects arising from a predetermined variable (gender), and given random classroom allocation, the concern that one's gender causes peers' gender is irrelevant.

[33]The median class size among the students in my estimating sample is 28.5 and the 75^{th} percentile is 33.

[34]On average, there are 22 students per school who belong to the YL survey and so I can observe their long-run views on gender roles.

[35]My main results hold under the more conservative choice of clustering errors at the school or the commune (third-degree administrative unit) level.

3.2 Validity of Exogenous Class Formation

The key identification assumption is that the exogenous assignment to peers within schools aimed by the principals was successful.[36] Thanks to the richness of my data I can perform a wide range of tests that consistently point at the validity of my empirical approach.

Balanced Characteristics. I take advantage of the longitudinal information from YL, which provides detailed individual and household information measured in 2007, hence *just predating* enrollment in the first year of primary education. In Figure 2 I show that, conditional on school fixed effects, age 5 characteristics are not correlated with the composition of peers in Grade 5. I find no evidence in favor of systematic differences prior to the formation of peer groups along a rich set of child's development dimensions (objective and subjective physical health, quantitative and verbal cognition), parental investments (measured by log consumption of food and non-food), parental locus of control (e.g. whether they believe that they can help their child if sick, which could determine both the incentives and the ability to affect class assignment) nor in terms of parental expectations about mid-life outcomes of their child.[37]

Appendix Figure B.7 abounds on this by showing that the parent's desired level of education for their child does not differ, nor their perceived probability that such level eventually is reached. There are no differences either in the reasons that parents allude to for having children, such as to provide additional economic or old care support help. Moreover, I compute the relative contribution of hours devoted to home production by the mother (i.e. the ratio of her time provision over the sum of her and her husband's contribution) as a relevant measure of traditionalism, and find no systematic relationship with class gender composition. Finally, female peer composition is not related either to a rich set of observable teacher characteristics such as their gender, tenure, the type of contract, their beliefs that teachers can succeed at helping even the worst students, and, notably, their pedagogical ability[38] (Table A.6).

Moreover, I take advantage of the gender attitudes elicited contemporaneously for the subset of young cohort students that *did not* participate in the school survey to obtain estimates of the impact of numerous observables[39] on gender attitudes. I then use these estimates to predict gender attitudes for those children in the young cohort that did participate in the school survey.

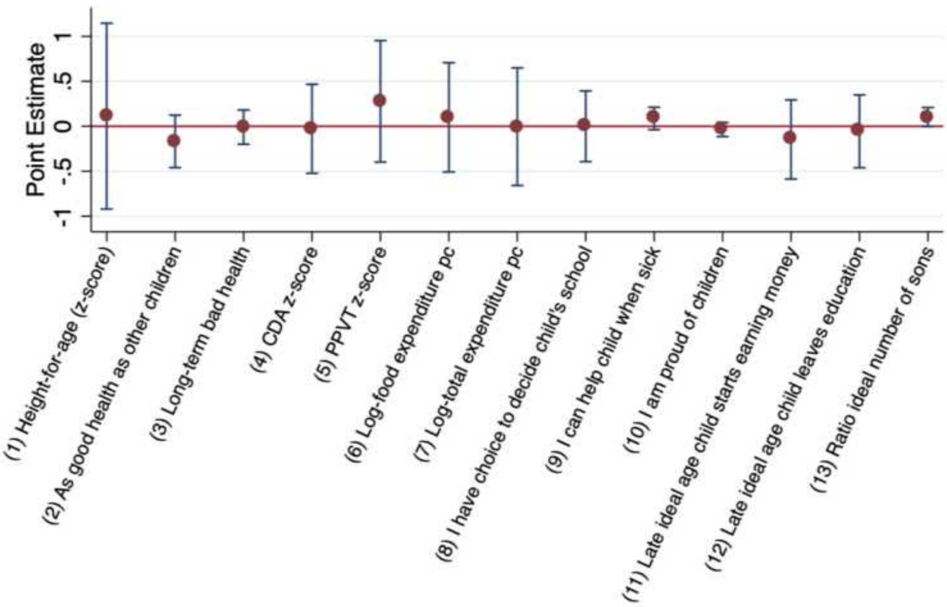
[36] For a similar setting see, for instance, Carlana (2019).

[37] I verify that this balance in observables does not mask heterogeneity across genders in two ways: (1) by not finding any statistically significant differences when estimating the regressions separately for males and females, and (2) when explicitly controlling for gender (available upon request).

[38] This is a unique piece of information from the SS by which teachers were given fictitious wrong answers from students to certain questions and were asked to identify the reasons why such mistakes most likely happened. This allows me to compute a measure of teachers' pedagogical abilities.

[39] These are: household size, gender, wealth index, gender of household head, household head's education, ethnic minority status, rural status, maternal education, religion, height-for-age and community fixed effects.

Figure 2: Balance of Pre-determined Covariates Measured at YL Round 2 (age 5)



Notes. All outcome variables are measured in Round 2 of the Young Lives survey, which was conducted in 2007. 2 and 3 are indicators for the parent saying that the child has at least as good health as same age children and that has some long-term health problems, respectively. 4 and 5 use standardized scores in the Cognitive Development Assessment – Quantitative test (CDA-Q) and the Peabody Picture Vocabulary Test. Both cognitive measures are the raw scores from two multiple choice tests with highest scores of 14 and 28 points, respectively. Outcome 8 is measured as the number of years of education that the parents state that they wish their child to achieve. 8-10 are indicators for whether parents agree or strongly agree with the following two statements: (a) "I can have a choice about which school to send my child to", and (b) "I can do little to help my child do well in school no matter how hard I try", and (c) "I feel proud of my children". 11 and 12 are indicators for whether the parents believe that their child should fulfill the statements after ages 25 and 23, respectively. The sample includes all YL children who were also surveyed in the School Survey and whose class-assignment was defined as "random". Regressions control for school fixed effects. 90% confidence intervals are reported.

Regressing these predicted attitudes on the proportion of female peers yields a point estimate of -0.017 with a p-value of 0.638. Hence, as much as a wide range of observables can predict gender norms, there was no difference between those children that were assigned to more or less females.

For the final balance check, I build upon the fact that there generally are multiple YL children per class (on average there are four). Because the YL children sampled in the first wave of the longitudinal study were selected as a random sample of the individuals in their cohort, I can verify the non-systematic allocation of students to classes by testing whether the distribution of YL children across classes within schools is consistent with an exogenous assignment.[40] I first replicate the analysis in Figure 2 where the dependent variable takes the value of one if the child is a participant in YL and zero otherwise. I obtain a point estimate of 0.067 with a p-value of 0.564. I additionally follow the literature in performing Fisher's exact test for whether students' observables and class assignment are statistically independent (e.g. Lavy and Sand,

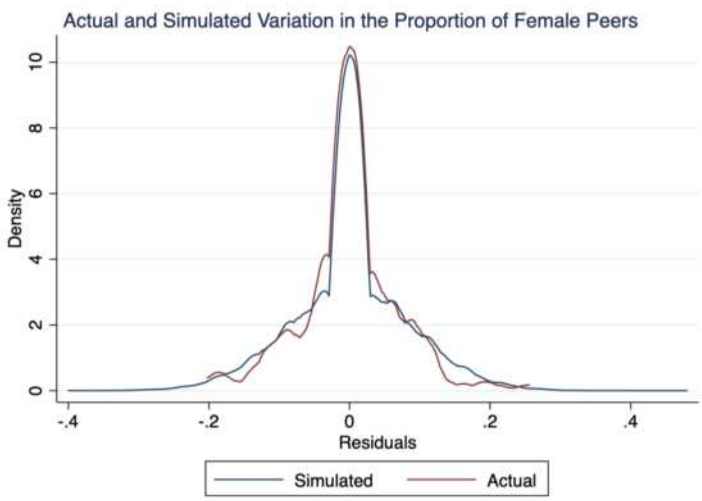
[40] It is relevant to note that there is very little variation in class size within schools. In my sample, the average gap in size between the largest and smallest class (normalized by the size of the largest class) is only 3%, the standard deviation being 0.05.

2019; Fruehwirth and Gagete-Miranda, 2019). There are only three schools where the p-value is below 0.05. As I discuss in the robustness section, excluding these schools does not influence the magnitude nor the significance of the main estimates.[41]

The empirical distribution of females is consistent with a randomized setting. I follow Bietenbeck (2020) in obtaining the residuals from a regression of the main independent variable — the proportion of females in a class — on school fixed effects. This is the variation that is exploited in my main analysis. I replicate this exercise 1,000 times, in each of them reshuffling students across classes within their original school, while respecting the schools' actual number of sections and the total number of female students. Figure 3 shows that the actual distribution of the residuals closely matches the one obtained in the Monte Carlo exercise, which strengthens the confidence that the variation in female peers exploited is indeed quasi-random.

Further evidence in favor of an exogenous assignment of students to classes is provided in Appendix C.3.2.

Figure 3: Distribution of Actual and Simulated Variation in Female Peers



[41]A double-check of the lack of systematic difference between YL children and peers, which reinforces the claim of lack of systematic class assignment, was already provided in Table A.2. My main results hold when restricting the estimating sample to schools with a lower proportion of YL children (e.g. 20%). This suggests that my results will not be driven by neighborhood effects, rather than school effects.

4 Main Results

4.1 Effects on the Composite Measure of Traditionalism

I present in Table 3 the results from estimating Equation 1, both without correcting for the partial observability of peers (Panel A) and correcting for it (Panel B).[42]

Baseline Specification (Uncorrected for Partial Observability). Column 1 simply controls for the student's gender and school fixed effects. There is a clear and strongly statistically significant negative correlation. In terms of economic magnitude, a ten percentage points increase in the proportion of female classmates decreases the degree of agreement with traditional norms in a tenth of a standard deviation (which is 0.32). While the quasi-experimental assignment of students to classes implies that these results are causal and unbiased, in column 2 I add a rich set of controls at the individual level — as listed in the Table's notes — in order to increase precision and to verify that the point estimate in column 1 is not sensitive to this inclusion, as should be the case under exogenous assignment. Both of these theoretical predictions hold.

Peers' Gender or Correlated Characteristics? Given that females may systematically differ from males along other dimensions than gender, it is important to account for a wide range of characteristics — both of the households and, arguably more importantly, of personality traits — at the peer level. In column 3 I start by adding the same controls as in column 2, this time for peers' leave-out attributes. The fact that the point estimate is virtually unchanged suggests that we are indeed capturing the effects of peers' gender, and not of other characteristics correlated with it. This is further reinforced in column 4, where I exploit a wide range of non-cognitive and attitudinal dimensions (e.g. effort exerted at school or the perception that by working hard the child can achieve college) in order to additionally control for usually-unobservable traits likely correlated with the outcome and that tend to differ across genders. In column 5 I explore whether the effects are heterogeneous for males and females. As expected from Figure B.12 this is not the case: while our measure of the proportion of female peers continues being significant and of about the same magnitude, its interaction with the female dummy is neither significant nor of a large economic size.

Main Specification (Corrected for Partial Observability). Reassuringly, the results from Panel B are qualitatively the same as in Panel A but they are an order of magnitude larger, as expected – the effect is now 16.25% of a standard deviation — and are precisely estimated. In Table A.7 I show that the results hold under alternative sample selections based on class formation (e.g. only fully "random" assignment). In Figure B.13 I abound on the stability of the

[42]For a visual analysis showing raw correlations between the proportion of females and views on gender roles in round 5 after residualizing both variables from school fixed effects see Figure B.12.

estimates by showing the 95% confidence interval from running 1023 ($2^{10}-1$) separate regressions featuring every possible combination out of the ten non-maintained controls[43] included in column 7. Alternatively, variable selection following Belloni et al. (2014)'s double machine learning selection algorithm (which implements two lassos: one for the main independent variable and one for the outcome variable) for which I also contemplate employing the dimensions analyzed in the balance tests (Figures 2 and B.7) and teacher characteristics (Table A.6) only suggests to control for the peers' wealth index (which was done in Table 3's columns 3 and 8).

4.2 Which Dimensions of Gender Norms Are Particularly Affected?

While using our summary measure of agreement with traditional gender norms has the benefit of being simple and parsimonious, it prevents us from more accurately understanding which of its dimensions were most affected. This is relevant for better pinpointing the mechanisms in place — and hence leading to potentially different policy implications. I follow Anderson (2008) in constructing variance-weighted indices for the four attitudinal categories after dichotomizing each of their items as in Table 2. Table 4's Panel A shows negative estimates over the board. Females are no longer expected to have traditional life goals, are not regarded as being less able, and are expected to interact with males in a more equal manner.

Heterogeneity by Gender. Although the interaction term in the baseline regression showed no average difference in treatment effects across genders, it is relevant to explore whether there is heterogeneity in the dimensions affected. In Table 4's panels B and C I repeat the same analysis separately for females and males, respectively. As expected, the impact on females' attitudes are large, and extend across all dimensions. Males experience similarly-sized impacts on "life purpose" and "cross-gender interactions". This is a key finding as traditional Confucian values emphasize the role of women in taking care of the house and the children, as well as the husband's family (Jiang, 2009).[44] Their views on females' overall ability is negative but not statistically significant. Unreported ordered probit regressions on the four-category outcome variable shows that this masks differences between their perceptions on women's intelligence (which is significantly impacted) and those on leadership (which are less affected). I complement this analysis by considering the possibility that being exposed to female teachers might lead to a similar shift in attitudes as exposure to female peers in Online Appendix C.4 (Dee,

[43] I do not include the control on numbers of books because it enters as a categorical variable in my main specification, which would largely increase the tuples generated. The maintained controls in my main specification (i.e. gender and school fixed effects) are always featured in every regression. Including further controls in the tuples, such as baseline cognitive ability or school interest of both the individuals and their peers yield a very similar pattern to the one in the figure.

[44] More generally, under the set of moral principles known as "Three Obediences and Four Virtues", women are expected to subordinate to males at the different stages of her life: first to the parent as a maiden daughter, then to the husband as a chaste wife, and finally to the son in widowhood (Gao et al., 2012).

Table 3: Effects of the Proportion of Female Peers on Long-term Views on Gender Norms

Panel A: Without Correction for Partial Observability of Peers					
	(1)	(2)	(3)	(4)	(5)
	Mean Agreement with Traditional Gender Norms at R5 (Age 15)				
Proportion Female Peers	-0.311**	-0.277**	-0.267**	-0.281*	-0.292*
	(0.139)	(0.128)	(0.128)	(0.148)	(0.170)
Female	-0.175***	-0.177***	-0.174***	-0.146***	-0.191*
	(0.021)	(0.021)	(0.021)	(0.021)	(0.107)
Female*Prop. Female Peers					0.030
					(0.215)
Mean (sd) Dep. Var.	2.18 (0.32)	2.18 (0.32)	2.18 (0.32)	2.18 (0.32)	2.18 (0.32)
Additional Controls	None	Indiv.	Indiv.+Peers	I.+P+Non-cognitive	Indiv.
Observations	880	880	880	792	880
R-squared	0.233	0.248	0.258	0.303	0.248
Panel B: With Correction for Partial Observability of Peers					
	(6)	(7)	(8)	(9)	(10)
	Mean Agreement with Traditional Gender Norms at R5 (Age 15)				
Proportion Female Peers	-0.526***	-0.486***	-0.424***	-0.489***	-0.505***
	(0.109)	(0.103)	(0.148)	(0.157)	(0.151)
Female	-0.185***	-0.186***	-0.181***	-0.157***	-0.204*
	(0.021)	(0.021)	(0.022)	(0.022)	(0.114)
Female*Prop. Female Peers					0.038
					(0.228)
Mean (sd) Dep. Var.	2.18 (0.32)	2.18 (0.32)	2.18 (0.32)	2.18 (0.32)	2.18 (0.32)
Additional Controls	None	Indiv.	Indiv.+Peers	I.+P+Non-cognitive	Indiv.
Observations	880	880	880	792	880
R-squared	0.274	0.290	0.299	0.349	0.290

Notes. All regressions control for gender and school fixed effects. Column 2 adds the following individual controls: age, an indicator for paternal literacy, an indicator for being an ethnic minority, the educational resources and household wealth indices, and the categorical measure of number of books in the household. Column 3 additionally adds controls for the average peer characteristics for the same dimensions included as individual controls. Column 4 adds non-cognitive characteristics both at the individual and peer levels: worry about exams, low school effort, high interest in school, difficult to complete homework, feel pressure at school, if I work hard I can go to college, and baseline mathematics score (all of them are defined in Section D). Column 5 estimates the same specification as in column 2, but adding the interaction between a female indicator and the proportion of female classmates. All controls are measured in the baseline round of the 2011 School Survey. The mean of the dependent variable is 2.18 and its standard deviation is 0.32. Panel B proceeds similarly but accounting for the partial observability of peers by including the addition of the interaction between school fixed effects and the proportion of peers observed in the class. The sample contains all YL students with non-missing values present in the School Survey whose class-assignment satisfied the sample selection criteria outlined in Section 2.1 and who are also observed in round 5 of the YL longitudinal dataset. Standard errors clustered at the class level (151 clusters) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

2007; Muralidharan and Sheth, 2016).

Table 4: Effects of Proportion of Females on Broad Categories of Gender Attitudes (R5, Age 15)

Panel A: Full Sample	(1) Life Purpose	(2) Abilities	(3) Cross-Gender Interactions	(4) Intra-household Decisions
Proportion Female Peers	-0.976***	-1.026***	-1.315***	-0.456
	(0.289)	(0.352)	(0.415)	(0.412)
Observations	865	865	865	865
R-squared	0.254	0.228	0.206	0.161
Panel B: Females	(9) Life Purpose	(10) Abilities	(11) Cross-Gender Interactions	(12) Intra-household Decisions
Proportion Female Peers	-1.187**	-1.660**	-1.380**	-0.998
	(0.524)	(0.727)	(0.550)	(0.807)
Observations	418	418	418	418
R-squared	0.321	0.306	0.239	0.294
Panel C: Males	(5) Life Purpose	(6) Abilities	(7) Cross-Gender Interactions	(8) Intra-household Decisions
Proportion Female Peers	-0.919***	-0.478	-1.317**	0.259
	(0.331)	(0.510)	(0.633)	(0.866)
Observations	430	430	430	430
R-squared	0.292	0.306	0.303	0.192

Notes. All regressions replicate Table 3's column 7. The outcome variables are variance-weighted indices. I first construct binary variables for each item taking the value 1 if the answer is "agree"/"strongly agree" with traditional views. After converting each individual binary variable into z-scores, I obtain the weight for each dimension within an attitudinal category from the inverse covariance matrix of all relevant items. The final index is again normalized to have mean 0 and standard deviation of 1. I restrict the sample to individuals with available information on all attitudinal items. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.3 Nonlinearities

So far I have estimated an average marginal effect common to all increases in the proportion of females. It is relevant from a policy perspective, however, to explore whether these marginal effects may be nonlinear. This is because, among other reasons, the number of males and females in the society is fixed, but the presence (or absence) of nonlinearities will have different implications for optimal class mixing.

Table 5 builds upon the baseline analyses and starts replacing the level in the proportion of female peers by an indicator for whether the child is exposed to more than 50% of female peers. I take this as a benchmark, as it reflects gender parity. In column 1 I find that children exposed to a majority of females reduce their traditionalism significantly more. Second, I construct three indicators for the child being exposed to: (i) up to 45% of female peers; (ii) between 45 and 55%; and (iii) over 55%. This choice of cutoffs was suggested by again aiming at having as a reference category one that reflects close to gender parity — category (ii) — as well as from visual inspection of the estimation of a partially-linear semiparametric model (Robinson, 1988). Again, I find that increased intensity in the proportion of females is translated into larger shifts towards less traditional views (column 2). In particular, the marginal effect for group (iii) is 0.085 points — around a fifth of the baseline effect, -0.486 — larger than for category (iii) than for any of the other two and precisely estimated. The p-value of the F-test of equality of the estimates for categories (i) and (iii) is 0.019.

The above results hold when in column 3 we take into account that the category for below 45% could mask considerable heterogeneity by restricting the sample to schools with at least 35% of female students — and hence expected to be more comparable to the rest of the schools than those with lower proportions. In column 4 I push this comparability idea even further by additionally limiting the sample to cases where the proportion of female students was below 65%. Moreover, the same qualitative reading arises both for males and females (columns 5 and 6). Given that all this discussion is based on relative comparisons, in column 7 I restrict my sample to those individuals whose peer composition lies within categories (i) and (ii) in order to explore whether level effects are also present in this part of the support of the proportion of females.[45] This is the case, the point estimate being smaller in absolute value than the baseline one (-0.367 vs. -0.486), as expected given the results in the previous columns.[46]

Class or cohort effects? The nonlinearities above suggest that there are potential Pareto improvements from manipulating the gender composition of classes. For this to actually be a plausible policy approach we need to make sure that the results found are indeed *class effects* (i.e. arising from the class composition) and not *cohort effects* (i.e. arising from the Grade gender composition at the school). If it were not the case, one would have to aim at affecting the full gender distribution of schools, as opposed to a simple reorganization of gender composition across classes. To explore this issue I create three categories of proportion of female students

[45]Note that, by doing so, I am allowing the impact of the controls to be different from those in the previous columns.

[46]I complement this evidence by allowing for different slopes at different points in the distribution of female peers by fitting linear splines in Figure B.15, which plots the predicted values from such model (after netting out school fixed effects). As expected, the slope up to the node at 0.4 is larger in absolute value than that for 0.4-0.55, so that we do not find statistical differences between these two groups in column 1. The slope then becomes steeper for the part of the support with high female presence (above 0.55).

in a class with a 3 p.p. window around gender parity: (i) below 47%; (ii) between 47 and 53%, and (iii) above 53%, the second category aiming at capturing parity in gender composition. In Table A.10 column 1 I restrict the sample to schools with classes belonging to at least two of these categories. Column 2 restricts the sample to schools with at least one class belonging to category (i) and one to category (iii), i.e. the more extreme categories. Given that the qualitative results are not affected, this provides evidence that the impact of gender norm is indeed class-driven and not cohort-driven. In Section 5 I provide a more detailed discussion on how, given my main results, one could think of feasible policy interventions that may successfully impact outcomes.

Table 5: Nonlinear Effects of the Proportion of Female Peers on Views on Gender Norms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean Agreement with Traditional Gender Norms (YL R5, Age 15)						
Proportion Female Peers ($> 50\%$)	-0.093***						
	(0.025)						
Proportion Female Peers ($0 - 45\%$)		0.018	0.001	0.002	-0.010	0.011	
		(0.028)	(0.031)	(0.031)	(0.048)	(0.045)	
Proportion Female Peers ($> 55\%$)		-0.085**	-0.081**	-0.076**	-0.155***	-0.118*	
		(0.034)	(0.034)	(0.035)	(0.048)	(0.070)	
Proportion Female Peers (Level)							-0.367**
							(0.177)
Sample	Full	Full	35% $\leq$ Prop. Fem.	35% $\leq$ Prop. Fem.$\leq$65%	Males	Females	Prop. Fem.$\leq$55%
Observations	886	886	796	744	446	440	707
R-squared	0.294	0.291	0.309	0.317	0.358	0.301	0.305

Notes. Replication of the specification in Table 3's column 7. Column 1 replaces the proportion of female peers by an indicator for the proportion of female peers being above 50%. Columns 2-5 use instead three indicators for the female proportion being: (i) below 45%; (ii) between 45 and 55% (the omitted category); and (iii) above 55%. Column 3 restricts the sample to classes where the proportion of female peers was at least 35%, while column 4 imposes the additional condition that the proportion of female peers is below 65%. Columns 5 and 6 restrict the sample to males and females, respectively. Column 7 fully replicates Table 3's column 7 for the subsample of individuals exposed to up to 55% of female peers. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.4 Are the Results in Line with Actual Behavior?

While the above analyses already provide novel and meaningful results, one important aspect to consider is whether stated preferences map onto actual outcomes that children can decide upon, at least partly. I explore two main ones, each expected to be particularly relevant for one gender : (i) female's academic outcomes (long-term enrollment) and (ii) male's contribution to home production.[47]

[47] Other interesting behaviors that have a clear mapping with the dimensions elicited in the gender norms questions are generally unobservable at the age of 15. For example, these adolescents are still too young to have formed a household so that we have no information on their chosen household arrangements or actions taken with respect to their children.

4.4.1 Academic Outcomes: Females' Long-term School Enrollment

I study the impacts on early (secondary education) dropout as well as on progression into tertiary education. Table 6 shows that there is none by age 15, a time when, while education is no longer compulsory (it finishes at at Grade 9, i.e. age 14) the dropout rate is still relatively low (about 20%). I am able to take advantage, however, of an ad hoc round of YL conducted in 2020 with the aim of collecting information on how the Covid-19 pandemics disrupted the lives of the participants. While this round mostly elicits information on Covid-related aspects, it does have information on pre-pandemics and contemporaneous enrollment status — individuals are asked both whether they were enrolled prior to the outbreak and if they are currently attending school or plan to continue attending after the pandemics is over — at the time that the children are already 19 years of age. Column 3 shows that females who were more exposed to other female peers are significantly more likely to remain at school at age 19 (52% of the children are still enrolled, generally at university[48]), while no effects are found for males (column 4).[49] What it is more, in column 5 I explore whether the nature of the major undertaken by girls at university is more traditionally male-dominated. This is indeed the case.[50] These findings are relevant given the inequalities in access to tertiary education across genders — even in the presence of the same observable characteristics (Mergoupis et al., 2018) — and that female minority status at traditionally male-dominated majors has been shown to be a cause of higher female dropout (Shan, 2021).[51]

4.4.2 Time Use: Males' Home Production

I employ detailed information on children's time use collected in rounds 4 and 5 of the YL survey. Children were asked to indicate the number of hours devoted to domestic tasks and caring for others on a normal week day.

In Table 7 I focus on the subsample of male YL children, whose reduced traditionalism has

[48]Official statistics from the Ministry of Education and Training show that there is close to parity in access to university across genders.

[49]Because YL asks full retrospective questions about school enrollment status by year, it is actually possible to explore the exact timing where the proportion of female peers starts mattering for the enrollment decision of female children. In unreported results from estimating Cox hazard models I find that the proportion of female peers did not significantly influence the decision to finish high school (Grade 12) but it did for the transition towards tertiary education.

[50]I consider male-dominated majors to be: mathematics and statistics; computer and communications sciences; construction and architecture engineering; other technologies; agriculture, forestry and aquaculture; transports; food processing; and armed force and police.

[51]I superficially investigate *who* are the girls that are more likely to remain at school when exposed to a higher proportion of females. Breaking the female sample between those who got a positive score in the cognitive z-score from round 2 provides suggestive evidence that, while both high an low ability females benefit from increased exposure, the effect may be somewhat larger among those one with a lower initial cognitive score (results available upon request).

Table 6: Is Long-run School Enrollment Affected by Peers' Female Composition?

	(1) Currently Enrolled R4 (Age 12)	(2) Currently Enrolled R5 (Age 15)	(3) Currently Enrolled Round Covid (Age 19)	(4) Currently Enrolled Round Covid (Age 19)	(5) Female-dominated Major (Age 19)
Proportion Female Peers	0.002	-0.157	0.615**	-0.217	-0.301**
	(0.013)	(0.168)	(0.245)	(0.382)	(0.126)
Sample	All	All	Females	Males	Females at University
Observations	883	883	408	378	192
R-squared	0.111	0.230	0.366	0.388	0.384

Notes. Replication of the specification in Table 3's column 2 for enrollment rates of YL children in rounds 4 (column 1) and 5 (column 2). Columns 3 and 4 explore the enrollment status in 2020 separately for females and females, respectively. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

the potential to make them more willing to contribute more time to chores[52]. Column 1 shows that being exposed to a higher fraction of female peers in Grade 5 causes a larger amount of hours devoted to home production two years later, although it is imprecisely estimated. Similar readings are obtained when looking at either the extensive or the intensive margin in columns 2 and 3, respectively. Returning to the same outcomes in round 5 (when children tend to provide more home production and there is more variation in the outcome) confirms the higher contributions to home production along all margins.[53] Table A.13 shows that the effects are significantly larger among individuals exposed to more than 50% of female peers, particularly for the outcomes in round 5. Allowing for nonlinearities also helps in increasing the precision of the estimates.

4.4.3 Cognitive Outcomes

In contrast with gender attitudes, which are novel outcomes in the peer effects literature, cognitive outcomes have been widely analyzed. I conclude this section by also addressing them because: (i) they are crucial in providing a full picture of the impact of exposure to females, which is necessary for policy, and (ii) they could be a potential mechanism behind females' increased enrollment.

Short-run Cognitive and School-related non-Cognitive Skills[54] Table 8 explores the effects

[52] A similar exercise for females does not uncover any systematic differences between children more or less exposed to female peers (results available upon request). One should note that at this age, even if girls have less traditional views there is little room for them to ask to contribute fewer hours than expected by the parents.

[53] A regression of time allocation in round 3, i.e. prior to the School Survey using the same specification as in the Table 7 does not yield significant estimates of female peer composition, which provides further evidence that class assignment in my sample was exogenous and that the divergence in outcomes occurs after class formation. In particular the point estimate is -0.06 (p-value of 0.80) when the outcome is an indicator for providing a positive number of home or care hours and it is 0.09 (p-value 0.46) when the outcome is an indicator for contributing over 2 hours per day.

[54] In Online Appendix C.6 I also explore the impact on other non-cognitive skills that are not directly related to academics, namely attitudes towards risk. I find that males who were exposed to a higher proportion of female peers are willing to pay less

Table 7: Effects of the Proportion of Female Peers on the Contribution to Home Production

	(1) R4 (Age 12) Total Home Hours	(2) R4 Any Home Hours	(3) R4 Over 3 Home Hours	(4) R5 (Age 15) Total Home Hours	(5) R5 Any Home Hours	(6) R5 Over 3 Home Hours
Proportion Female Peers	0.606	0.158	0.569**	2.126**	0.427*	0.489**
	(0.799)	(0.276)	(0.272)	(0.992)	(0.216)	(0.213)
Mean (sd) Dep. Var.	1.36 (1.24)	0.78 (0.42)	0.15 (0.36)	1.64 (1.42)	0.86 (0.35)	0.18 (0.38)
Observations	433	433	433	433	433	433
R-squared	0.301	0.334	0.257	0.283	0.352	0.260

Notes. All regressions replicate the specification in Table 3's column 7 for the subsample of male children. Home hours are computed as the number of hours devoted on a normal weekday to either home chores (etching water, firewood, cleaning, cooking, washing and shopping) or caring for relatives (younger children, ill household members) in rounds 4 and 5 of the longitudinal YL survey (collected in 2013 and 2016, respectively). Columns 1 and 4 use the levels of home hours as outcome while the remaining ones use the relevant indicator according to the column's header. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

on other school-related, cognitive and non-cognitive, outcomes. The latter are interesting both in themselves and also to better understand potential mechanisms.[55] Given that in this exercise I am purely interested in those changes taking place in the short-run (i.e. exclusively during the School Survey academic year) I can gain statistical power by expanding my sample to all students in the class, irrespective of whether they are YL children (results are similar when using exclusively YL children). Column 1 in Panel A shows that male children exposed to more females increase significantly more their score in Vietnamese over one academic year (I control for baseline cognition at the start of the year). Positive effects are also found for mathematics and among girls, but they are imprecisely estimated. Columns 3 and 4 suggest that part of these academic spillovers are channeled through increased effort and self-confidence in own academic ability. Such an endogenous spillover arising from female classmates, who are themselves significantly more inclined to put effort at school[56], has previously been found in the literature (e.g. Gong et al., 2019).[57].

for a risky lottery. This is accompanied by them implementing more preventive measures in the context of the Covid-19 pandemics. No effects are found for females.
Additionally, in Appendix C.7 I document the effects on fertility preferences. I find again that it is males' views that are predominantly shifted — and, once more, in the direction of becoming more in line to those of women.

[55] Having said this, it is an almost impossible endeavour to attempt to explore all possible non-cognitive channels. Indeed, an important candidate, namely improved class atmosphere, cannot be studied as measures for this dimension are unavailable to me. Still, valuable lessons can be learnt.

[56] In my sample, my "low effort" variable has an average value of -0.15 for females and 0.11 for males — the difference being significant at the 1% level — showing that females are systematically more prone to put effort at school.

[57] Replacing the proportion of female peers in Panel A's column 4 for an indicator of the child having above 50% of female peers shows that the effects on Vietnamese are larger for that group (by 25.696 points, significant at the 5% confidence level.)

Table 8: Effects of the Proportion of Female Peers on Short-term, Value-added Cognitive and Non-cognitive Academic Outcomes (Second Round of the School Survey, Age 10-11)

Panel A: Males					
	(1)	(2)	(3)	(4)	(5)
	Vietnamese SS Retake	Mathematics SS Retake	Low School Effort	Low Confidence Own Academic Ability	Working Hard will Not Help me to Attend College
Proportion Female Peers	130.066**	98.774	-1.062**	-1.441**	0.080
	(60.618)	(96.063)	(0.439)	(0.629)	(0.627)
Mean (sd) Dep. Var.	477.03 (113.92)	490.97 (118.03)	0.11 (1.29)	0.02 (1.22)	1.62 (0.77)
Observations	1,463	1,463	1,426	1,421	1,449
R-squared	0.377	0.356	0.179	0.172	
Panel B: Females					
	(1)	(2)	(3)	(4)	(5)
	Vietnamese SS Retake	Mathematics SS Retake	Low School Effort	Low Confidence Own Academic Ability	Working Hard will Not Help me to Attend College
Proportion Female Peers	100.720	99.016	-0.450	-0.525	0.199
	(62.148)	(95.116)	(0.383)	(0.766)	(0.637)
Mean (sd) Dep. Var.	502.87 (106.37)	492.69 (110.13)	-0.15 (1.21)	0.004 (1.19)	1.64 (0.78)
Observations	1,335	1,335	1,306	1,297	1,328
R-squared	0.418	0.385	0.197	0.192	

Notes. All regressions replicate the specification in Table 3's column 7 with outcome variables collected at the second round of the 2011 School Survey. The sample uses also non-YL students. Cognitive scores are standardized to have a mean of 500 and a standard deviation of 100 (the regressions in columns 1 and 2 are augmented by controlling also for baseline cognitive scores). Column 5 in both panels estimates an ordered probit model. A description of the other outcome variables used is provided in Section D. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Long-run Cognitive Outcomes. Given the findings on short-term cognition, I now exploit another strength of the YL dataset by which participants' mathematics and verbal skills were tested in every round. This is an attractive feature since it provides actual cognitive ability, rather than having to resort to proxying for the effects on human capital through grade completion or school enrollment, which are noisy measures, particularly in developing countries (Singh, 2020). Crucially, these cognitive scores are available for every YL student, *irrespective* of her school enrollment status.

Table 9 shows that there is little evidence in favor of increased cognitive performance in the mid-run, and the point estimates even turn negative among males — albeit they are not significant. This is consistent with ample existing evidence in the literature documenting that initially significant cognitive spillovers soon turn non-significant (e.g. Carrell et al., 2018; Bietenbeck, 2020).

5 Mechanisms

5.1 Mechanisms for Decreased Female Dropout

Given YL's aim to track children's development, very detailed attitudinal information on academic and professional aspirations, as well as the expected returns to education were elicited over time. This grants a unique opportunity to explore the potential channels behind females' decreased dropout in a comprehensive way.

Academic Aspirations. Given the finding in Table 4 that females' agreement with traditional life goals was largely decreased, an intuitive first mechanism to explore is whether females' own academic aspirations increased upon stronger exposure to females. In Table A.14 in the Online Appendix I construct an indicator variable taking the value of 1 if the child claims, in round 5 of the YL (three years before actually enrolling at university), that his/her desired level of education is at least a college degree, and 0 otherwise. While I do not find any significant impact on males, there is a strong positive effect for females.[58]

Perceived Returns to Education. Another potential — and complementary — channel is that the increased professional ambition is closely related with a shift in the expected returns

At this point, one may wonder whether the effects on academic scores are indeed driven by gender or they could simply be capturing the fact that, in Vietnamese primary schools, females perform better than males. This is a reasonable possibility because the baseline gender gap in raw tests scores is significantly larger in Vietnamese (19.38 for females vs. 16.27 for males) than in mathematics (16.27 vs. 16.19). However, in Recall, however, that my specifications control for a wide range of peer characteristics, including peers' baseline test scores. Therefore, differences in peers' ability do not explain the effects of increased exposure to female classmates.

[58]Finding effects of female peers on boys' academic outcomes but not on girls' has been documented by, for instance, Lavy and Schlosser (2011) and Brenøe and Zölitz (2020).

to education. If children believe — whether accurately or not — that returns to education are higher, they will be more likely to remain at school (Jensen, 2010; Attanasio and Kaufmann, 2014). Indeed, the connection between expected returns and enrollment among the Young Lives' Peruvian young cohort has been recently shown by Favara et al. (2021).

I explore this possibility by making use of Round 5's information from the following questions: (i) assuming that you complete university, what do you think is the minimum amount you can earn per month at age 25?; (ii) assuming that you complete university, what do you think is the maximum amount you can earn per month at age 25?; (iii) assuming that you complete university, what do you think is the probability that your earnings at age 25 will be at least [midpoint between max and min level]? Under the standard assumption of a perceived triangular distribution of wages (Guiso et al., 2002; Coibion et al., 2021) I can use the above three pieces of information to compute the expectation and the variance of expected earnings for each individual.

In Table A.15 I show that females who are exposed to a higher proportion of female peers display both a higher (log) mean and a higher (log) variance for expected earnings at age 25 (columns 1 and 3), while there is no change among males (columns 2 and 4, as one would have expected). In column 5 I find that the mean expected return is largely predictive of enrollment at age 19. I find a negative coefficient for the variance, as consistent with theory, but small in economic size and not significant. This can explain why, while both the expected mean and variance increased in column 1, the net effect on female enrollment is positive, since variance does not seem to play a large role in the decision. These results, which investigate a novel outcome within the peer effects literature, provide a mechanism rationalizing both why females who are more exposed to female peers are more prone to still be enrolled by age 19 and why we do not observe any change for males.[59]

Professional Expectations. There remains, however, the question of why girls' expectations on academic returns increase, and why males' views do not change much. As a possible explanation, and based on the above results pointing at girls changing their desired careers but males not doing so, I make use of an additional question in round 5 of the YL survey asking: "when you are about 25 years old, what job would you like to be doing?", where children were able to choose among a very wide range of professions. I categorize these professions into three blocks: (i) traditionally female (e.g. secretary); (ii) gender neutral (e.g. artist), and (iii) traditionally male (e.g. soldier). Additionally, I construct an indicator for a profession not being

[59]It is relevant to note that there is a significant negative relationship between females' views on gender roles and expected earnings at age 25 when children are asked to assume that they have obtained college education. This holds after saturating the regression with controls such as past cognitive performance and academic aspirations. Instrumenting views on gender roles with the proportion of female peers in the class abounds on the negative relationship (results available upon request).

traditionally female-dominated — i.e. takes value 1 if the desired profession belonged to (ii) or (iii).[60] Table A.16 shows some evidence that while girls who were majority in the class are more likely to wish to have less traditional female jobs (using the binary indicator in column 3[61]; there is no effect with the three-category outcome variable in 1), males do not change their expectations. Given that female-dominated professions are traditionally lower-paying, these results can relate the increase in expected returns to education to the change in desired life goals, which also shifts the *sector* of jobs to which female students aspire.

Increased Academic Self-confidence. In Table A.17's column 1 I show that, by round 5, males who were more exposed to female peers do not show a significantly higher confidence in their own abilities, which was identified as a plausibly source of the short-term cognitive gains. Crucially, the short-term gains in confidence that were not statistically significant for females now are strongly so (column 2). This not only highlights the importance of tracking outcomes over time to better understand their dynamics and true effects, but also provides another potential channel for why females increased their school enrollment in the long-run.[62]

Mediation Analysis

The above discussion has identified multiple potential channels at work: (i) fall in acceptance of traditional career goals for females; (ii) increased aspirations to reach university; (iii) shifts in expected returns to university: higher mean and lower variance of expected earnings; (iv) increased confidence. All of these mechanisms, whose relative importance can be assessed through a mediation analysis, are consistent both with female peers' being more encouraging towards less traditional careers and male peers exerting less stereotypical views.[63]

[60] My proposed categorization seems to match YL children's responses well. Among males, 43.22% wish to have a category (iii) job, 47.57% a category (ii) and only 9.21% a category (i). Conversely, 33.51% of girls choose a category (i) job, 63.03% a category (ii) and only 3.46% a category (iii) one.

[61] Unreportedly, I find that this effect is mostly driven by category (ii) jobs — females become more likely to belong to this category.

[62] Why do females raise their confidence? One may only speculate at this point, but one possible reason is that, by being more in contact with other girls they can make more accurate comparisons of themselves with respect to other girls. Another possibility is that they gain self-confidence when being less of a minority within a group.

[63] As a first step, it is relevant to note that only females' probability of achieving tertiary education was affected by the proportion of female peers. This was not ex ante obvious, as one could think that an increase in female attendance to university could have crowded-out males at the margin of academic ability. This is indeed what the negative sign on actual enrollment and on academic aspirations among males might capture — although there are no heterogeneous effects by distance to peers' ability nor by whether the school has on average higher performers than other schools (unreported). In any case, thee result is not surprising because none of the three mechanisms was impacted either among males. The most plausible reason for my finding is that, while my interest is on whether males change their attitudes towards females in a more inclusive way, this is unlikely to come at the expense of boys turning to less favoring outcomes from themselves, particularly in terms of outcomes that are largely independent from what females do, most notably academics (unlike within household power, where the male Pareto weight is the complement of the female's). One pertinent remark is that I find virtually no correlation between risk attitudes — which are modified by female exposure — and enrollment rates.

Table 9: Effects of the Proportion of Female Peers on Long-term Academic Outcomes

Panel A: Males				
	(1)	(2)	(3)	(4)
	Vietnamese	Math	Vietnamese	Math
	R4 (Age 12)	R4 (Age 12)	R5 (Age 15)	R5 (Age 15)
Proportion Female Peers	49.364	52.213	-27.517	-26.949
	(50.682)	(75.439)	(50.150)	(74.592)
Observations	428	429	429	429
R-squared	0.425	0.450	0.403	0.432
Panel B: Females				
	(1)	(2)	(3)	(4)
	Vietnamese	Math	Vietnamese	Math
	R4 (Age 12)	R4 (Age 12)	R5 (Age 15)	R5 (Age 15)
Proportion Female Peers	45.547	25.153	-9.459	73.867
	(66.733)	(65.276)	(61.652)	(71.527)
Observations	419	419	429	427
R-squared	0.361	0.460	0.367	0.435

Notes. All regressions replicate the specification in Table 3's column 7. The outcome variables are the standardized (mean of 500, standard deviation of 100) Vietnamese and mathematics scores from the fourth and fifth rounds of the YL data (collected in 2013 and 2016, respectively). Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Gelbach (2016) provides a decomposition of the overall treatment effect of interest into the relative importance of each covariate that is independent of the order in which they are added. His approach, inspired by the well-known formula defining the population omitted variable bias arising from excluding a relevant set of controls in an OLS regression, consists of two steps. In the first one I regress one-at-a-time the five main mechanisms found above (I consider the mean and the variance in (iii) separately) on the leave-out mean of the proportion of female students in the class, together with the baseline controls. I denote each of these mechanisms by "p". In particular, such regressions take the following form:

$$p^j_{ics2} = \beta_0 + \beta_j PF_{-ics1} + \beta_2 X_{ics1} + \lambda_s + \epsilon_{ics}, \tag{2}$$

where, as stated, $j = \{$life goals, aspirations for university, mean expected income from university, variance of expected income from university, confidence$\}$. This yields, for each p^j, an estimate (β_j) of how the proportion of female classmates shifts each of these mediators.

In the second step I run my main specification including all the variables in p as controls. This will yield estimates for the individual effect of each of them on enrollment in 2020 (α^j) above-and-beyond the role of female composition (α_M).

$$Enrolled_{ics2020} = \gamma_0 + \alpha_M PF_{-ics1} + \sum_J \alpha^j p^j_{ics2} + \gamma_1 X_{ics1} + \lambda_s + \epsilon_{ics}. \tag{3}$$

In Table 10 I report the estimates from Equation 3 (column 5) where for completeness I also show how — an arbitrary — ordering in the inclusion of the mechanisms affect the point estimate attached to the proportion of female peers. As consistent with the claim that the suggested mechanisms are indeed relevant, one can appreciate the fall in economic magnitude and the loss in statistical significance experimented by the proportion of female friends from column 1 (I call it θ) to 5 (α_M), while all the channels considered relate significantly (except confidence) — and in the expected direction — with enrollment in 2020.

One can then quantify the relative importance (RI) of each mechanism in explaining the size difference in the estimates of the proportion of female peers in column 1 (0.621) and 5 (0.16) by computing, separately for each variable j, the following ratio: $\text{RI}_j = \frac{\alpha^j \beta_j}{\theta}$, i.e. the fraction of the treatment effect of female peers that is explained by each mediator.

Building upon the estimates of Equation 2 previously reported in various tables throughout the text, Figure 4 shows that the five channels considered are able to explain almost 75% of the full effect on enrollment rates: 1-(0.16/0.621). This is mostly driven by increased aspirations for university (35% of the total effect) and by higher expected earnings (26%).[64] An additional 9%

[64] These estimates should be interpreted with caution if one were to think that the mechanisms are measured with error or that some channel that also correlates with the included ones has not been taken into account. The fact that we are able to explain three fourths of the total effect should reduce this concern.

Table 10: Mediation Analysis for Female's Increased Enrollment in 2020

	(1)	(2)	(3)	(4)	(5)
	Enrolled at School in 2020 (Age 19)				
Proportion Female Peers	0.621**	0.462*	0.247	0.209	0.160
	(0.249)	(0.264)	(0.267)	(0.267)	(0.271)
Expected Earnings from University (in 10^{-6})		23.237**	20.973**	18.714*	17.805*
		(9.996)	(10.085)	(9.823)	(9.837)
Variance Earnings from University (in 10^{-6})		-0.010**	-0.010**	-0.009*	-0.008*
		(0.005)	(0.005)	(0.004)	(0.004)
Aspires to University			0.355***	0.332***	0.330***
			(0.076)	(0.080)	(0.080)
Traditional Life Goals				-0.076**	-0.077**
				(0.035)	(0.036)
Confidence					0.078
					(0.096)
Observations	402	402	402	402	402
R-squared	0.370	0.382	0.437	0.446	0.447

Notes. All regressions replicate the specification in Table 3's column 7 using enrollment in 2020 as an outcome and introducing mediating dimensions in a stepwise manner. The sample is restricted to females. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

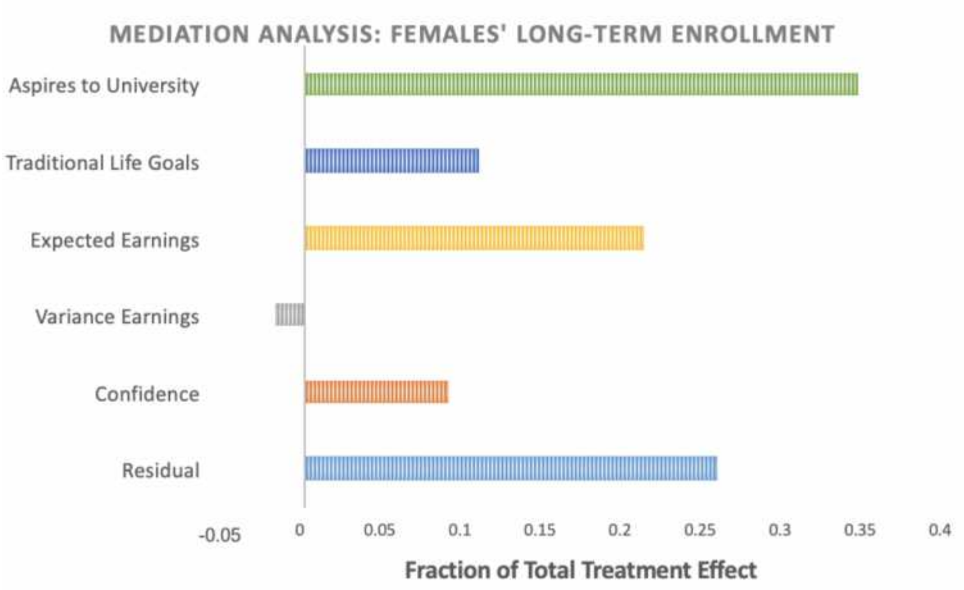
is explained by higher self confidence and 11% by having less traditional life goals, which can be closely connected to the change in aspirations. Finally, consistently with previous results, the increased in variance plays almost no role on enrollment (the effect is actually negative as a higher proportion of female peers led to a higher expected variance in earnings, which negatively relates to enrollment).[65]

5.2 Mechanism for Males' Decreased Traditionalism in Views on Norms and Ability: Cross-gender Contact and Friendship Formation

Table 4 clearly points towards males becoming more progressive in their views of the preferable careers for females, and in the way their personal interactions should take place. This suggests that an important mechanism behind this is likely to be increased actual interactions across

[65] A similar mediation analysis for the number of children shows that, while a large fraction of the variation remains unexplained, the strongest observed channel is through cross-gender relationship (10% of the total variation).

Figure 4: Mediation Analysis: Breakdown of Relative Importance of Channels in Decreased Female Dropout



genders. To validate this claim I exploit the school survey friend nominations for every surveyed classmate. I focus on the male subsample and ask the following question: are children who share the classroom with a higher proportion of female classmates disproportionally more likely to form friendship groups with a higher fraction of females?

In Table 11 I provide the results of estimating Equation 1 for different definitions of friendship. Columns 1-3 make use of the question "How would you describe your friendship with this classmate?", which allows for the following answers: (a) Not close friends; (b) A little/sometimes friends; (c) Close friends; (d) Very close friends, and construct three indicators that are increasing in the strength of the friendship.[66] Columns 4-6 proceed similarly with respect to the following question: "How much do you do things with this classmate outside of school?" where the options were: (a') None; (b') Not very much; (c') Quite a lot; (d') A lot. Therefore, while columns 1-3 consider self-perceived friendship, columns 4-6 acknowledge that: (i) there may still be interactions outside school with individuals that are not considered to be a friends, and (ii) that there is room for peers to have varying degrees of influence based on the total amount of interactions. Finally, columns 7-9 require both dimensions to hold simultaneously for the outcome variable to take value 1. For instance, the dependent variable in column 7 will be 1 if the answer to the first question is (b), (c) or (d) and, at the same time, the answer for the second one is (b'), (c') or (d').

[66]For instance, column 1's "at least a little close" is an indicator taking the value 1 if the child responded option (b), (c) or (d) and 0 otherwise. In a similar vein,"at least close" takes the value 1 if the answer is either (c) or (d).

The results are consistent across all these definitions. There is no significant relationship between the proportion of female classmates and the proportion of "close" friends or strong outside-school interactions (i.e. columns 2, 3, 5, 6, 8, and 9) whereas the effect is large and significant in the "somewhat" friends and "above-not very much" contact cases (columns 1, 4, and 7). In order to interpret the magnitude, a useful counterfactual to consider is that, if friendship formation was purely random, the estimated coefficient would be 1. Therefore, given the mean values of the dependent variables reported in the table, it becomes clear that a higher proportion of female peers significantly increases the fraction, and the total number, of nominated female friends. Indeed, in Section C.5[67] I argue that, under reasonable assumptions, these estimates suggest that increasing the fraction of females classmates raises the proportion of female friends *more than proportionally* when using the current friendship formation patterns as counterfactual. Importantly, the nature of such interactions is relatively limited: male children are more prone to interact with females but not to the point where there is an increase in the proportion of them that become very close friends. This is confirmed through a fixed-effects model in which I exploit within student variation in the gender of classmates. I find that males in classes with more than 50% of females are over 4 percentage points more likely to become friends with girls.

Overall, this unique opportunity of observing friendship nominations provides evidence that there is an actual behavioral change on the part of male children and also highlights that increased interaction of male children with female peers, even if not to the point of becoming close friends, potentially reshaped their views towards females' abilities and identity roles through fostered familiarity and friendship. A remaining question is whether non-friend female classmates have an independent impact beyond that of friends. This is likely the case given that only the weakest definition of friendship yielded significant estimates. I provide evidence supporting this claim by using the estimates from Table 11 to predict the proportion of female friends among the male population and conducting my baseline analysis of gender attitudes separately for those males whose actual proportion of female friends is below the predicted one (and hence are not expected to have changed their friendship patterns much) or above. The results hold for both subsamples (available upon request).

5.3 Additional Evidence on Mechanisms

In Online Appendix C.8 I provide additional evidence consistent with the above discussion. In particular, I: (i) show that being surrounded by objectively more able female peers' — as

[67] This section also provides a parallel analysis for females.

Table 11: Effects of the Proportion of Female Peers on the Female Composition of Male Children's Friendship Networks (Second Round of School Survey, Age 10-11)

Panel A: Degree of Friendship			
	(1)	(2)	(3)
Female Friends/Total Friends defined as...	At Least a Little Close	At Least Close	Very Close
Proportion Female Peers	0.657***	0.116	-0.020
	(0.208)	(0.158)	(0.187)
Mean (sd) Dep. Var.	0.35 (0.19)	0.21 (0.21)	0.12 (0.23)
Observations	441	436	382
R-squared	0.447	0.364	0.203

Panel B: Contact Outside School			
	(4)	(5)	(6)
Female Friends/Total Friends defined as...	At Least not Many	At Least Quite a Lot	A Lot
Proportion Female Peers	0.573***	0.165	0.089
	(0.153)	(0.188)	(0.164)
Mean (sd) Dep. Var.	0.33 (0.21)	0.20 (0.24)	0.15 (0.26)
Observations	432	425	355
R-squared	0.399	0.338	0.243

Panel C: Combined Measures			
	(7)	(8)	(9)
Female Friends/Total Friends defined as...	1+4	2+5	3+6
Proportion Female Peers	0.588***	0.077	0.030
	(0.178)	(0.152)	(0.170)
Mean (sd) Dep. Var.	0.30 (0.22)	0.16 (0.23)	0.10 (0.22)
Observations	432	419	312
R-squared	0.417	0.299	0.201

Notes. All regressions replicate the specification in Table 3's column 2 for the subsample of male children. The outcomes are measured in the second round of the 2011 School Survey. The dependent variables in columns 1-3 refer to friendship nomination. Columns 4-6 refer to the frequency of doing things together. Columns 7-9 require both definitions to be satisfied at the same time. The number of observations varies as the dependent variable is the proportion of female friends *among total friends*. Therefore, if a child does not nominate any friend, regardless of the gender, for a given category this situation will lead to a missing value. Standard errors clustered at the class level (132 clusters) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

measured by the beginning-of-the-year test of the School Survey — is an independent driver or males' shifts towards less traditional views on female abilities. This is in line with a revision of pre-conceived negative views on them on the males' part upon exposure, and (ii) investigate in more detail *who* experience larger treatment effects. I find that the main source of heterogeneity is pre-exposure traditionalism either at the household level (measured as the proportion of important decisions that the female spouse undertakes) or at a more aggregate level (measured as the overall traditionalism in the community of residence from the old cohort's reported views). This analysis shows that the larger effects are found among more traditional individuals, which points at an overall convergence in attitudes. A more nuanced exploration through quantile regressions suggests in Table A.25 suggests that the impact is largest around the median, and that the very right tail of traditionalism is not impacted.

6 Room for Policy: Discussion

Is there room for effective policy interventions aiming at changing gender norms? Absent my results from above, one may tempted to think that the answer is no: the status quo is universal gender mixing in all primary education classes, only 2% of male children do not report to have any female friends and, yet, their gender attitudes are strongly traditional. This is why the presence of nonlinearities in my main results, together with the fact that outcomes for classes with low proportions of female students are not significantly different from those where there is close to gender parity (but still male-dominated), is an encouraging finding.

For a potential policy to be operational, I require it to satisfy the following characteristics: (i) it should *only* rely on the *actual* school gender composition, i.e. it should not involve attracting new students nor forcefully displacing them into a school, (ii) it should not create significant gender-imbalances across classes (i.e. maintain *close to* gender parity), and (iii) it should be framed in terms of easily observable and fixed characteristics (like gender), not on particular characteristics of individuals that may be endogenous (including the degree of interpersonal interactions) or that may not be easily portable or developed across peer groups (Carrell et al., 2013).

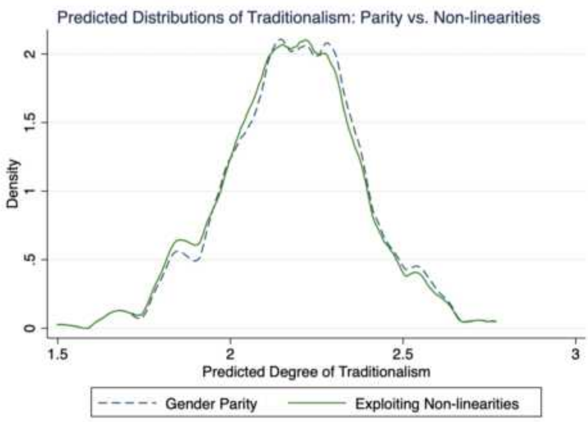
In what follows I compare the outcomes from two allocative processes: (1) full gender parity, and (2) a policy that relies on the 45 and 55% cutoffs non-linearities uncovered in Table 5 while still preserving close to gender parity at the class level. The rationale for the second thought experiment is that female-dominated classes are disproportionally beneficial for the important outcomes discussed throughout the text, while for the slightly male-dominated classes the treatment effects are more linear. One should note, however, that the proposed policy described in more detail below simply aims at being illustrative of the potential for interventions to be successful while, at the same time, preserving equality in outcomes. I therefore do not aim at making optimality claims, which would require a more flexible understanding of the existing non-linearities than the relative coarse modelling of Table 5.

To be specific, I proceed under the following algorithm (always respecting the actual number of classes, students per class, and gender distribution at the school level):

1. **Counterfactual allocation of students**
   - Counterfactual 1: Full parity
     - If a school only has one class, the counterfactual distribution is the same as the actual one.
     - If a school has multiple classes, assign the first girl to the first class, the second one to the second class, and so on until all classes have one girl. Repeat the operation for the second girl in each class.
   - Counterfactual 2: Exploiting non-linearities
     - If a school only has one class, the counterfactual distribution is the same as the actual one.
     - If a school has multiple classes, and it is possible to have one class with 55% females while ensuring that all other sections have at least 40% of females, do so. If there are still ”additional” girls, assign them to the next class so that it achieves the thresholds of either 45% or 55% females.
     - If a school has multiple classes and it is not possible to have one class with 55% females while ensuring that all other sections have at least 40% of females, equalizing the number of girls across all sections.
2. **Prediction of counterfactual gender norms**
   - Based on the estimates of my nonlinear model (three categories for the proportion of females) in Table 5, use the counterfactual proportion of female classmates generated under the two counterfactual scenarios above to obtain the corresponding two values of predicted of predicted gender norms.

By construction, the second counterfactual aims at employing the non-linearities to decrease the degree of traditionalism whenever possible, without imposing a trade-off that negatively impacts attitudes in the other cases. One can appreciate in Figure 5 that, indeed, the main gains under the non-linearities counterfactual are concentrated on the left-tail of the distribution, whereas the right one remains almost unaltered. In terms of size, the mean difference in predicted traditionalism under both policies is over 2% of the standard deviation of the predicted norms under the true class configuration.[68] Kolmogorov-Smirnov's test of equality of the two distributions rejects the null with at the 4% confidence level. These effects are particularly meaningful given that recent evidence suggests that convergence in gender equality in Vietnam is slow and, in some cases, even regressive despite large governmental gender-equality measures such as the National Program on Gender Equality (ISDS, 2016).

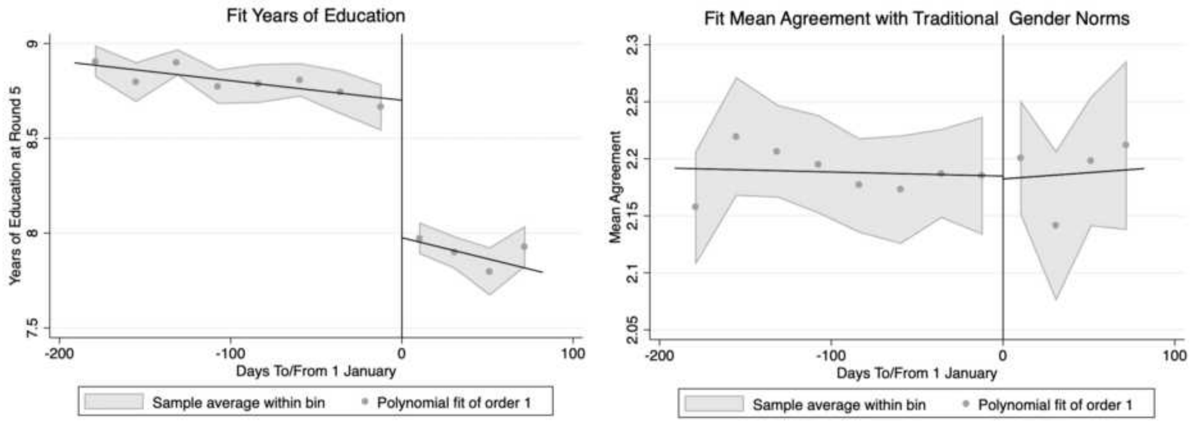
Figure 5: Distribution of Predicted Outcomes Under the Two Counterfactual Scenarios



Benchmarking the Effects of the Policy: Comparison with the Returns to One Year of Education. The estimated gains from my proposed policy, which are sizable in themselves, become particularly attractive when noting that, in my context of interest, one additional year of education does not significantly translate into less traditional gender roles, which is a novel result within the literature on the non-pecuniary returns to education (Oreopoulos and Salvanes, 2011). In particular, I exploit the fact that the educational regulation provides an exogenous source of variation in total years of education attained by round 5 between children born before and after 1 January, as those born after that date are systematically more likely to enrol in education one academic year later — in a fuzzy manner. This strategy has recently been used with YL data by Singh (2020), who offers a compelling argumentation in favor of the validity of this empirical approach. In Figure 6 one can appreciate that, while we do have the expected discontinuity in educational achievement (individuals to the right of the threshold are born after 1 January), this does not translate into different support towards traditional gender

[68] One way of thinking of the validity of the model proposed is by looking at the R^2 in the baseline nonlinear specification. It can be seen from Table 5 that it is high: around 30%. I also verify that the predicted and actual norms are on the 45 degree line and the cloud of dots is tight around it and evenly distributed.

Figure 6: No Returns to Education on Gender Norms



(a) First Stage: Education

(b) Reduced Form: Norms

roles.[69] Formal estimation delivers a point estimate of 0.054 (indicating that younger children, and hence with one less year of formal education, have more traditional views) although it is not statistically significant (p-value is 0.621).

In order to provide an even more complete picture of the relative size of my estimates, I also consider an interesting comparison: how does the impact of female classmates compare to the impact of female siblings? In order to deal with the potential endogeneity arising from unobserved time-invariant heterogeneity in son preference common in Confucianism-related cultures, I obtain 2SLS estimates of the impact of having at least one sister on gender norms where I instrument the main independent variable with an indicator of whether the first child born to the parents of the YL kid was a male, a common strategy in the literature (e.g. Li and Wu, 2011; van Lent, 2020). Under such an event, the likelihood that families have a female sibling is lower as their incentives for having a male fall. I find a precisely estimated zero impact of female siblings. This speaks to the difficulty of changing gender norms in the Vietnamese society, including within the household, and highlights the relevance of the sizable effects estimated to arise from the exposure to peers.

7 Robustness Checks and Cross-country Extensions

In this section I provide evidence on the the robustness of my main empirical results as well as their external validity.

[69]More generally, this analysis, together with the reduced form study of the impact of males home hours and perceptions of females' abilities, suggest that the three key concerns identified in the National Program on Gender Equality could be partially alleviated through my proposal: (a) low male contribution to home hours; (b) females are disproportionally more likely to leave formal education in order to care for others, and (c) unfair treatment in the labor market.

Randomization Based Inference: Placebo Peer Allocation. In order to verify that the effects uncovered are not mechanical nor driven by unobserved factors, I provide a falsification test based on simulating random allocations of students to classes within schools, without replacement, and maintaining the original size of the classes and number of females in each class. I then run the specification in Table 3's column 6 and replicate this exercise 1,000 times. The strength of this exercise is that it allows the researcher to be agnostic about asymptotic distribution and rather empirically construct the nonparametric distributions of test statistics via replication of the randomization procedure (Imbens and Wooldridge, 2009). If we were to find that our baseline effect is consistent with those ones obtained in this placebo analysis, where students are assigned to classmates who were not their actual peers, we may conclude that the main effects were not actually driven by interacting with those peers.

Reassuringly, the distribution of estimated treatment effects reported in Figure 7 shows that the baseline estimate of -0.526 (signaled by the red line) lies outside the 90% confidence interval obtained from the simulations, which is [-0.211, 0.187], as indicated by the dashed green lines.[70]

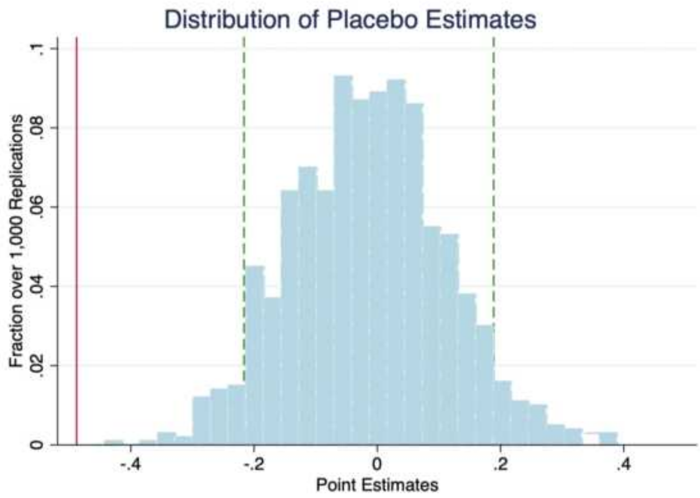
Additionally, in Figure B.17 in the Appendix I proceed in a similar spirit, but this time instead of reshuffling peers, I pool the gender attitudes of all students in a given school and I reshuffle them across all students in the school. If there were broad school-level factors driving my main results, we would find that the statistically significant effects are still present in the counterfactual scenario. One can appreciate that the results are very similar to those in Figure 7.

Coefficient Stability to Selection on Unobservables. Although a standard approach in the empirical literature to strengthen claims of causality is to show that the inclusion of additional controls does not alter the estimate of interest, Oster (2019) cautions against not scaling the changes in point estimates by changes in the R^2 as more controls are included. I compute the coefficient of proportionality δ that defines the importance that unobservables should have relative to observable characteristics in determining our main explanatory variable in order to render its estimated coefficient not statistically different from zero. This computation requires the researcher to take a stand on the highest R^2 that one could achieve under a feasible set of controls. I follow Oster (2019)'s recommendation of setting it to 1.3 times the baseline R^2 (0.29 in Table 3's column 7).[71] I find that selection on unobservables would have to be significantly larger than selection on observables to render my main estimate non-significant — $\delta \approx 2$, well

[70] Complementary, I obtain the empirical p-value as the proportion of total cases in which the absolute value of the point estimate is bigger than the one obtained in my estimation using actual data (Athey and Imbens, 2017). There is no such cases, hence leading to a p-value of 0.000, as in the baseline specification.

[71] She uses a set of studies published in top economics journals and shows that under this heuristics 90% of the studies relying on a randomized control trial remain significant while only 45% of the non-randomized ones continue being so.

Figure 7: Distribution of Placebo-Generated Estimates of the Proportion of Female Peers on Gender Norms (empirical 90% CI indicated by green lines)



above the rule-of-thumb value of 1 established in the literature (e.g. Altonji et al., 2005; Oster, 2019). Similarly, I compute Oster (2019)'s bounds for our estimate of interest under $\delta = 0$ and $\delta = 1$ to be [-0.485, -0.245], which does not contain zero.[72]

Accounting for Influential Observations. Contemporaneous work by Broderick et al. (2020) highlights the potential for estimates obtained even in ex ante very clean empirical setups (e.g. RCTs) to be driven by as few as one observation. I apply their proposed metric to find the set of observations that are most influential in my analysis. This allows me to compute: (i) the proportion of observations that should be discarded from my estimating sample to make my estimate of the proportion of female peers change signs, and (ii) the proportion of observations that should be dropped to make the estimate not only change sign, but also become statistically significant. I find these figures to be 4% and 9%, respectively. These are very large numbers (while the authors are cautious of indicating a rule-of-thumb cutoff, they do suggest that, in practice, 5% for case (ii) is already quite a robust result[73]). Hence, this constitutes strong supportive evidence in favor of my findings[74].

No Selection out of Sample. Section C.1 has shown that the sample employed does not show systematic differences from the original one. It could still be, however, that there is selective attrition in between the school survey and the time when I measure gender-related views

[72]Note also that -0.245, the bound computed under $\delta = 1$ is within 2.8 times the standard error of the estimate under $\delta = 0$ (i.e. 0.103), which is an alternative measure of coefficient stability proposed by Oster (2019).

[73]Indeed, they describe how for some very influential papers, such as Finkelstein et al. (2012)'s Oregon health insurance experiment, dropping just 0.05% of the data is enough to flip the sign and its significance.

[74]In particular, Broderick et al. (2020) emphasize the strength of their approach in acting as a robustness checks for inference (i.e. for the use/treatment of standard errors and their ability to account for non-random sampling).

(the fifth round of the YL survey). I start by noting that this is unlikely: in the fifteen-year period between the first and last waves of the YL survey, the attrition rate was remarkably low, standing at 2.5% (Young Lives, 2017). In my particular case of interest, out of the 1,138 YL children interviewed in the school survey, only 17 (1.5%) did not participate in the fifth round, and only 1 out of the 17 was part of my estimating sample for the main analyses.

Table A.21 formally shows that there are no systematic differences in predetermined characteristics by successful-round-5-matching status. More specifically, none of the dimensions considered is predictive of matching status when introduced separately in columns 1-6 nor jointly in column 7 (the p-value of the F-test of joint significance being 0.631).[75]

External Validity: Cross-country Extensions. Traditional gender norms acting as a source of frictions for economic activity is widespread, as suggested by Figure B.5. One may nevertheless wonder how externally valid my findings are. A first argument in favor of their validity is that while Vietnam displays significant traditionalism relative to Western nations, it is still more modern that other closer countries such as Thailand or Malaysia. Having said this, a particular strength of my study is the unique richness of my data and the adequacy of the empirical setting favoring credible identification. This hinders replicability to other countries using currently available datasets.

Fortunately, Young Lives has the unique feature of providing harmonized information also for Peru, India, and Ethiopia. The School Surveys were, however, conducted with large degrees of freedom across countries, as different needs/policy foci were identified for each nation. While this prevents me from systematically replicating the analyses across all countries, I am able to follow a similar strategy for Ethiopia, for which I do observe the *full* class composition but the sample size is considerably reduced. In Figure B.19 I show that, after residualizing both the gender norms views and the indicator for the class having at least 50% of females (which creates the variation in the horizontal axis instead of only having just two mass points) and netting out school fixed effects, I find again a negative correlation between exposure to females and traditionalism among girls, but no effect for males. This is confirmed through formally estimating Equation 1, where girls with over 50% of female classmates display views on gender norms that are 19% of a standard deviation less traditional three years later after exposure (p-value 0.024; N=200). This suggests that my findings for Vietnam have the potential to be — at least partly — externally valid across very different cultures and institutional systems.

Additional Robustness Checks. In Online Appendix section C.9 I show that the main

[75] Another form of selection would be if children change schools during Grades 1-5 as a response to the gender composition of their class. In my estimating sample, only 18 out of 946 (2%) change schools between those Grades. Applying semiparametric sample selection models like D'Haultfœuille et al. (2018) yields consistent results to those in the baseline model.

results are not sensitive to, among others: (i) placebo outcomes that should not be affected by the treatment (e.g. family resources, height, sibling's outcomes); (ii) discarding the three schools for which the Fisher's exact test rejected independence of observable characteristics across classes; (iii) further accounting for potential selection into schools (beyond school fixed effects) through re-estimating the baseline equation for the subsample of schools with and without other primary schools located nearby that could also attract families, and through the inclusion of spatially more aggregate fixed effects (community-level).

8 Conclusion

Identity norms play a key role throughout a person's life. In the aggregate, they have important consequences on welfare as they can, for instance, lead to severe misallocation of talent across occupations. In this paper I provide the first evidence within the social interactions literature answering the question: does increased exposure to females at schools during childhood shift attitudes towards gender norms and, if so, how persistent are these changes and how do they operate?

My focus is on Vietnam, a country with marked gender gaps and strong traditional gender roles, even among the current youth. I take advantage of an ideal setting that allows me to estimate the causal effects of female classmates on the long-run views on gender norms in the absence of selection into groups and other commonplace threats to identification in the interpersonal contact literature (Manski, 1993). A rich battery of tests lends strong support to the identification assumption based on exogenous assignment of students to classes within schools. The main result is that a ten percentage points increase in the proportion of female classmates leads to a reduction of about a tenth of a standard deviation in the agreement with more traditional views both among males and females.

While a concern could be that survey responses may not translate into actual behavior, I exploit rich information on child's outcomes to show that, on the female side, there is a large increase in college enrollment nine years after the exposure to my group of peers of interest. Mediation analysis suggest that this is driven by increased expectations (both academic and in terms of the returns to tertiary education), higher self-confidence, and more ambitious career goals (including undertaking more male-dominated jobs). On the male side, home production, a key gendered-outcome among adolescents, is significantly higher. Using friendship nominations I show that exposure and familiarity with females is likely an important channel explaining the shifts in views and behavior uncovered: male children spend more time and develop more ties with female peers. This is in line with the fact that the spillovers from female peers into academic outcomes — both cognitive and noncognitive — are relatively minor and short-lived, which is at odds with a story of pure improvements in education leading to more inclusive views.

Overall, my findings provide a rich picture of the role of inter-gender interactions and integration, both in the scope of outcomes explored and in the extended time-frame along which they are observed. Both of these dimensions are needed in order to better inform policy. For instance, going beyond exclusively accounting for academic spillovers highlights the limitations that obviating socio-emotional gains arising from certain class configurations can yield. Moreover, finding sustained impacts over time suggests that policy can indeed have a sizable impact. This is crucial as the existing evidence on whether policy can impact culture is scarce (Bau, 2021).

Based on my results, I propose to benefit from the nonlinearities in treatment effects that I uncover through a simple and feasible policy for gender-mixing that, while preserving close to gender parity and not requiring compositional changes of students across schools, would significantly reduce traditionalism and increase female education. This step towards women facing less traditional gender roles in the society is prone to further help increase female labor force participation and career development opportunities, particularly in the numerous sectors where they are under-represented, which is a long-standing concern in a country where one in every five job postings makes explicit gender requirements (International Labour Organization, 2015). This, in turn, would have implications for the female spouse's bargaining power within the household, as well as intergenerational spillovers arising from the transmission of attitudes from parents to children and the fact that child outcomes have been shown to be enhanced when the mother has a heavier weight in family decisions (e.g. Hoddinott and Haddad, 1995; Duflo, 2003). At the aggregate level, higher female power in then expected to cause economic growth (Doepke et al., 2012). Designing empirical frameworks specifically targeted at further exploring the nonlinear gains from the proposed class allocations as well as the extent to which such benefits persist if a child's assignment to a male or female-dominated class is changed across years emerges as an exciting avenue for future research.

References

Alesina, Alberto and Nicola Fuchs-Schündeln, "Goodbye Lenin (or not?): The effect of communism on people's preferences," *American Economic Review*, 2007, *97* (4), 1507–1528. (return to page 7)

Alger, Ingela and Donald Cox, "The evolution of altruistic preferences: mothers versus fathers," *Review of Economics of the Household*, 2013, *11* (3), 421–446. (return to page Appendix - 47)

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 2005, *113* (1), 151–184. (return to page 45)

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495. (return to page 22)

Anelli, Massimo and Giovanni Peri, "The Effects of High School Peers' Gender on College Major, College Performance and Income," *The Economic Journal*, 2019. (return to page 6)

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik, "The effect of teacher gender on student achievement in primary school," *Journal of Labor Economics*, 2015, *33* (1), 63–89. (return to page Appendix - 43)

Ashraf, Nava, Erica Field, and Jean Lee, "Household bargaining and excess fertility: an experimental study in Zambia," *American Economic Review*, 2014, *104* (7), 2210–37. (return to page Appendix - 47)

_ , Natalie Bau, Corinne Low, and Kathleen McGinn, "Negotiating a better future: How interpersonal skills facilitate intergenerational investment," *The Quarterly Journal of Economics*, 2020, *135* (2), 1095–1151. (return to page 6)

Athey, Susan and Guido W Imbens, "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, 2017, *31* (2), 3–32. (return to page 44)

Attanasio, Orazio P and Katja M Kaufmann, "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender," *Journal of Development Economics*, 2014, *109*, 203–216. (return to page 32)

Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor, "Does affirmative action work? Caste, gender, college quality, and academic success in India," *American Economic Review*, 2016, *106* (6), 1495–1521. (return to page 2)

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Women's empowerment in action: evidence from a randomized control trial in Africa," *American Economic Journal: Applied Economics*, 2020, *12* (1), 210–59. (return to page 8)

Bau, Natalie, "Can policy change culture? Government pension plans and traditional kinship practices," *American Economic Review*, 2021, *111* (6), 1880–1917. (return to page 2, 7, 48)

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova, "Powerful women: does exposure reduce bias?," *The Quarterly Journal of Economics*, 2009, *124* (4), 1497–1540. (return to page 2)

Becker, Gary S, "An economic analysis of fertility," in "Demographic and economic change in developed countries," Columbia University Press, 1960, pp. 209–240. (return to page Appendix - 47)

Behr, Helen, "Comparing rural and urban primary education in the Mekong Delta," 2005. (return to page 11)

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 2014, *81* (2), 608–650. (return to page 22)

Bertrand, Marianne, Emir Kamenica, and Jessica Pan, "Gender identity and relative income within households," *The Quarterly Journal of Economics*, 2015, *130* (2), 571–614. (return to page 11)

_ , Patricia Cortes, Claudia Olivetti, and Jessica Pan, "Social Norms, Labor Market Opportunities, and the Marriage Gap Between Skilled and Unskilled Women," *The Review of Economic Studies*, 2020. (return to page 2)

Bettinger, Eric P and Bridget Terry Long, "Do faculty serve as role models? The impact of instructor gender on female students," *American Economic Review*, 2005, *95* (2), 152–157. (return to page Appendix - 43)

Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian, "Gender stereotypes about intellectual ability emerge early and influence children's interests," *Science*, 2017, *355* (6323), 389–391. (return to page 4)

Bietenbeck, Jan, "The long-term impacts of low-achieving childhood peers: Evidence from project STAR," *Journal of the European Economic Association*, 2020, *18* (1), 392–426. (return to page 6, 20, 31)

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes, "Under pressure? The effect of peers on outcomes of young adults," *Journal of Labor Economics*, 2013, *31* (1), 119–153. (return to page 6)

Blundell, Jack and Erling Risa, "Income and Family Background: Are we Using the Right Models?," *Available at SSRN 3269576*, 2019. (return to page Appendix - 39)

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794. (return to page 2)

Brenøe, Anne Ardila and Ulf Zölitz, "Exposure to more female peers widens the gender gap in stem participation," *Journal of Labor Economics*, 2020, *38* (4), 1009–1054. (return to page 5, 31, Appendix - 47)

Broderick, Tamara, Ryan Giordano, and Rachael Meager, "An Automatic Finite-Sample Robustness Metric: Can Dropping a Little Data Change Conclusions?," *arXiv preprint arXiv:2011.14999*, 2020. (return to page 16, 45)

Brundrett, Mark and Mai Thi Thuy Dung, "The challenge of ensuring gender equality in Vietnamese and English high schools," *International Journal of Comparative Education and Development*, 2018. (return to page Appendix - 43)

Carlana, Michela, "Implicit stereotypes: Evidence from teachers' gender bias," *The Quarterly Journal of Economics*, 2019, *134* (3), 1163–1224. (return to page 18, Appendix - 44, Appendix - 52)

Carrell, Scott E, Bruce I Sacerdote, and James E West, "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 2013, *81* (3), 855–882. (return to page 5, 40)

_ , Mark Hoekstra, and Elira Kuka, "The long-run effects of disruptive peers," *American Economic Review*, 2018, *108* (11), 3377–3415. (return to page 6, 31)

_ , _ , and James E West, "The impact of college diversity on behavior toward minorities," *American Economic Journal: Economic Policy*, 2019, *11* (4), 159–82. (return to page 7)

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 2011, *126* (4), 1593–1660. (return to page Appendix - 42)

Chiplunkar, Gaurav and Pinelopi K Goldberg, "Aggregate Implications of Barriers to Female Entrepreneurship," Technical Report, National Bureau of Economic Research 2021. (return to page 5)

Choudhury, Suparna, Sarah-Jayne Blakemore, and Tony Charman, "Social cognitive development during adolescence," *Social Cognitive and Affective Neuroscience*, 2006, *1* (3), 165–174. (return to page 6)

Chung, Bobby W, "Peers' parents and educational attainment: The exposure effect," *Labour Economics*, 2020, *64*, 101812. (return to page Appendix - 53)

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, Geoff Kenny, and Michael Weber, "The Effect of Macroeconomic Uncertainty on Household Spending," Technical Report, National Bureau of Economic Research 2021. (return to page 32)

Cools, Angela, Raquel Fernández, and Eleonora Patacchini, "Girls, boys, and high achievers," Technical Report, National Bureau of Economic Research 2019. (return to page 8)

Croson, Rachel and Uri Gneezy, "Gender differences in preferences," *Journal of Economic Literature*, 2009, *47* (2), 448–74. (return to page Appendix - 46)

Dahl, Gordon, Andreas Kotsadam, and Dan-Olof Rooth, "Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams," *Quarterly Journal of Economics*, forthcoming. (return to page 6)

Dang, Hai-Anh H and Paul W Glewwe, "Well begun, but aiming higher: A review of Vietnam's education trends in the past 20 years and emerging challenges," *The Journal of Development Studies*, 2018, *54* (7), 1171–1195. (return to page 14)

Daruich, Diego, "The macroeconomic consequences of early childhood development policies," *FRB St. Louis Working Paper*, 2018, (2018-29). (return to page 7)

Dee, Thomas S, "Teachers and the gender gaps in student achievement," *Journal of Human Resources*, 2007, *42* (3), 528–554. (return to page 22)

Dhar, Diva, Tarun Jain, and Seema Jayachandran, "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India," Technical Report, National Bureau of Economic Research 2018. (return to page 4, 7)

Doepke, Matthias and Michèle Tertilt, "Families in macroeconomics," in "Handbook of Macroeconomics," Vol. 2, Elsevier, 2016, pp. 1789–1891. (return to page Appendix - 47)

__ , **Michele Tertilt, and Alessandra Voena**, "The economics and politics of women's rights," *Annual Review of Economics*, 2012, *4* (1), 339–372. (return to page 48)

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner, "Individual risk attitudes: Measurement, determinants, and behavioral consequences," *Journal of the European Economic Association*, 2011, *9* (3), 522–550. (return to page Appendix - 46)

Duflo, Esther, "Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa," *The World Bank Economic Review*, 2003, *17* (1), 1–25. (return to page 48)

D'Haultfœuille, Xavier, Arnaud Maurel, and Yichong Zhang, "Extremal quantile regressions for selection models and the black–white wage gap," *Journal of Econometrics*, 2018, *203* (1), 129–142. (return to page 46)

Favara, Marta, Paul Glewwe, Catherine Porter, and Alan Sanchez, "Expecting Better? How Young People Form Their Earnings Expectations," 2021. (return to page 32)

Feld, Jan and Ulf Zölitz, "Understanding peer effects: On the nature, estimation, and channels of peer effects," *Journal of Labor Economics*, 2017, *35* (2), 387–428. (return to page 12, Appendix - 42)

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti, "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299. (return to page 2)

__ **and** __ , "Fertility: The role of culture and family experience," *Journal of the European Economic Association*, 2006, *4* (2-3), 552–561. (return to page 2)

Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Moore, "On her own account: How strengthening women's financial control impacts labor supply and gender norms," *American Economic Review*, 2021, *111* (7). (return to page 2, 7, 8)

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group, "The Oregon health insurance experiment: evidence from the first year," *The Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106. (return to page 45)

Fortin, Nicole M, "Gender role attitudes and women's labor market participation: Opting-out, aids, and the persistent appeal of housewifery," *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 2015, (117/118), 379–401. (return to page 5)

Fruehwirth, Jane Cooley and Jessica Gagete-Miranda, "Your peers' parents: Spillovers from parental education," *Economics of Education Review*, 2019, *73*, 101910. (return to page 20, Appendix - 39)

Gao, Ersheng, Xiayun Zuo, Li Wang, Chaohua Lou, Yan Cheng, and Laurie S Zabin, "How does traditional Confucian culture influence adolescents' sexual behavior in three Asian cities?," *Journal of Adolescent Health*, 2012, *50* (3), S12–S17. (return to page 22)

Gelbach, Jonah B, "When do covariates matter? And which ones, and how much?," *Journal of Labor Economics*, 2016, *34* (2), 509–543. (return to page 5, 33)

Gong, Jie, Yi Lu, and Hong Song, "The effect of teacher gender on students' academic and noncognitive outcomes," *Journal of Labor Economics*, 2018, *36* (3), 743–778. (return to page 6)

_ , _ , and _ , "Gender peer effects on students' academic and noncognitive outcomes: Evidence and mechanisms," *Journal of Human Resources*, 2019, pp. 0918–9736R2. (return to page 4, 29)

Gould, Eric D, Victor Lavy, and M Daniele Paserman, "Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence," *The Economic Journal*, 2009, *119* (540), 1243–1269. (return to page 6)

Goussé, Marion, Nicolas Jacquemet, and Jean-Marc Robin, "Marriage, labor supply, and home production," *Econometrica*, 2017, *85* (6), 1873–1919. (return to page 11)

Guilmoto, Christophe, "Recent change in the sex ratio at birth in Viet Nam: a review of evidence," 2009. (return to page 12)

Guiso, Luigi, Tullio Jappelli, and Luigi Pistaferri, "An empirical analysis of earnings and employment risk," *Journal of Business & Economic Statistics*, 2002, *20* (2), 241–253. (return to page 32)

Guryan, Jonathan, Kory Kroft, and Matthew J Notowidigdo, "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 34–68. (return to page Appendix - 42)

Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe, "Do risk preferences change? Evidence from the great East Japan earthquake," *American Economic Journal: Applied Economics*, 2018, *10* (2), 298–330. (return to page Appendix - 46)

Hanushek, Eric A, "The failure of input-based schooling policies," *The Economic Journal*, 2003, *113* (485), F64–F98. (return to page Appendix - 42)

Hoddinott, John and Lawrence Haddad, "Does female income share influence household expenditures? Evidence from Côte d'Ivoire," *Oxford Bulletin of Economics and Statistics*, 1995, *57* (1), 77–96. (return to page 48)

Hospido, Laura, Luc Laeven, and Ana Lamo, "The gender promotion gap: evidence from central banking," *Review of Economics and Statistics*, 2019, pp. 1–45. (return to page 2)

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474. (return to page 2, 5)

Hwang, Jisoo, Chulhee Lee, and Esther Lee, "Gender norms and housework time allocation among dual-earner couples," *Labour Economics*, 2019, *57*, 102–116. (return to page 2, 4, 7)

Hyun, Yeseul, "Does Intrafamily Time Allocation Respond to Economic Shocks? The Role of Gender Norms," 2020. (return to page 7)

Imbens, Guido W and Jeffrey M Wooldridge, "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 2009, *47* (1), 5–86. (return to page 16, 44)

International Labour Organization, "Gender Equality in Recruitment and Promotion Practices in Viet Nam," ILO's Policy Brief 2015. (return to page 48)

ISDS, "Social determinants of gender inequality in Vietnam," 2016. (return to page 42)

Iyer, Padmini, Obiageri Bridget Azubuike, and Caine Rolleston, "Young Lives School Survey 2016-17: Evidence from Vietnam," *Country Report, Oxford: Young Lives*, 2017. (return to page Appendix - 54)

James, Zoe, "Young Lives Vietnam school survey round 1 (2011-2012): Data user guide," 2013. (return to page Appendix - 39)

Jayachandran, Seema, "The roots of gender inequality in developing countries," *economics*, 2015, *7* (1), 63–88. (return to page 4, 5)

— , "Microentrepreneurship in developing countries," 2020. (return to page 6)

Jensen, Robert, "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, 2010, *125* (2), 515–548. (return to page 32)

Jiang, Xinyan, "Confucianism, women, and social contexts," *Journal of Chinese philosophy*, 2009, *36* (2), 228–242. (return to page 22)

Kazianga, Harounan, Dan Levy, Leigh L Linden, and Matt Sloan, "The effects of" girl-friendly" schools: Evidence from the BRIGHT school construction program in Burkina Faso," *American Economic Journal: Applied Economics*, 2013, *5* (3), 41–62. (return to page 5)

Krueger, Alan B, "Experimental estimates of education production functions," *The Quarterly Journal of Economics*, 1999, *114* (2), 497–532. (return to page 17)

Lavy, Victor and Analia Schlosser, "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, 2011, *3* (2), 1–33. (return to page 6, 31)

— **and Edith Sand**, "On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases," Technical Report 2019. (return to page 19)

Le, TTBT, "Labour division and decision-making in families," *Tran. TVA and Nguyen, HM (eds.) Gender equality in Vietnam. Hanoi: Social Science Publisher*, 2008, pp. 142–173. (return to page Appendix - 30)

Levitt, Steven D and John A List, "What do laboratory experiments measuring social preferences reveal about the real world?," *Journal of Economic Perspectives*, 2007, *21* (2), 153–174. (return to page 7)

Li, Lixing and Xiaoyu Wu, "Gender of children, bargaining power, and intrahousehold resource allocation in China," *Journal of Human Resources*, 2011, *46* (2), 295–316. (return to page 43)

Lim, Jaegeum and Jonathan Meer, "The impact of teacher–student gender matches random assignment evidence from South Korea," *Journal of Human Resources*, 2017, *52* (4), 979–997. (return to page Appendix - 43)

Liu, Amy YC, "Sectoral gender wage gap in Vietnam," *Oxford Development Studies*, 2004, *32* (2), 225–239. (return to page Appendix - 38)

Lovenheim, Michael F and C Lockwood Reynolds, "The effect of housing wealth on college choice: Evidence from the housing boom," *Journal of Human Resources*, 2013, *48* (1), 1–35. (return to page Appendix - 39)

Lowe, Matt, "Types of contact: A field experiment on collaborative and adversarial caste integration," *American Economic Review*, forthcoming. (return to page 7)

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542. (return to page 3, 16, 47)

Martorell, Reynaldo and Jean-Pierre Habicht, "Growth in early childhood in developing countries," *Human growth: a comprehensive treatise*, 1986, *3*, 241–262. (return to page Appendix - 50)

McPherson, Miller, Lynn Smith-Lovin, and James M Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, 2001, *27* (1), 415–444. (return to page Appendix - 44)

Mergoupis, Thanos, Van Phan, and John G Sessions, "Puzzle me this? The Vietnamese reverse gender education gap," Technical Report, WIDER Working Paper 2018. (return to page 27)

Mousa, Salma, "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq," *Science*, 2020, *369* (6505), 866–870. (return to page 6)

Muralidharan, Karthik and Ketki Sheth, "Bridging education gender gaps in developing countries: The role of female teachers," *Journal of Human Resources*, 2016, *51* (2), 269–297. (return to page 24)

Nguyen, Cuong, "Gender Equality in Education, Health Care, and Employment: Evidence from Vietnam," 2012. (return to page 12)

Nguyen, Ngoc P, "An assessment of the Young Lives sampling approach in Vietnam," 2008. (return to page Appendix - 38)

Nguyen, The Cuong, Abdul Hafeez-Baig, Raj Gururajan, and Nam C Nguyen, "The hidden reasons of the Vietnamese parents for paying private tuition fees for public school teachers," *Social Sciences & Humanities Open*, 2021, *3* (1), 100105. (return to page 14)

Oanh, Nguyen Hoang and Nguyen Hong Ngoc, "Gender pay gap in Vietnam: a propensity score matching analysis," *Journal of Economics and Development*, 2020. (return to page Appendix - 38)

Oreopoulos, Philip and Kjell G Salvanes, "Priceless: The nonpecuniary benefits of schooling," *Journal of Economic Perspectives*, 2011, *25* (1), 159–84. (return to page 8, 42)

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204. (return to page 16, 44, 45)

Paluck, Elizabeth Levy, Seth A Green, and Donald P Green, "The contact hypothesis re-evaluated," *Behavioural Public Policy*, 2019, *3* (2), 129–158. (return to page 6)

Quinn, Simon and Christopher Woodruff, "Experiments and entrepreneurship in developing countries," *Annual Review of Economics*, 2019, *11*, 225–248. (return to page 6)

Rao, Gautam, "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools," *American Economic Review*, 2019, *109* (3), 774–809. (return to page 7)

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar, "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender," *The Economic Journal*, 2017, *127* (604), 2153–2186. (return to page 8)

Robinson, Peter M, "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, 1988, pp. 931–954. (return to page 25)

Rolleston, Caine, Zoe James, Laure Pasquier-Doumer, and Tran Ngo Thi Minh Tam, *Making progress: Report of the Young Lives school survey in Vietnam*, Young Lives, 2013. (return to page 12)

Sacerdote, Bruce, "Peer effects in education: How might they work, how big are they and how much do we know thus far?," in "Handbook of the Economics of Education," Vol. 3, Elsevier, 2011, pp. 249–277. (return to page 5)

Shan, Xiaoyue, "Does Minority Status Drive Women Out of Male-Dominated Fields?," *Unpublished working paper*, 2021. (return to page 27)

Singh, Abhijeet, "Learning more with every year: School year productivity and international learning divergence," *Journal of the European Economic Association*, 2020, *18* (4), 1770–1813. (return to page 3, 5, 31, 42)

Sojourner, Aaron, "Identification of peer effects with missing peer data: Evidence from Project STAR," *The Economic Journal*, 2013, *123* (569), 574–605. (return to page 17)

Stinebrickner, Todd and Ralph Stinebrickner, "Learning about academic ability and the college dropout decision," *Journal of Labor Economics*, 2012, *30* (4), 707–748. (return to page 4)

Tuan, Nguyen Thanh, Pham Duy Tuong, and BM Popkin, "Body mass index (BMI) dynamics in Vietnam," *European Journal of Clinical Nutrition*, 2008, *62* (1), 78–86. (return to page Appendix - 23)

van Lent, Max, "Fathering Daughters and Personality," *Available at SSRN 3742024*, 2020. (return to page 43)

Westoff, Charles F et al., "Desired number of children: 2000-2008.," *DHS Comparative Reports*, 2010, (25). (return to page Appendix - 47)

Wiswall, Matthew and Basit Zafar, "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 2015, *82* (2), 791–824. (return to page 4)

World Bank, "Vietnam: Breaking gender stereotypes that hinder women's empowerment," World Bank's East Asia Pacific on the Rise Series 2015. (return to page 3)

Young Lives, "Young Lives Survey Design and Sampling in Vietnam," 2014. (return to page 8, Appendix - 38)

_ , "Young Lives Survey Design and Sampling (Round 5): Viet Nam," 2017. (return to page 46)

Zölitz, Ulf and Jan Feld, "The effect of peer gender on major choice in business school," forthcoming. (return to page 4)

A Additional Tables (For Online Publication)

Table A.1: Balance of Characteristics at Round 2 (Age 5) by Status for School Survey Inclusion

Panel A						
	(1)	(2)	(3)	(4)	(5)	(6)
	Rural	Female	Low-BMI-for-age	Long-term Health Issue	Enrolled	Father's Education
In School Survey	-0.004	0.012	0.009	-0.010	0.257***	0.007
	(0.003)	(0.025)	(0.009)	(0.015)	(0.020)	(0.162)
Mean (sd) Dep. Var.	0.79 (0.40)	0.49 (0.50)	0.04 (0.19)	0.09 (0.029)	0.83 (0.38)	6.59 (3.57)
Observations	1,970	1,970	1,958	1,967	1,911	1,894
R-squared	0.922	0.006	0.020	0.035	0.171	0.302

Panel B						
	(7)	(8)	(9)	(10)	(11)	(12)
	Mother's Education	Father's age	Mother's age	Household Size	Atheist	Wealth Index
In School Survey	0.033	0.136	0.363	-0.044	-0.015	0.007
	(0.155)	(0.291)	(0.283)	(0.076)	(0.017)	(0.007)
Mean (sd) Dep. Var.	6.10 (3.64)	34.06 (5.99)	31.19 (5.77)	4.67 (1.51)	0.86 (0.19)	0.52 (0.19)
Observations	1,942	1,907	1,955	1,970	1,969	1,949
R-squared	0.356	0.095	0.087	0.089	0.125	0.490

Notes. Regressions of the indicated outcome variables (measured in 2007 for Round 2 of the YL survey) on an indicator taking the value 1 if the Young Lives child was sampled for the School Survey and 0 otherwise and sentinel site fixed effects. The sample are all Young Lives students present in round 2. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Balance of Characteristics between YL Children and their Classmates at the 2011 School Survey (Age 10)

Panel A

	(1)	(2)	(3)	(4)	(5)	(6)
	Female	Any Parent Illiterate	Ethnic Minority	No Health Problems	Home Educational Resources	Wealth Index
YL Child	0.026	-0.011	-0.012	0.012	-0.006	-0.003
	(0.020)	(0.010)	(0.009)	(0.019)	(0.010)	(0.006)
Mean (sd) Dep. Var.	0.47 (0.50)	0.11 (0.31)	0.12 (0.33)	0.71 (0.46)	0.71 (0.27)	0.52 (0.18)
Observations	2,956	2,965	2,962	2,971	2,971	2,971
R-squared	0.020	0.439	0.680	0.096	0.329	0.385

Panel B

	(7)	(8)	(9)	(10)	(11)
	Number Books	If I Work Hard I can go to College	No Reinforcement Class	Enjoy School	Worry about Exams
YL Child	-0.046	-0.024	-0.002	-0.004	-0.011
	(0.042)	(0.032)	(0.017)	(0.011)	(0.028)
Mean (sd) Dep. Var.	1.79 (1.22)	1.63 (0.78)	0.63 (0.48)	1.05 (0.24)	1.91 (0.71)
Observations	2,953	2,942	2,971	2,945	2,942
R-squared	0.271	0.084	0.350	0.051	0.094

Notes. Regressions of the indicated outcome variables (measured in the first round of the school survey) on an indicator taking the value 1 if the child belongs to the longitudinal Young Lives study and 0 otherwise and school fixed effects. The sample contains all students with non-missing values present in the School Survey whose class-assignment satisfied the sample selection criteria outlined in Section 2.1. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Descriptive Statistics of Attitudes towards Gender Norms for the Old Cohort

	(1) Mean	(2) Standard Deviation	(3) Min	(4) Max	(5) Count
Life Purpose					
Life Goals	0.56	0.50	0	1	910
Freedom	0.05	0.22	0	1	909
Family Encouragement for College	0.16	0.37	0	1	908
Importance of Good Academic Performance	0.15	0.36	0	1	910
Abilities					
Leadership Ability	0.45	0.50	0	1	908
Intelligence	0.06	0.23	0	1	908
Cross-Gender Interactions					
Asking for a Date	0.11	0.32	0	1	909
Pay for a Date	0.55	0.50	0	1	908
Male-dominated Sports	0.07	0.26	0	1	909
Swearing	0.58	0.49	0	1	908
Intra-household Decisions					
Authority in Household Decisions	0.49	0.50	0	1	909
Chores	0.05	0.21	0	1	909
Mean Score (Full Sample)	2.16	0.31	1	3.08	910
Males	2.22	0.30	1.09	3.08	447
Females	2.10	0.31	1	3.08	463

Notes. All variables are indicators taking the value one if the child agrees or strongly agrees with a traditional view on gender norms for each dimension of interest. The exact questions were reported in the "Views on Gender Norms (longitudinal YL)" subsection. Respondents are from the Old Cohort and are interviewed in the fifth round of the YL survey (2016, age 22).

Table A.4: Does Peers' Female Composition Predict Observable Individual Characteristics?

Panel A

	(1)	(2)	(3)	(4)	(5)
	Ethnic Minority	Repeater	Any Health Issue	Any Parent Illiterate	Number of Books at Home
Proportion Female Peers	-0.007	-0.018	0.148	-0.049	0.084
	(0.039)	(0.037)	(0.122)	(0.044)	(0.419)
Mean (sd) Dep. Var.	0.12 (0.32)	0.04 (0.20)	0.29 (0.45)	0.10 (0.31)	1.79 (1.22)
Observations	2,491	2,488	2,498	2,493	2,483
R-squared	0.545	0.045	0.096	0.218	0.216

Panel B

	(6)	(7)	(8)	(9)
	Wealth Index	Home Educational Resources	Takes Private Classes	No Reinforcement Class
Proportion Female Peers	0.053	0.035	0.008	-0.019
	(0.050)	(0.056)	(0.236)	(0.234)
Mean (sd) Dep. Var.	0.52 (0.18)	0.71 (0.27)	0.38 (0.48)	0.63 (0.48)
Observations	2,498	2,498	2,491	2,498
R-squared	0.343	0.205	0.348	0.347

Notes. All outcome variables in Panel A are indicators (except the 4-category variable for number of books at home). All outcomes were measured in the first round of the 2011 School Survey. Regressions control for school fixed effects. The sample contains all students with non-missing values present in the School Survey whose class-assignment satisfied the sample selection criteria outlined in Section 2.1. Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.5: Does Peers' Female Composition Correlate with YL Children's Characteristics as Reported by Teachers?

	(1) Home Academic Support	(2) Academic Ability
Proportion Female Peers	0.016	0.040
	(0.499)	(0.406)
Mean (sd) Dep. Var.	2.56 (0.95)	2.49 (0.81)
Observations	2,894	2,894

Notes. Both outcomes are measured in the baseline round of the 2011 School Survey as perceived by each students' head teacher. Teachers could choose one of the following five categories: very high; high; medium; low; very low (which I code 1 to 5). The models estimated are ordered probits. School fixed effects are included. Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.6: Are Teacher Characteristics Correlated with Peers' Female Composition?

Panel A	(1)	(2)	(3)	(4)	(5)
	Female	Highest Training	Temporary Contract	Has Multiple Jobs	Current Years of Tenure
Proportion Female Peers	0.397	0.108	-0.066	-0.005	-4.498
	(0.253)	(0.538)	(0.146)	(0.255)	(7.233)
Mean (sd) Dep. Var.	0.75 (0.43)	3.27 (0.74)	0.01 (0.12)	0.25 (0.43)	9.76 (7.86)
Observations	2,894	2,875	2,874	2,894	2,894
R-squared	0.569	0.597	0.261	0.582	0.445

Panel B	(6)	(7)	(8)	(9)	(10)
	Total Years of Tenure	Wealth Index	Winner of Excellent Teacher Award	Teacher Believes Possible to Help Worst Students	Pedagogical Ability
Proportion Female Peers	2.519	0.125	0.246	-0.106	-0.005
	(6.254)	(0.081)	(0.218)	(0.762)	(0.126)
Mean (sd) Dep. Var.	17.48 (8.43)	0.70 (0.11)	0.94 (0.25)	-0.06 (1.22)	0.595 (0.123)
Observations	2,924	2,924	2,924	2,904	2,693
R-squared	0.519	0.688	0.379	0.576	0.401

Notes. Regressions of the indicated teacher characteristics (measured in the first round of the 2011 School Survey) on the proportion of female peers, gender, and school fixed effects. The sample contains all students with non-missing values present in the School Survey whose class-assignment satisfied the sample selection criteria outlined in Section 2.1. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Robustness to Different Sample Selections

	(1)	(2)	(3)	(4)
	Mean Agreement with Traditional Gender Norms at R5 (Age 15)			
Proportion Female Peers	-0.486***	-0.489***	-0.372**	-0.556***
	(0.103)	(0.102)	(0.154)	(0.104)
Sample	Random+location+single	Random+location	Random	Full
Observations	880	843	728	1,027
R-squared	0.290	0.286	0.304	0.290

Notes. Replications of the specification in Table 3's column 7 for different subsamples. Column 1 uses individuals assigned "randomly", "by location of residence" or "there was only one Grade 5 class" (i.e. as in column 7 in Table 3). Column 4 employs the full sample of individuals, irrespective of the assignment procedure. Tests of equality of the estimated coefficients in columns 2-4 relative to column 1 do not reject the null (p-values of 0.766, 0.142, and 0.290 for 1 vs. 2, 1 vs. 3, and 1 vs. 4, respectively). Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Heterogeneous Effects of Exposure to Female Teachers for Female Children

Panel A						
	(1)	(2)	(3)	(4)	(5)	(6)
	Life Goals	Freedom	College Encouragement	Importance Academics	Leadership Abilities	Intelligence
Proportion Female Peers	-1.763***	-0.414	0.178	-0.933*	-0.823	0.304
	(0.597)	(0.706)	(0.963)	(0.549)	(0.722)	(0.525)
Female Teacher	0.957***	-0.214	0.799***	0.266	0.547	0.608***
	(0.309)	(0.355)	(0.207)	(0.360)	(0.414)	(0.204)
Teacher believes his/her job will influence the kids' life	0.422**	-0.095	0.270	-0.245	0.019	0.534***
	(0.173)	(0.159)	(0.181)	(0.185)	(0.188)	(0.096)
Female Teacher*Job Can Influence	-0.426***	0.009	-0.359**	0.048	-0.175	-0.343***
	(0.154)	(0.136)	(0.172)	(0.176)	(0.188)	(0.077)
Observations	423	424	423	422	424	423
R-squared	0.281	0.289	0.335	0.308	0.312	0.313
Panel B						
	(7)	(8)	(9)	(10)	(11)	(12)
	Ask for Date	Pay for Date	Sports	Swearing	Household Authority	Chores
Proportion Female Peers	0.400	-2.304***	-2.167**	0.627	0.644	0.623
	(0.795)	(0.855)	(0.851)	(0.989)	(0.911)	(0.855)
Female Teacher	0.125	0.492*	1.121***	0.401	0.866**	1.075***
	(0.293)	(0.286)	(0.353)	(0.487)	(0.345)	(0.256)
Teacher believes his/her job can influence kids' lives	0.130	0.148	0.341*	0.250	0.327*	0.551***
	(0.131)	(0.178)	(0.195)	(0.237)	(0.183)	(0.126)
Female Teacher*Job Can Influence	-0.239**	-0.147	-0.354*	-0.440*	-0.343**	-0.632***
	(0.120)	(0.167)	(0.182)	(0.232)	(0.172)	(0.124)
Observations	414	419	424	423	423	424
R-squared	0.334	0.277	0.329	0.180	0.353	0.263

Notes. Replication of the specification in Table 3's column 7 for the various dimensions of gender roles. "Teacher believes his/her job can influence kids' lives" is the principal component out of the degree of agreement (after recoding the responses when needed so that higher values mean more agreement with teachers being able to influence students' outcomes and standardizing the responses previously given in a four-point scale) with the following three statements: (i) "if I try hard I can get through to even the most difficult or unmotivated students"; (ii) "the amount a student can learn is primarily related to family background"; (iii) "when I really try, I can get through to most difficult students"; (iv) "I am very limited in what I can achieve because a student's home environment is a large influence on his/her achievement"; (v) "teachers are not a very powerful influence on student achievement when all factors are considered", and (vi) "even a teacher with good teaching abilities may not reach many students". The sample is female children. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Heterogeneous Effects of Exposure to Female Teachers for Male Children

Panel A						
	(1)	(2)	(3)	(4)	(5)	(6)
	Life Goals	Freedom	College Encouragement	Importance Academics	Leadership Abilities	Intelligence
Proportion Female Peers	-0.405	-0.856*	-0.315	-1.410*	-0.629	-1.211**
	(0.551)	(0.439)	(0.591)	(0.807)	(0.488)	(0.540)
Female Teacher	-0.329	-0.505***	-0.920***	-0.729***	-0.016	-0.343*
	(0.239)	(0.146)	(0.265)	(0.236)	(0.271)	(0.187)
Teacher believes his/her job will influence the kids' life	-0.083	-0.310***	-0.336**	-0.413*	0.181	0.081
	(0.146)	(0.084)	(0.134)	(0.209)	(0.161)	(0.164)
Female Teacher*Job Can Influence	-0.172	0.419***	0.411***	0.318	-0.139	-0.016
	(0.129)	(0.065)	(0.130)	(0.204)	(0.156)	(0.152)
Observations	426	426	425	426	426	426
R-squared	0.289	0.271	0.273	0.276	0.322	0.309
Panel B						
	(7)	(8)	(9)	(10)	(11)	(12)
	Ask for Date	Pay for Date	Sports	Swearing	Household Authority	Chores
Proportion Female Peers	-0.825*	-1.316**	-2.549***	-1.800***	-1.269	0.465
	(0.463)	(0.620)	(0.714)	(0.568)	(0.772)	(0.629)
Female Teacher	-0.010	-0.157	-0.555*	-0.356*	-1.024***	-0.335
	(0.237)	(0.232)	(0.292)	(0.206)	(0.183)	(0.242)
Teacher believes his/her job can influence kids' lives	0.172	-0.397***	-0.478***	-0.166	-0.284**	-0.295**
	(0.118)	(0.115)	(0.140)	(0.125)	(0.142)	(0.128)
Female Teacher*Job Can Influence	-0.017	0.177*	0.489***	0.160	0.159	0.187
	(0.116)	(0.095)	(0.121)	(0.128)	(0.120)	(0.117)
Observations	424	426	426	426	426	426
R-squared	0.259	0.317	0.250	0.263	0.280	0.281

Notes. Replication of the specification in Table 3's column 7 for the various dimensions of gender roles. "Teacher believes his/her job can influence kids' lives" is the principal component out of the degree of agreement (after recoding the responses when needed so that higher values mean more agreement with teachers being able to influence students' outcomes and standardizing the responses previously given in a four-point scale) with the following three statements: (i) "if I try hard I can get through to even the most difficult or unmotivated students"; (ii) "the amount a student can learn is primarily related to family background"; (iii) "when I really try, I can get through to most difficult students"; (iv) "I am very limited in what I can achieve because a student's home environment is a large influence on his/her achievement"; (v) "teachers are not a very powerful influence on student achievement when all factors are considered", and (vi) "even a teacher with good teaching abilities may not reach many students". The sample is male children. Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.10: Exploring Class vs. Cohort Effects

	(1)	(2)
	Mean Agreement with Traditional Gender Norms (YL R5, Age 15)	
Proportion Female Peers	-0.465***	-0.467***
	(0.112)	(0.126)
Observations	548	248
R-squared	0.282	0.300

Notes. Replication of the specification in Table 3's column 7. I create three categories: (i) below 47%; (ii) between 47 and 53%, and (iii) above 53%. Column 1 restrict the sample to the subset of schools with classes belonging to at least two of these categories. Column 2 restricts the sample to schools with at least one class belonging to category (i) and one to category (iii). Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.11: Effects of the Proportion of Female Peers on the Female Composition of Female Children's Friendship Networks (Second Round of School Survey, Age 10-11)

Panel A: Degree of Friendship			
	(1)	(2)	(3)
Female Friends/Total Friends defined as...	At Least a Little Close	At Least Close	Very Close
Proportion Female Peers	1.049***	0.241	-0.004
	(0.154)	(0.147)	(0.106)
Mean (sd) Dep. Var.	0.63 (0.22)	0.81 (0.23)	0.89 (0.23)
Observations	438	432	368
R-squared	0.428	0.358	0.330

Panel B: Contact Outside School			
	(4)	(5)	(6)
Female Friends/Total Friends defined as...	At Least not Many	At Least Quite a Lot	A Lot
Proportion Female Peers	0.882***	0.421***	0.158
	(0.159)	(0.121)	(0.169)
Mean (sd) Dep. Var.	0.67 (0.23)	0.80 (0.25)	0.87 (0.26)
Observations	429	405	325
R-squared	0.379	0.351	0.268

Panel C: Combined Measures			
	(7)	(8)	(9)
Female Friends/Total Friends defined as...	1+4	2+5	3+6
Proportion Female Peers	0.947***	0.315**	0.128
	(0.166)	(0.129)	(0.142)
Mean (sd) Dep. Var.	0.71 (0.24)	0.85 (0.23)	0.91 (0.22)
Observations	427	391	282
R-squared	0.397	0.377	0.417

Notes. All regressions replicate the specification in Table 11 for the subsample of female children. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.12: Effects of the Proportion of Female Peers on Risk Attitudes and Behaviors (YL Covid Round, Age 19)

	(1) Above 10,000 Pay Lottery	(2) Above 10,000 Pay Lottery	(3) Mean Protective Measures R1	(4) Mean Protective Measures R2	(5) Mean Protective Measures R1	(6) Mean Protective Measures R2
Proportion Female Peers	-0.523*	0.021	0.180**	0.157***	0.097	0.012
	(0.293)	(0.233)	(0.087)	(0.051)	(0.087)	(0.082)
Sample	Males	Females	Males	Males	Females	Females
Mean (sd) Dep. Var.	0.34 (0.47)	0.23 (0.42)	0.43 (0.18)	0.28 (0.13)	0.48 (0.17)	0.29 (0.13)
Observations	372	394	381	377	417	410
R-squared	0.301	0.297	0.453	0.376	0.496	0.300

Notes. These regressions replicate the specification in Table 3's column 2 separated for the subsamples of male and female children. The dependent variable in columns 1 and 2 is an indicator taking the value of 1 if the child's stated willingness to pay for a risky lottery is above 10,000 VND. The outcome in columns 3-6 is the mean across multiple indicator variables taking the value of 1 if the child claims to be following a given protective measure against Covid-19, and 0 otherwise. R1 and R2 stand for the first and second round of Phone Surveys undertaken in 2020, respectively. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.13: Nonlinear Effects of the Proportion of Female Peers on the Contribution to Home Production

	(1) R4 (Age 12) Total Home Hours	(2) R5 (Age 15) Total Home Hours	(3) R4 Over 2 Home Hours	(4) R5 Over 2 Home Hours	(5) R4 Any Home Hours	(6) R5 Any Home Hours
Proportion Female Peers ($> 50\%$)	0.272	0.427**	0.078	0.099**	0.029	0.123**
	(0.189)	(0.197)	(0.056)	(0.047)	(0.058)	(0.051)
Mean (sd) Dep. Var.	1.36 (1.24)	1.64 (1.42)	0.15 (0.36)	0.18 (0.38)	0.78 (0.42)	0.86 (0.35)
Observations	433	433	433	433	433	433
R-squared	0.303	0.352	0.254	0.260	0.334	0.287

Notes. Replication of Table 7 with the replacement of the proportion of female peers by an indicator of having at least 50% of female peers. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.14: Effects of the Proportion of Female Peers on Academic Aspirations

	(1)	(2)
	Aspires to Reach University (YL R5, Age 15)	
Proportion Female Peers	-0.100	0.300**
	(0.303)	(0.133)
Sample	Males	Females
Observations	428	431
R-squared	0.384	0.365

Notes. These regressions replicate the specification in Table 3's column 2 separated for the subsamples of male and female children. The dependent variable is an indicator taking the value of 1 if the child stated that his/her desired level of academic achievement would be at least a college degree. Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.15: Mechanisms for Higher Female Enrollment at Age 19: Perceived Returns to Schooling (measured at YL's R5, Age 15)

	(1)	(2)	(3)	(4)	(5)
Log...	Mean Expected Return University	Mean Expected Return University	Variance Expected Return University	Variance Expected Return University	Currently Enrolled Round Covid (Age 19)
Proportion Female Peers	0.872***	0.307	1.771*	0.974	
	(0.216)	(0.202)	(0.903)	(0.760)	
Mean Expected Returns					0.165*
					(0.056)
Variance Expected Returns					-0.001
					(0.016)
Sample	Females	Males	Females	Males	All
Observations	424	439	425	439	775
R-squared	0.288	0.296	0.251	0.230	0.381

Notes. I obtain the mean and the variance of expected income based on the following three questions measured at round 5: (i) assuming that you complete university, what do you think is the minimum amount you can earn per month at age 25?; (ii) assuming that you complete university, what do you think is the maximum amount you can earn per month at age 25?; (iii) assuming that you complete university, what do you think is the probability that your earnings at age 25 will be at least [midpoint between max and min level]? Under the standard assumption of a perceived triangular distribution of wages, together with the assumption of a triangular density function. The outcome in column 5 is measured in 2020 — the Covid round — and take the value 1 if the child is enrolled or is currently unable to attend school because of Covid but plans to return once it is over, and 0 otherwise. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.16: Mechanisms for Higher Perceived Returns to Schooling: Professional Aspirations (measured at YL's R5, Age 15)

	(1)	(2)	(3)	(4)
	Three Categories Job Masculinity	Three Categories Job Masculinity	Indicator Job Not Female-dominated	Indicator Job Not Female-dominated
Proportion Female Peers ($\geq 50\%$)	0.107	-0.204	0.345**	0.014
	(0.174)	(0.141)	(0.174)	(0.230)
Sample	Females	Males	Females	Males
Observations	363	369	363	369

Notes. Information for this exercise comes from round 5 of YL's question on "when you are about 25 years old, what job would you like to be doing?". Columns 1 and 2 estimate ordered probit models where the outcome variable has three categories, indicating traditionally female jobs (coded as 1), gender neutral jobs (coded 2), and traditionally male jobs (coded 3). Columns 3 and 4 use instead a binary variable as outcome, which takes value one if the job was initially coded 2 or 3, and zero otherwise. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.17: Effects of the Proportion of Female Peers on Self-confidence in Abilities

	(1)	(2)
	Confidence in Self (YL R5, Age 15)	
Proportion Female Peers	0.177	0.405***
	(0.122)	(0.103)
Sample	Males	Females
Observations	433	431
R-squared	0.314	0.290

Notes. These regressions replicate the specification in Table 3's column 7 for a measure of self-confidence in own ability to deal with challenging situations (the exact variables used are described in Section D). Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.18: Effects of the Proportion of Female Peers on Fertility Preferences (YL R5, Age 15)

Panel A: Males				
	(1)	(2)	(3)	(4)
	Age First Child	Total Number of Children	Number of Sons	Number of Daughters
Proportion Female Peers	0.616	-0.861**	-0.337*	-0.321**
	(1.437)	(0.367)	(0.203)	(0.141)
Mean (sd) Dep. Var.	27.28 (3.02)	1.99 (0.39)	1.03 (0.25)	0.97 (0.28)
Observations	435	446	446	446
R-squared	0.328	0.235	0.162	0.257

Panel B: Females				
	(5)	(6)	(7)	(8)
	Age First Child	Total Number of Children	Number of Sons	Number of Daughters
Proportion Female Peers	1.556	0.024	0.136	0.022
	(1.394)	(0.262)	(0.138)	(0.188)
Mean (sd) Dep. Var.	26.72 (2.96)	1.95 (0.40)	0.97 (0.25)	1.00 (0.29)
Observations	418	440	439	440
R-squared	0.386	0.202	0.198	0.147

Notes. All regressions replicate the specification in Table 3's column 2, separately for male respondents (Panel A) and female respondents (Panel B). The outcomes are self-reported preferences for ideal quantities stated in the fifth round of the YL survey. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.19: Heterogeneous Effects of the Proportion of Female Peers on Long-term Views on Gender Norms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean Agreement with Traditional Gender Norms (YL R5, Age 15)						
Proportion Female Peers	-0.694***	-0.297**	-0.267**	-0.232*	-0.289**	-0.306**	-0.250**
	(0.254)	(0.124)	(0.130)	(0.128)	(0.125)	(0.140)	(0.122)
Mother is Decision Maker	-0.304*						
	(0.173)						
Mother is Decision Maker*Proportion Female Peers	0.643*						
	(0.365)						
Any Parent Illiterate*Proportion Female Peers		0.535					
		(0.420)					
Ethnic Minority*Proportion Female Peers			-0.293				
			(0.343)				
High Wealth Index*Proportion Female Peers				-0.080			
				(0.219)			
Not Better Student than Other*Proportion Female Peers					-0.042		
					(0.100)		
Low School Effort*Proportion Female Peers						-0.160	
						(0.097)	
High Interest School*Proportion Female Peers							0.024
							(0.091)
Sample	Males	All	All	All	All	All	All
Observations	445	886	886	886	855	850	857
R-squared	0.274	0.254	0.249	0.249	0.254	0.261	0.256

Notes. All regressions replicate the specification in Table 3's column 2 (although column 1 restricts the sample to male children). A description of the interacting variables used is provided in Section D. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.20: Supporting Evidence: Effects of Exposure to Females on Perceived Popularity

	(1)	(2)	(3)	(4)
	Own Popularity	Own Popularity	Own Popularity	Sibling's Own Popularity
	R4 (Age 12)	R4	R4	R4
Proportion Female Peers	0.383***	0.425**	0.417**	0.027
	(0.127)	(0.194)	(0.189)	(0.125)
Outcome Reported by	YL Child	Male YL Child	Female YL Child	YL's Sibling
Observations	879	433	430	393
R-squared	0.172	0.248	0.301	0.280

Notes. Information for this exercise comes from round 4 of YL. The outcomes are the simple mean across seven questions eliciting self popularity among peers (more details in Section D). Columns 1-3 use as outcome self-perceived measures of own popularity for the YL child, while column 4 uses as outcome YL's sibling's self-perceived own popularity. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.21: Non-selective Attrition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Matched with YL's Round 5 (Age 15)						
Female	-0.001						-0.002
	(0.008)						(0.007)
Ethnic Minority		0.004					-0.001
		(0.009)					(0.012)
Home Educational Resources			-0.018				-0.013
			(0.016)				(0.017)
Literate Mother				-0.026			-0.026
				(0.021)			(0.024)
Low School Effort					-0.001		-0.002
					(0.003)		(0.003)
Wealth Index						-0.042	-0.035
						(0.037)	(0.040)
In Estimating Sample							0.000
							(0.003)
P-value F-test Joint Significance							0.631
Observations	1,129	1,132	1,138	1,128	1,090	1,138	1,069
R-squared	0.183	0.183	0.184	0.184	0.183	0.185	0.186

Notes. Regressions of an indicator taking the value 1 if the YL child surveyed in the 2011 School Survey was also present ("matched") with the attitudinal information from YL's round 5 (collected in 2016) and 0 otherwise on the baseline characteristics indicated in the rows and measured in the first round of the school survey and school fixed effects. I do not report a separate regression for "In Estimating Sample" as only 1 out of the 17 children who are not matched was part of my estimating sample. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.22: Is There Selection into Subsequent Academic Peer Groups?

	(1) Proportion Female Peers 2017 (Age 16)	(2) Ever Transfers Schools
Proportion Female Peers 2011	0.174	-0.032
	(0.263)	(0.050)
Observations	221	884
R-squared	0.547	0.166

Notes. Replication of the specification in Table 3's column 7 for: (1) the proportion of female classmates in Grade 10, as measured in the 2017 School Survey (hence the observations are restricted to YL children who participated both in the 2011 and in the 2017 school surveys); (2) an indicator taking the value 1 if the child switches schools between Grade 6 and Grade 9 as retrieved from round 5 of the YL survey. Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.23: Robustness to Restricting the Sample to Children Enrolled in Large Schools

	(1)	(2)
	Mean Agreement with Traditional Gender Norms (YL R5, Age 15)	
Proportion Female Peers	-0.486***	-0.410**
	(0.122)	(0.189)
Female	-0.201***	-0.170***
	(0.027)	(0.045)
Observations	535	191
R-squared	0.249	0.308

Notes. Replication of the specification in Table 3's column 5 restricting the sample to children in schools with at least 3 classes (column 1) or at least 4 classes (column 2). Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.24: Are Parental Investments Affected by Peers' Female Composition? (YL R4, Age 12)

	(1)	(2)	(3)	(4)	(5)
	Any Extra Class	Hours Extra Classes	Money Extra Classes	Ideal Education Level Reached	Knows Names of Child's Friends
Proportion Female Peers	0.068	0.575	-95.251	0.114	0.006
	(0.135)	(1.978)	(342.070)	(0.353)	(0.143)
Observations	882	881	876	863	884
R-squared	0.376	0.359	0.220	0.178	0.169

Notes. Replication of the specification in Table 3's column 2 for parental investment information collected in round 4 of the YL survey. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.25: Quantile Regressions of the Effects of Peers' Female Composition on Views on Gender Norms

	(1)	(2)	(3)	(4)	(5)
	Mean Agreement with Traditional Gender Norms (Age 15)				
	Q. 0.2	Q. 0.4	Q. 0.5	Q. 0.6	Q. 0.8
Proportion Female Peers	-0.216	-0.356**	-0.394*	-0.464**	-0.018
	(0.247)	(0.177)	(0.215)	(0.202)	(0.266)
Observations	446	446	446	446	446
R-squared	0.134	0.204	0.224	0.209	0.161

Notes. Quantile regressions for the male subsample using the same controls as in Table 3's column 2. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.26: Robustness of Effects on Gender Norms to Standardizing the Outcome

	(1)	(2)
	Standardized Norms (Age 15)	
Proportion Female Peers	-0.856**	-1.524***
	(0.403)	(0.324)
Peers Correction	No	Yes
Observations	886	880
R-squared	0.252	0.292

Notes. Replication of the specification in Table 3's columns 2 and 6 where the outcome variable is constructed first standardizing each of its components, then obtaining their mean, and standardizing it again. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.27: Robustness of Effects on Gender Norms to the Exclusion of Schools with Potential non-Exogenous Assignment

	(1)	(2)
	Mean Agreement with Traditional Gender Norms (Age 15)	
Proportion Female Peers	-0.510***	-0.481***
	(0.085)	(0.088)
Observations	629	562
R-squared	0.266	0.271

Notes. Replication of the specification in Table 3's column 7 with the exclusion of students from the schools where the p-value of Fisher's Exact test is below 0.05 and 0.1 (columns 1 and 2, respectively). Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.28: Placebo Outcomes in Round 4 (Age 12)

	(1)	(2)	(3)	(4)	(5)	(6)
	Child's Height (cm)	Normal BMI	Self-Reported Good/Very Good Health	School/Work Days Missed due to Illness	Wealth Index	Frequent Internet User
Proportion Female Peers	-2.241	0.067	0.016	-0.059	0.020	-0.128
	(2.905)	(0.085)	(0.255)	(0.174)	(0.033)	(0.125)
Mean (sd) Dep. Var.	159.24 (7.87)	0.96 (0.19)	0.32 (0.47)	0.37 (0.53)	0.63 (0.11)	0.82 (0.39)
Observations	880	880	880	870	879	880
R-squared	0.334	0.170	0.144	0.163	0.487	0.245

Notes. Replication of column 7 in Table 3 when using as outcomes dimensions measured in round 4 of the YL survey that are not expected to change due to female peers. Normal BMI is defined based on the z-score on the BMI-for-age as consistent with the medical literature (Tuan et al., 2008). Column 3's outcome is an indicator taking the value 1 if self-reported health is good or very good. Column 4's outcome is a categorical variable taking the value of 0 if no days were missed, 1 if 1-5 days were missed, and 2 if more than 5 were missed. "Frequent Internet User" takes the value 1 if the child reports to have used internet many times in his/her life, and 0 otherwise. Standard errors clustered at the class level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.29: Robustness Check: Effects on Sibling's Outcomes

	(1) R4 (Age 12) Total Home Hours	(2) R4 Any Home Hours	(3) R5 (Age 15) Total Home Hours	(4) R5 Any Home Hours	(5) R4 PPVT Raw Score
Proportion Female Peers	0.646	0.064	0.505	0.026	-1.806
	(1.270)	(0.290)	(0.709)	(0.219)	(4.997)
Mean (sd) Dep. Var.	1.34 (1.99)	0.52 (0.50)	1.23 (1.65)	0.56 (0.50)	56.78 (10.93)
Observations	443	443	379	379	488
R-squared	0.422	0.485	0.502	0.421	0.577

Notes. All outcomes are obtained for one sibling per child belonging to the YL's longitudinal survey. Cognition is measured through the Peabody Picture Vocabulary Test (PPVT), which aims at quantifying receptive vocabulary ability. The proportion of female peers is the one corresponding to the main YL child as measured from the School Survey. The specifications are similar to those in Table 3's column 7 with the inclusion of a polynomial in sibling's age. Age of the siblings in columns 1-4 is restricted to be between 7 and 18. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.30: Robustness of Effects on Gender Norms to Only Using Schools with Low Gaps in the Proportion of Females Across Classes

	(1)	(2)	(3)
	Mean Agreement with Traditional Gender Norms (Age 15)		
Proportion Female Peers	-0.564*	-0.512*	-0.513***
	(0.299)	(0.284)	(0.171)
Sample	< 0.25	0.1 $<$difference < 0.25	$0.35 <$ level < 0.65
Observations	600	395	738
R-squared	0.302	0.307	0.311

Notes. Columns 1 and 2 replicate the specification in Table 3's column 7 when restricting the sample to schools where the difference between the highest and the lowest proportion of female peers is "low" (as indicated in the table). Column 3 restricts the sample instead to classes where the female composition is between 35 and 65%. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.31: Robustness Check: Accounting for Extreme Responses

	(1) Log Norms	(2) Winsorized Norms
Proportion Female Peers	-0.235***	-0.464***
	(0.054)	(0.102)
Observations	880	880
R-squared	0.282	0.291

Notes. Replication of column 7 in Table 3 when using as outcomes the logarithm of my measure of traditionalism (column 1) and when winsorizing the outcome for the top and bottom 1 percent of the distribution. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.32: Robustness of Effects on Gender Norms to the Inclusion of Higher-level Fixed Effects

	(1)	(2)
	Mean Agreement with Traditional Gender Norms (Age 15)	
Proportion Female Peers	-0.499***	-0.493***
	(0.103)	(0.104)
Aggregate Fixed Effects	Region	Community
Observations	880	859
R-squared	0.296	0.317

Notes. Replication of the specification in Table 3's column 7 with the inclusion of more aggregate-level fixed effects. Column 1 additionally controls for whether the location is rural or urban. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

B Additional Figures (For Online Publication)

Figure B.1: Distribution of the Proportion of Female (Class-level)

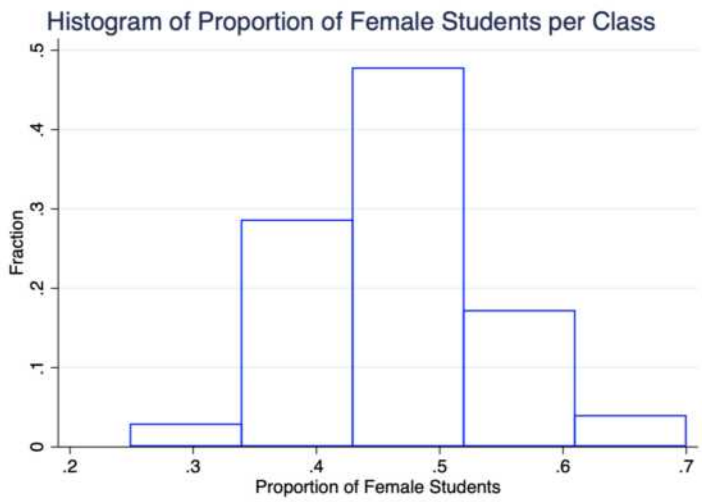
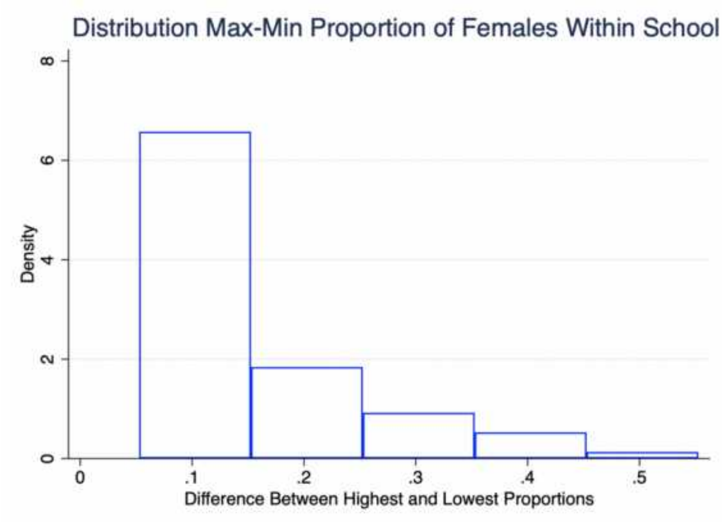


Figure B.2: Distribution of the Within-School Variation in the Proportion of Female Peers



[76]The horizontal axis labels are as follows: 1: strongly disagree; 2: disagree; 3: agree; 4: strongly agree. 95% confidence intervals are also reported.

[77]Ethiopia, one of the four countries in which YL undertook their studies, has an abnormally low level of traditionalism in the fifth wave of the WVS. The values in wave 7 are 16% and 35% for reaching university and business leaders, respectively. The question on women executives was not asked in Hong Kong.

Figure B.3: Distribution Views on the Different Dimensions of Traditionalism by Gender[76]

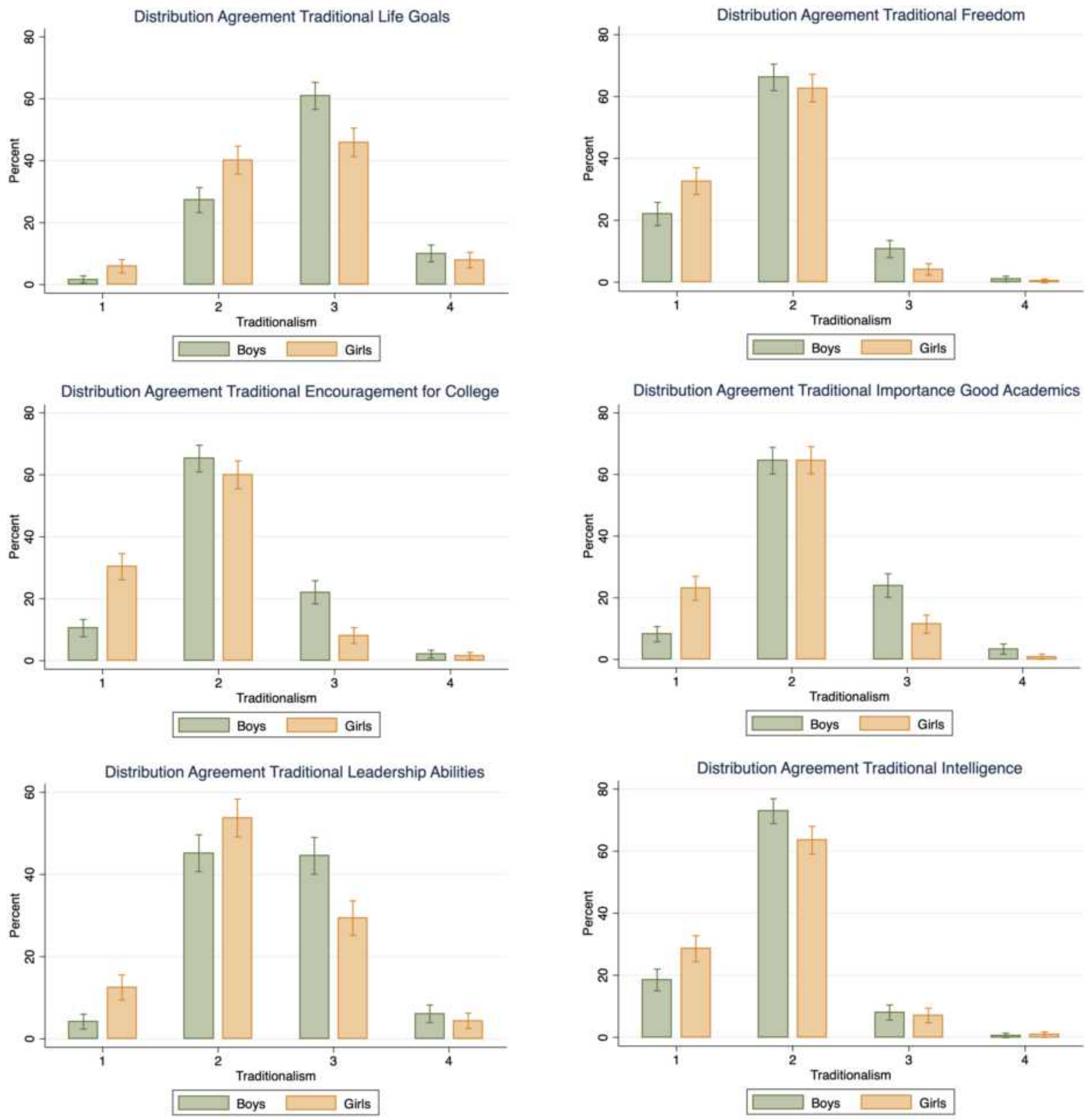


Figure B.4: Distribution Views on the Different Dimensions of Traditionalism by Gender (Continued)

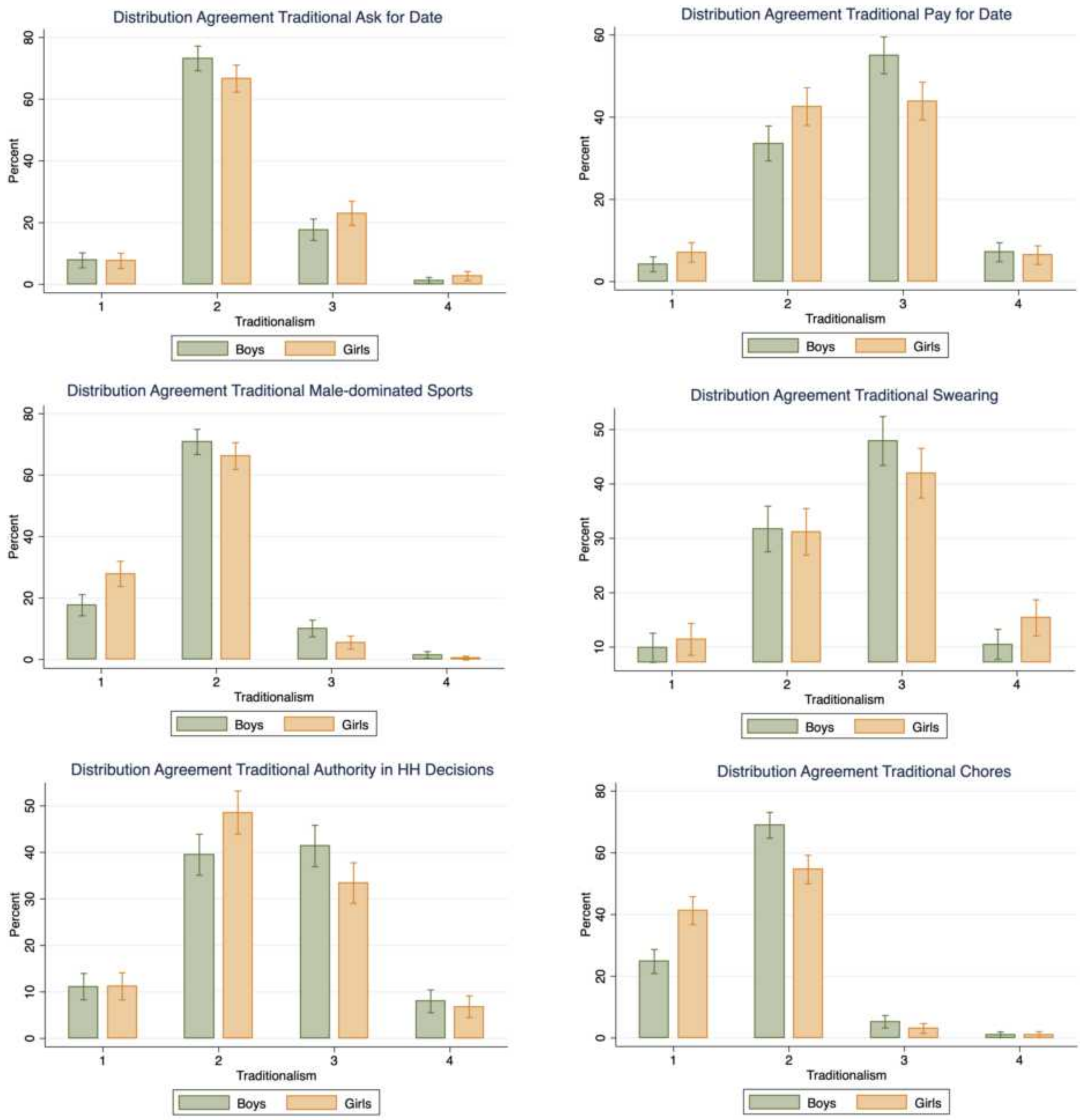


Figure B.5: Distribution of Traditional Gender Views Across Countries (Source: World Values Survey, Wave 5)[77]

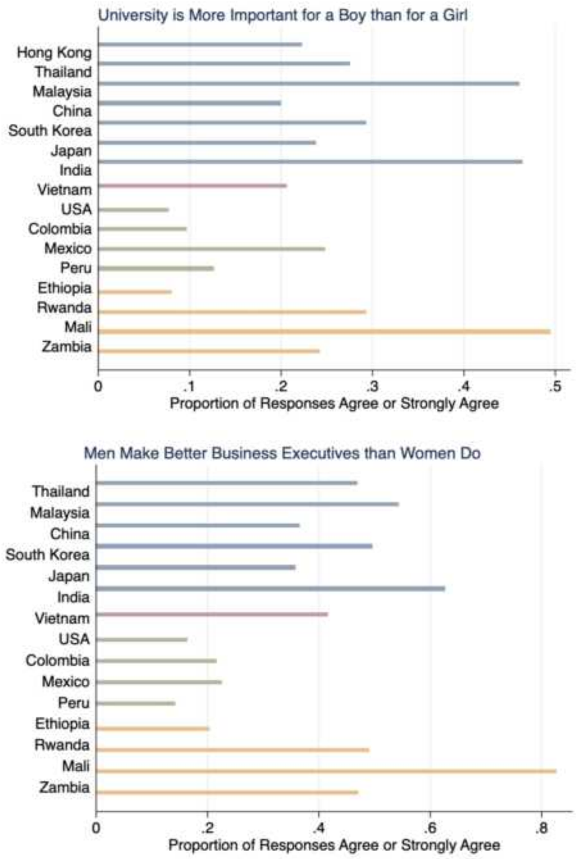


Figure B.6: Distribution of the Difference Between Actual and Simulated Proportion of Female Peers for One Randomly-selected Class per School

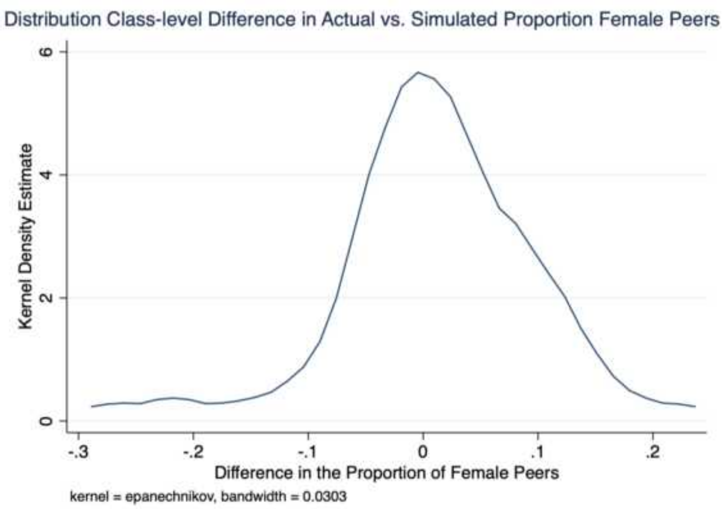
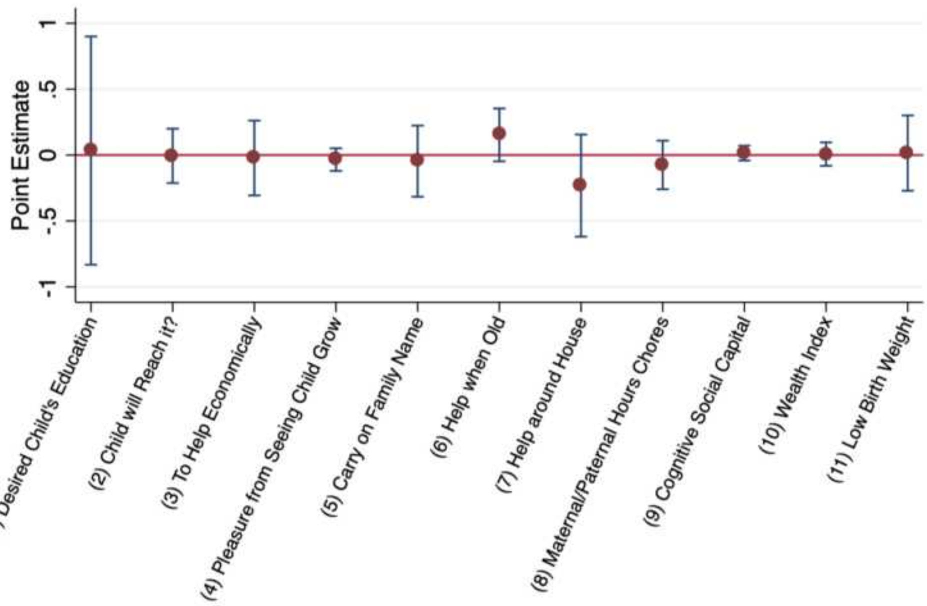
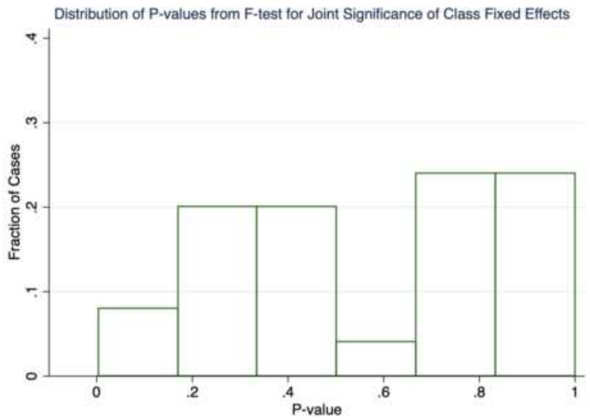


Figure B.7: Balance of Pre-determined Covariates: (i) Parental Expectations and Reasons for Having Children (Age 5); (ii) YL Basic Characteristics in Round 1 (Age 1)

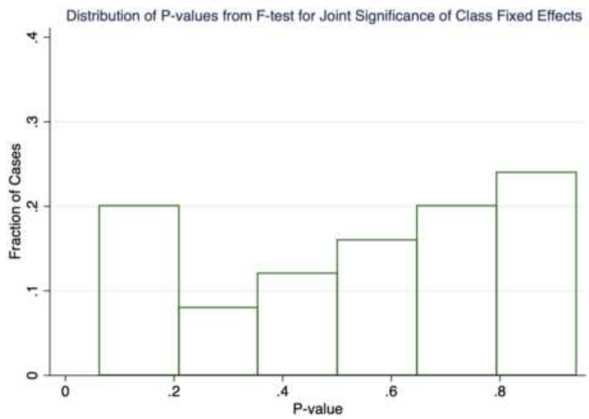


Notes. Regressions replicate those from Table 2 with the inclusion of a gender indicator as control (the qualitative and quantitative results are virtually identical if the gender indicator is not included). 1 uses as outcome the years of education that the parents wish their child to achieve (I code college education as 14). 2's outcome takes the value 1 if the parents believe that the child will actually achieve the desired level of education and zero otherwise. 3-7 are dummies taking the value of 1 if the parents stated that they think that the reason for having a child proposed in each column is "important" or "very important" and 0 otherwise — traditional Confucian-based values dictate that women should have son in order to carry on the parental family name (Le, 2008). 8 is the ratio of maternal contribution to home production (measured in yearly hours) over the sum of hours contributed by both spouses. 9 is an index constructed by YL's based on caregiver's responses on whether (s)he feels part of the community, whether (s)he feels people in general can be trusted, whether people generally get along with each other okay, and whether people would try and take advantage of them if they could. 11 is an indicator taking the value of 1 if the child was less than 2,900 grams at birth. The sample includes all YL children who were also surveyed in the School Survey and whose class-assignment was defined as "random". Regressions control for school fixed effects. 90% confidence intervals are reported.

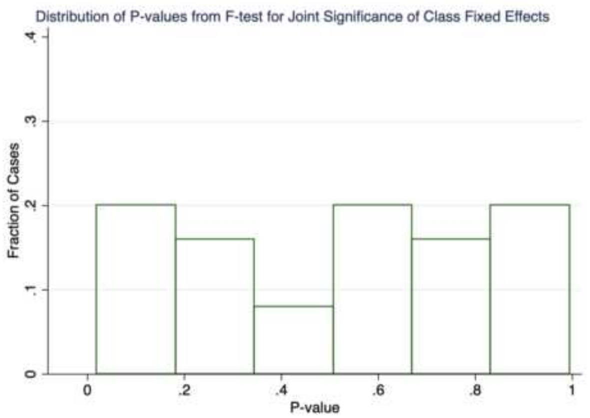
Figure B.8: Distribution P-value Joint Significance F-Tests for Class Fixed Effects: Parental Expectations Measured at Age 5 (I)



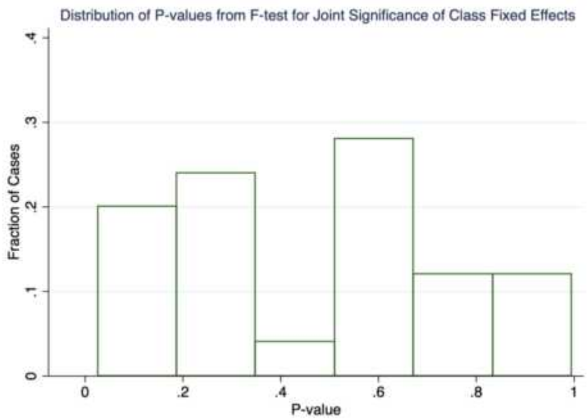
(a) Age Start Earning Money



(b) Age Leave Education

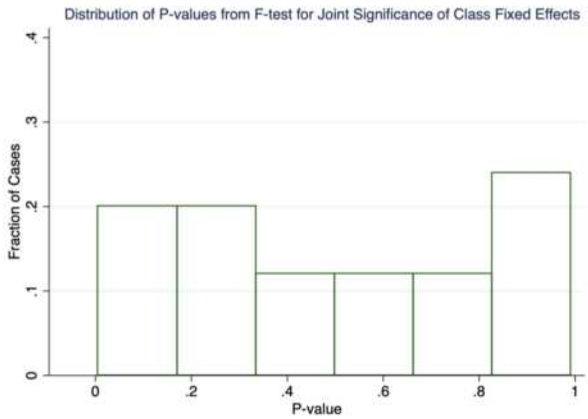


(c) Age Financially Independent

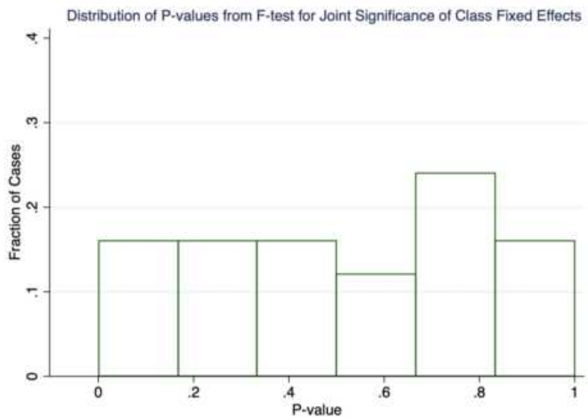


(d) Age Form Own Household

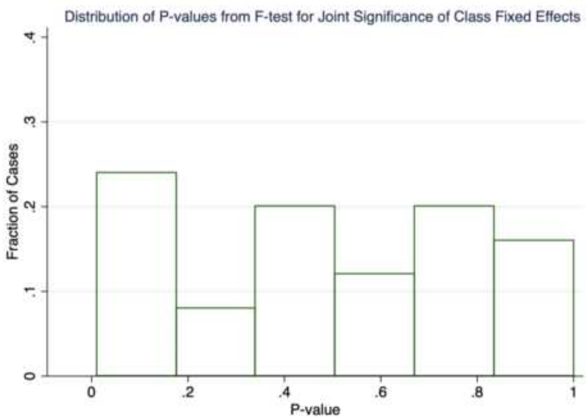
Figure B.9: Distribution P-value Joint Significance F-Tests for Class Fixed Effects: Parental Expectations Measured at Age 5 (II)



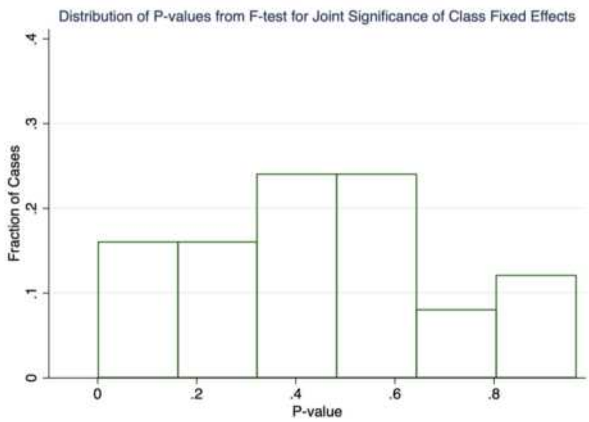
(a) Age Marriage



(b) Age Have Children

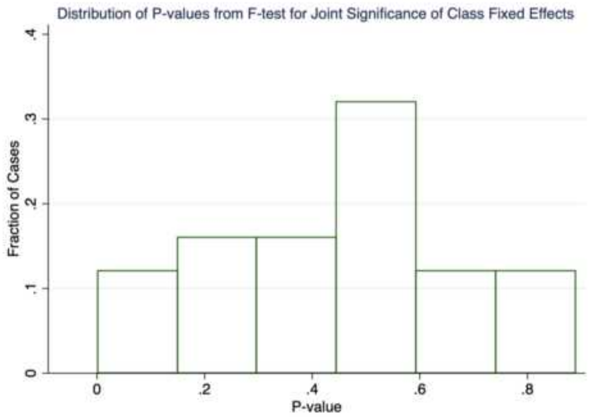


(c) Should Live Close

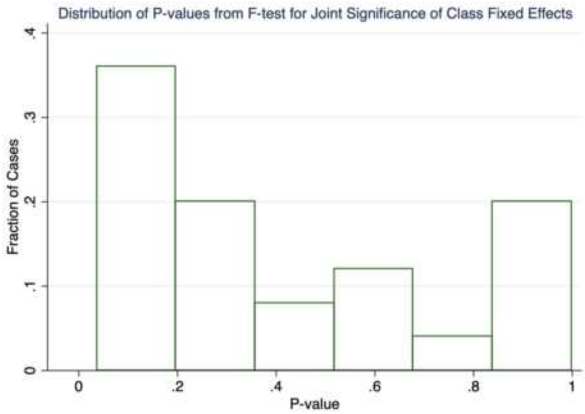


(d) Should Help Siblings Financially

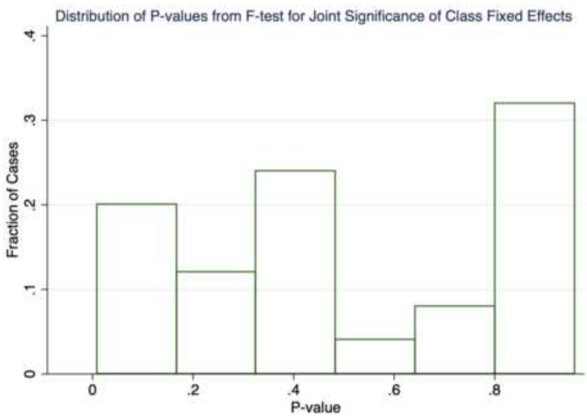
Figure B.10: Distribution P-value Joint Significance F-Tests for Class Fixed Effects: Parental Expectations Measured at Age 5 (III)



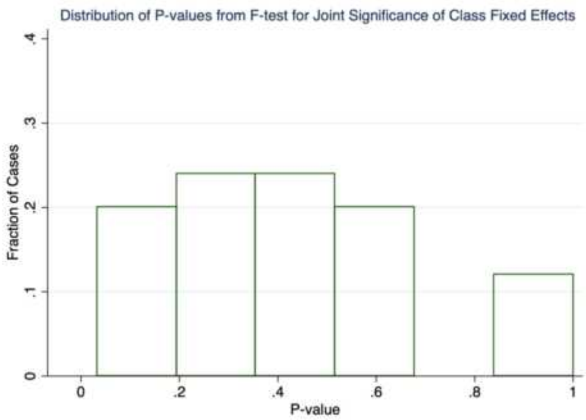
(a) Should Help Housework



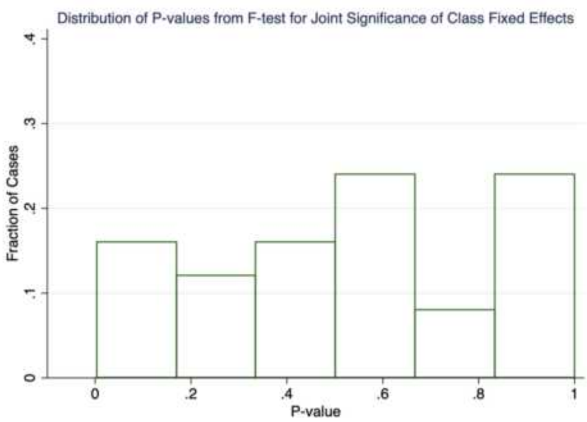
(b) Should Help Parents Financially



(c) Should Take Care of Siblings

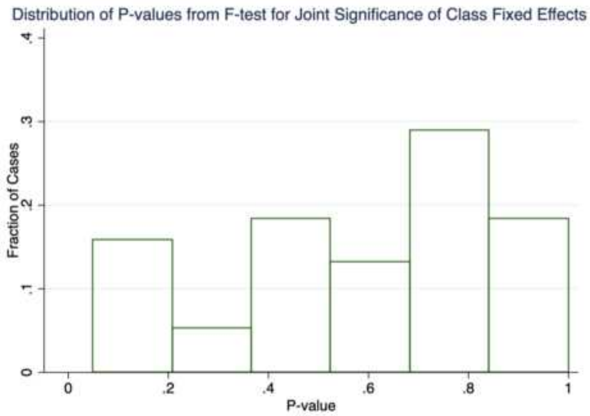


(d) Should Take Care of Parents When Old

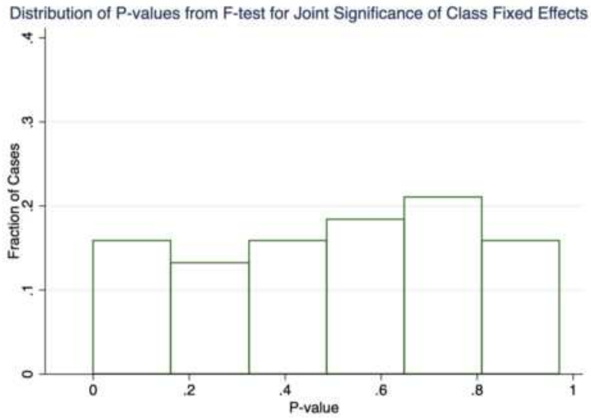


(e) Should Provide Emotional Support to Parents

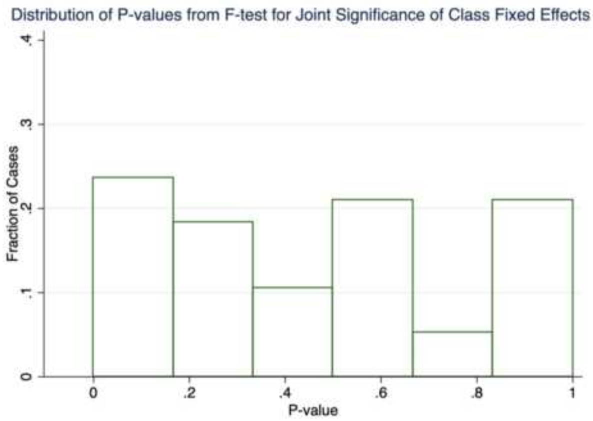
Figure B.11: Distribution P-value Joint Significance F-Tests for Class Fixed Effects (Pre-determined Characteristics, Measured at Age 5)



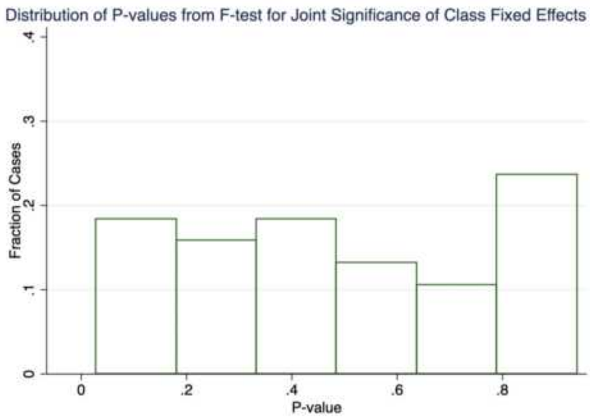
(a) Females



(b) Parental Education

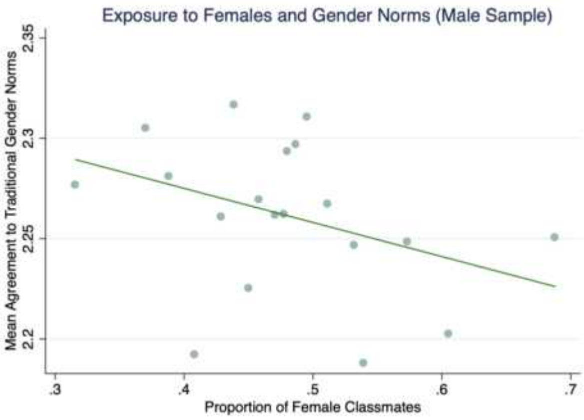


(c) Wealth

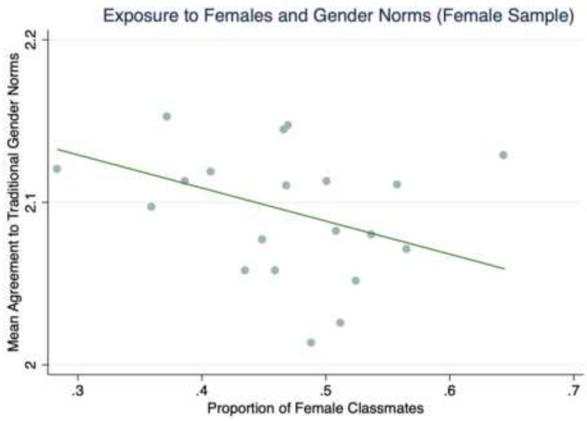


(d) Age

Figure B.12: Correlations of Proportion of Female Peers and Long-term Views on Gender Norms by Student Gender



(a) Males



(b) Females

Figure B.13: Estimates from All Combinations of Main Controls

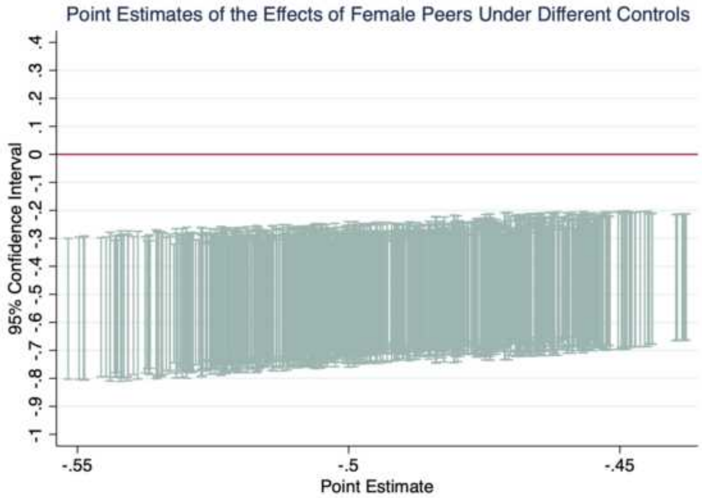
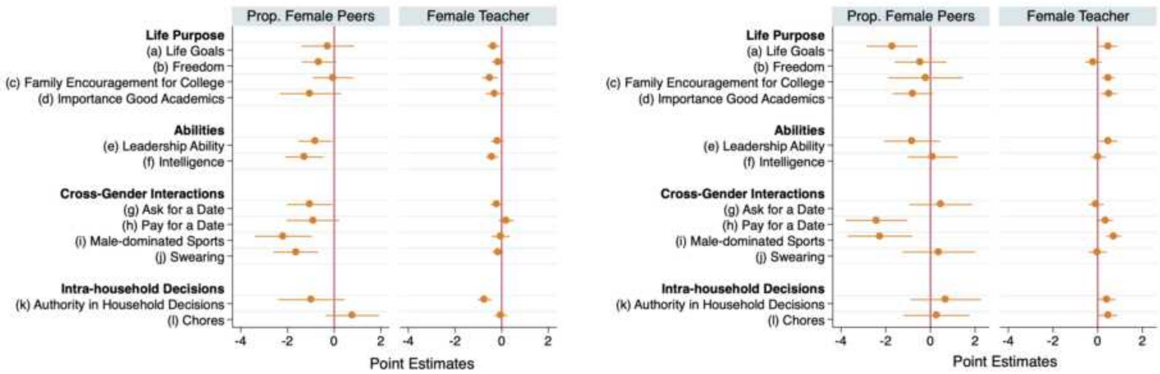


Figure B.14: Effects by Gender of the Proportion of Female Peers and Exposure to Female Teacher On Views on Gender Norms



(a) Males

(b) Females

Notes. Each row collects two coefficients from the same regression: the one on the proportion of female peers and the one for an indicator of the main teacher being a female.

Figure B.15: Allowing for Different Slopes: Predicted Values from Model with Linear Splines

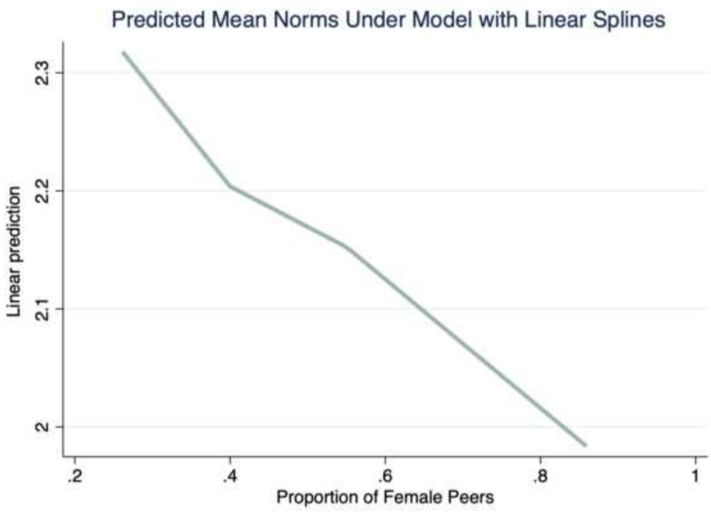
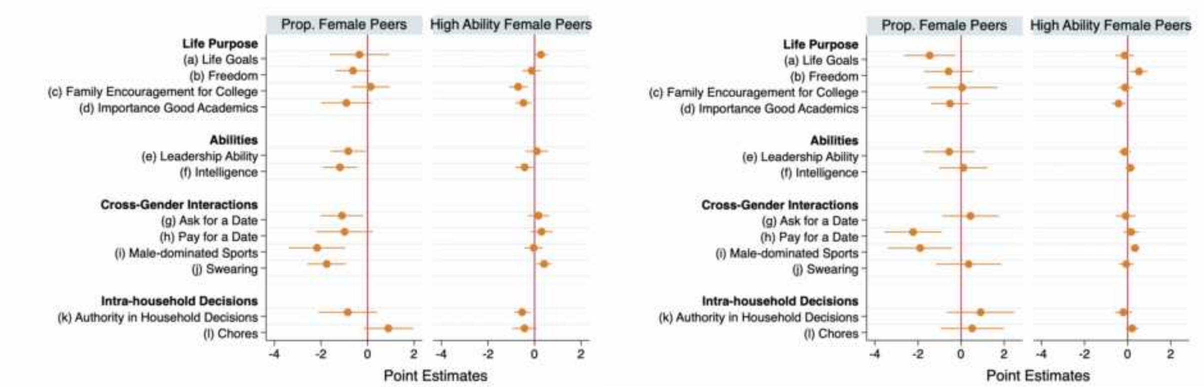


Figure B.16: Effects by Gender of the Proportion of Female Peers and Exposure to High-Ability Female Peers On Views on Gender Norms



(a) Males

(b) Females

Figure B.17: Distribution of Placebo-Generated Estimates of the Proportion of Female Peers on Gender Norms (empirical 90% CI indicated by green lines) When Reshuffling Attitudes Within Schools

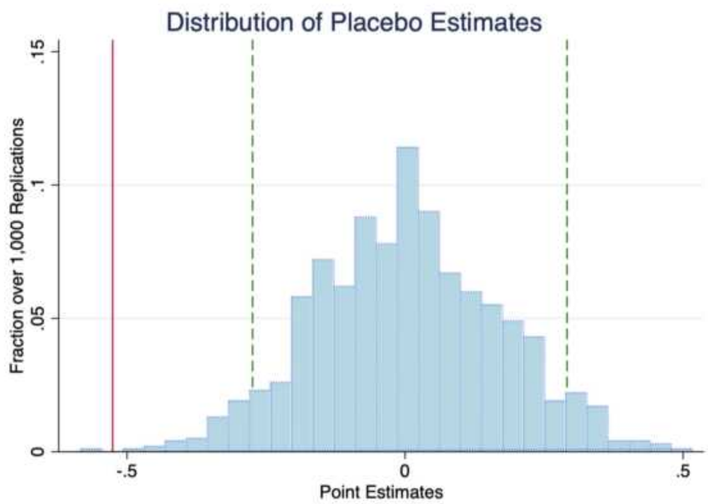
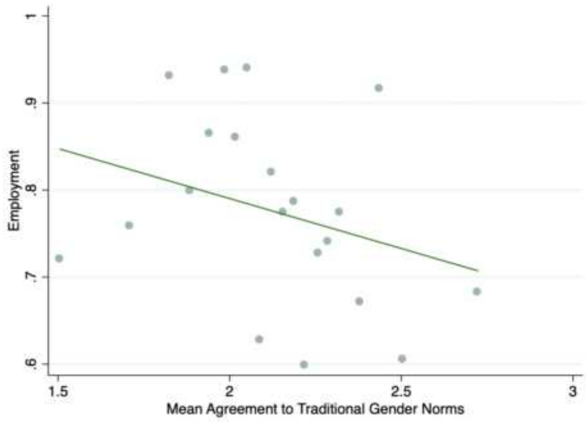
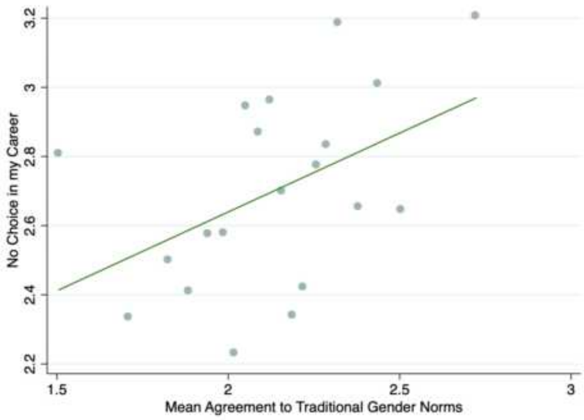


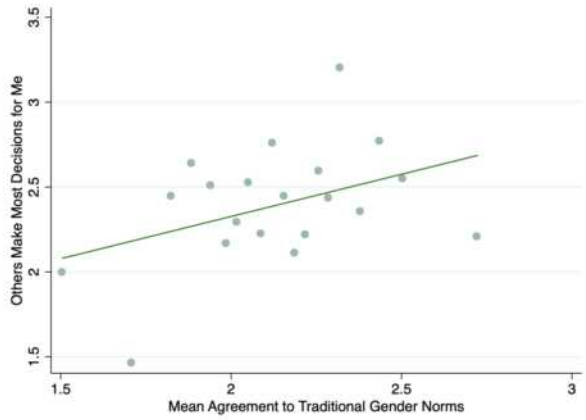
Figure B.18: Relationship between Gender Traditionalism and other Outcomes and Attitudes (Old Cohort Females)



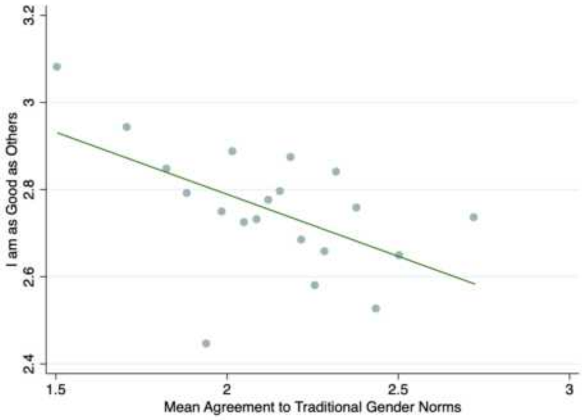
(a) Labor Force Participation



(b) No Other Career Choice Available

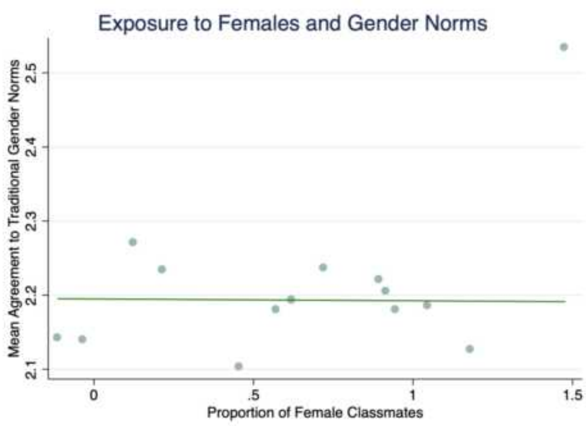


(c) Others Make Most Decisions for Me

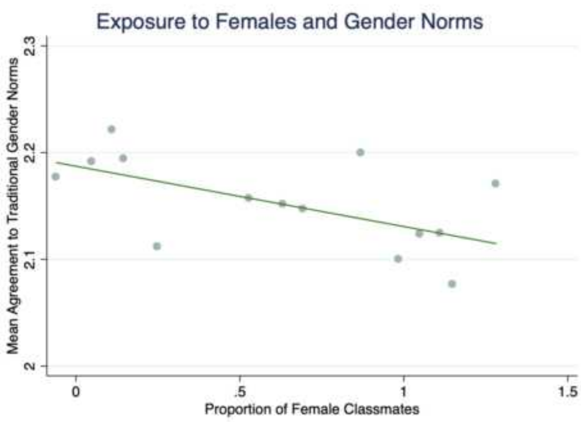


(d) I am as Good as Others

Figure B.19: Correlations of Proportion of Female Peers and Long-term Views on Gender Norms by Student Gender: Ethiopia



(a) Males



(b) Females

C Extended Analyses (For Online Publication)

C.1 Representativity of the Sample

As mentioned, the Young Lives survey was designed with a pro-poor approach. Therefore, while it is well-suited to study the dynamics of the children's living conditions and development process, the final dataset is slightly poorer than nationally representative ones (Young Lives, 2014). It is representative, however, at the regional level for households with children aged 1 or 8, the targeted population (Nguyen, 2008). Indeed, throughout the main text I have highlighted that those key outcomes of interest for my study that can be contrasted with nationally representative studies (e.g. attitudes to several gender norms and fertility preferences, in Figure B.5 and footnote 100, respectively), yield very close figures.[78]

In this section I show that: (i) the subsample of children that participate in the school survey is not a selected one within the Young Lives data; (ii) the YL children in the school survey are not different from their non-YL-participating classmates.

For the first point, I use the full YL sample available at the second wave[79] and regress a battery of observable characteristics on an indicator of whether the child was also included in the school survey. Table A.1 shows that the majority of the estimates are small in magnitude and not statistically significant. This suggests that those YL children that were selected for the school survey were indeed a random subsample of the original sample. As expected, those children included in the school survey are slightly more likely to be enrolled in school (4%), since this was a requirement to be eligible to be part of the school survey in the first place.

For the second point, I turn to the school survey and I focus, as in the main analysis, on exogenously formed classes, but I do not impose that the child needs to be part of the YL main survey. Instead, the interest here is on keeping also the classmates and exploring whether an indicator of being a YL child is statistically significant. This would indicate that YL children are systematically different from their peers. Table A.2 shows that this is not the case.[80]

C.2 Additional Details on Available Survey Information

Sociodemographics (SS). Standard questions on the gender, age, parental education, ethnicity, and household size were asked. Moreover, an interesting piece of information pertains to the households' ownership of various relevant assets, which allows me to compute a wealth index[81]. This is useful because wealth has been shown to be important for addressing omitted

[78]Another example is the gender wage gap, which stands at slightly above 20% both in the Young Lives sample and in nationally representative ones such as the Vietnam Household Living Standards (e.g. Liu, 2004; Oanh and Ngoc, 2020).

[79]The original cross-sectional dimension (i.e. at the first round) of this cohort was 2,000.

[80]See Section D for more details on the variables used.

[81]I compute such an index as the proportion of positive answers to the questions of whether the household owns each of the following eleven items: phone; television; electric fan; computer; mobile phone; bicycle; air conditioning; internet; radio; motorbike;

variable biases in academic contexts (Lovenheim and Reynolds, 2013) and for predicting intergenerational transmission of outcomes (Blundell and Risa, 2019). Relevant to our context, there is also information on the number of books at home as well as on physical resources available at home, which are likely an input in academic production (owning: a calculator; a study desk; a study chair, and a own space at home to study). I construct another index, this time for academic resources at home, as the proportion of affirmative responses to such questions.

Cognitive Information (SS). In order to evaluate the students' academic progress during one school year, YL designed, distributed, and supervised the undertaking of mathematics and Vietnamese language tests at each of the two visits (i.e. at the very beginning of the school year and right before its end). These tests consisted of 30 multiple-answer questions aiming at testing country-wide official Grade 5 curriculum knowledge and containing a subset of items common to both rounds. Two attractive features are: (a) given the multiple-choice nature of the tests and their implementation by YL fieldworkers, the scores do not suffer from differential grading neither within nor across schools, which maximizes their comparability, and (b) since the second round is conducted right before the end of the school year, differential knowledge losses over the holidays are avoided (Fruehwirth and Gagete-Miranda, 2019).[82] I work both with the total raw score and, to ease interpretation, also with their standardized version (mean of 500 and a standard deviation of 100), in order to explore whether class composition affected academic performance in the short- and the long-run.

Peers' Information and Friendship Nominations (SS). The School Survey collects rich information on the subset of randomly-selected peers — I provide evidence on the successful randomization in Table A.2.[83] This allows me to compute leave-out-means for the various dimensions of interest *at the class level*, which is a finer level of aggregation than what it is possible in the absence of random assignment. Importantly, the combination of a survey designed ex profeso to study potential channels of academic spillovers (e.g. non-cognitive aspects) sets this study apart from the rather limited information typically available on peers' characteristics, which is usually limited to those readily available in administrative data (such as gender and age).

Moreover, a unique feature of the school survey is that, at the end of the academic year, every child was asked to answer the following two questions with respect to *each and every* surveyed classmate[84]: (1) how would you describe your friendship with this classmate?, and

car.

[82] Importantly, on the date of the second round (an unannounced one), only 2% of children did not attend school.

[83] More details on the process followed and its success are provided in James (2013).

[84] This is in contrast, for instance, with the AddHealth dataset, where children are asked to nominate up to five male and five female friends within the school.

(2) how much do you do things with this classmate outside of school?[85] I use this information to explore friendship formation and cross-gender interactions.

Teacher Characteristics (SS). The school survey also collected detailed background characteristics of the teachers. These included, among others, gender, ethnicity, highest educational level, years of experience, and wealth. While this information is useful in reinforcing the evidence that exogenous allocation of students also extended to their assigned teachers, it is unfortunate that the survey did not collect longitudinal information on teacher practices that could have provided direct evidence on whether teacher behavior is adaptive to class composition.

Non-cognitive Information (SS). One strength of the school survey is that it collected non-cognitive information for all students including, among other, self-reported level of effort and of perceived performance at school. This is important for the study of the mechanisms in place. Indeed, although I do not have self-reported behavioral information from the teachers, I do observe students' perceptions on the treatment by their teachers both at the beginning and at the end of the year. This is relevant for the study of peer effects since students are likely to be affected by and act upon such *perceptions*, regardless of actual behavioral changes on the part of the teachers.

Sociodemographics and Cognition (longitudinal YL). The longitudinal tracking of the two YL cohorts offers a rich depiction of the dynamics of the livelihoods of the children. Moreover, language and mathematics tests were consistently implemented in every YL round. I take advantage of this in two main ways. First, it allows me to provide supportive evidence in favor both of the random allocation of students to classes and of the non-selected choice of the subset of YL children that were sampled for the school survey. Second, it provides valuable information to explore the mechanisms behind my main effects as well as alternative outcomes of interest. For instance, I am able to investigate whether dropout rates systematically differ across individuals who faced different degrees of exposure to female peers, and whether cognitive effects are persistent or fade away over time.

Views on Gender Norms. In the 2016 round of YL, when the young cohort was aged 15, asked the following battery of questions, to which the students had to answer in a four-point scale (strongly disagree, disagree, agree, strongly agree) and that I classify into four categories according to the broad gender-related aspect that was elicited. **(1) "Life purpose"** covers the following statements: (a) girls should be more concerned with becoming good wives and mothers than desiring a professional or business career; (b) girls should have the same freedoms as boys; (c) more encouragement in a family should be given to sons than to daughters to go to

[85] The options available for question (1) were: not close friends; a little/sometimes friends; close friends; very close friends. For (2) they were: none; not very much; quite a lot; a lot.

college; (d) it is more important for boys than for girls to do well in school. **(2) "Abilities"** contains: (e) boys are better leaders than girls; (f) on the average, girls are as smart as boys. Note that these dimensions, which are particularly interesting, do not strictly lie within the broad category of gender roles, so I will refer to them separately when relevant. **(3) "Cross-gender interactions"** features: (g) it is all right for a girl to ask a boy out on a date; (h) on a date, the boy should be expected to pay all expenses; (i) it is all right for a girl to want to play rough sports like football; (j) swearing is worse for a girl than for a boy. Finally, **(4) "intra-household decisions"** addresses: (k) in general, the father should have greater authority than the mother in making family decisions; (l) if both husband and wife have jobs, the husband should do a share of the housework such as washing dishes and doing the laundry.

C.3 Further Evidence on Exogenous Class Assignment

C.3.1 Permutation Tests

In order to verify that the average characteristics at the class level are in line with those obtained from randomizing the existing students in a school across the available sections, I proceed as follows.

First, I form synthetic classes of the same size as in the actual data by randomly allocating actual students in a school across sections. I then compute the average characteristics at the class level for a series of important dimensions such as gender, parental education, the wealth index, and non-cognitive skills (e.g. interest in school). I repeat this exercise 1,000 times. For each class and simulation, I compute the distance between the actual value of each characteristic in the original class configuration and its simulated counterpart. Calculating the average distance between the original and simulated values across the 1,000 repetitions yields absolute values of about 0.0001 for all the observables considered, with very small standard deviations. For example, for gender, the mean distance is -0.0001 and the standard deviation is 0.001, while for the students' degree of interest in school, a plausibly important dimension along which endogenous selection across classes could take place, the figures for the mean and the standard deviation are -0.001 and 0.007, respectively. This analysis therefore provides further support to the exogenous allocation of students to classes within schools.[86]

[86] To check whether these very low mean values of the distance variable are a consequences of averaging out larger differences across classrooms within schools, in Figure B.6 I plot thee kernel density estimate of the distribution of the gap between the actual and the 1,000 simulated classroom formations of the proportion of female peers for the first class within each school (such class ordering was randomly selected by the School Survey fieldworkers). The mean of this variable is of similar size as found when averaging across all sections in a school, hence pointing against towards a lack of non-random assignment into classes.

C.3.2 Complementary Checks

I complement the evidence in favor of an exogenous assignment of students to classes discussed in Section 3.2 in various ways.

First, one strong piece of additional evidence is that running separate regressions for each school in which the dependent variable is one of the parental attitudes/expectations towards their children and controlling for class fixed effects one would expect the distribution of p-values of the F-tests for joint significance of the class fixed effects to follow a uniform distribution (Feld and Zölitz, 2017). In order to be able to carry out this procedure for the attitudinal characteristics measured in round 2 (which are only available for YL children), I exploit the fact that it is frequent in my sample to have multiple YL children enrolled in the same class (on average there is four) [87] — and they are a random sample of the actual class composition — to carry out this analysis using only information from the YL's round 2 round. The results in Figures B.8. B.9, and B.10 show the expected patterns consistent with an exogenous allocation of children to classes.

Second, I show that the proportion of female classmates is not correlated with any of a wide range of individual observables like ethnicity, parental education, child's health conditions, and educational resources at home.[88] Table A.4 shows that this is not the case.[89] I also exploit rarely-available information reported by each student's head teacher during the first round of the 2011 School Survey on how strong (s)he perceives the academic ability and the parental support for academics to be. This is arguably an a priori set of key individual characteristics that the school would have weighted-in had the allocation of students to classes not been exogenous. Table A.5 shows, once again, that these individual characteristics are not systematically correlated with the proportion of female peers each student is exposed to.

Third, I follow Chetty et al. (2011) in regressing students' gender on school and class fixed effects. Given random allocation of female students — a predetermined variable — across classes within schools, an F-test of the joint significance of the class dummies should not show a significant relationship after the inclusion of school fixed effects. This is indeed the case, with the test yielding a p-value of 0.97.[90]

[87] As consistent with randomization, the proportion at the school level has the same mean and standard deviation as the one at the class level.

[88] As a comparison, one of the most studied settings benefiting from randomized class allocation, Project STAR, only allows for the exploration of baseline balance for age, gender, race, and free-lunch status. This leaves the door open to potential differences in other dimensions for students across classes and treatment arms (Hanushek, 2003; Chetty et al., 2011).

[89] I also check for differences in gender. For this, I follow Guryan et al. (2009) in controlling for the proportion of female peers at the school level to account for the fact that there is a bias arising from the sampling of peers without replacement. No relationship is found at convention levels of significance (p-value of 0.111).

[90] I additionally again follow Feld and Zölitz (2017) in running separate regressions per each school for the main predetermined characteristic (e.g. gender, age — which captures grade repetition/grade skipping) and the regressors are class dummies. Figure B.11 reinforces the claim that class allocation is as good as random.

Finally, I perform Fisher's exact test for all multiple-class schools (38) and for five different dimensions (gender, wealth, parental education, age, and health) yields 4.8% of the total number of p-values computed to be below or equal to 5% — and none for the specific case of gender. This means that only for 5% of the groupings I am not able to reject the presence of non-random formation.

C.4 Exploring the Role of Teachers' Gender

From the school survey I also observe the gender of each class' head ("homeroom") teacher. Past research has relied on this as a proxy for the presence of role models and/or stereotypes (Bettinger and Long, 2005; Antecol et al., 2015). Given that 95% of the homeroom teachers in my sample are responsible of instructing both mathematics and Vietnamese to the students, the interactions with them are intense and prolonged.[91]

In Figure B.14 in the Online Appendix I provide the estimates from the same analysis as in Figure 4 separately for each attitudinal question and with the inclusion of an indicator for being exposed to a female teacher. The results show that while this event further reduces the traditionalism in gender views of males (negative and significant point estimates on the right part of the (a) panel), the impact is negative for some dimensions among females — panel (b). This finding, which highlights the differential effect between being exposed to same-gender peers (horizontal relation) and same-gender role models (teachers, vertical relation) is not surprising in light of the facts that teaching primary students is a traditionally female-dominated occupation, particularly in primary education (Brundrett and Dung, 2018): 75% of the classes in the sample had a female head teacher, and 45% of female YL children responded in Round 3 that they wanted to be a teacher, compared to 10% among boys. This makes female teachers likely to instill their own traditional views on gender roles on the students (e.g. Lim and Meer, 2017). At the same time, male teachers are a selected subgroup of the male population in terms of traditional views.[92] For instance, in my sample, while the value computed from the principal component of six questions eliciting how much teachers believe that teachers can influence children's lives and academic outcomes is 0.14 for male teachers, it is only -0.05 for female teachers (higher values reflect more agreement with teachers' being able to influence students' achievements; more details on this variable are on the notes of Table A.8).[93] This idea that teacher

[91] In my sample, the total weekly time of instruction of this two subjects is, on average, 10 hours.

[92] It is worth noting that, while school teachers are mostly college-educated and more educated people do hold more modern beliefs, traditional views are still strongly present among well-educated individuals. For instance, using the old cohort's views in round 5 and their education level (which is unlikely to be censored as they are already 22 years of age) I find that, controlling for regional fixed effects, individuals with a college degree have an average agreement with traditional gender norms that is 0.15 points lower than those individuals with primary education and below (the mean is 2.16 and the standard deviation is 0.31). Interacting the education category with a gender dummy does not show statistically significant differences.

[93] Note in passing that this is the opposite to what we found for gender attitudes, where females are significantly less traditional.

gender is correlated with important stereotyping dimensions that can influence students' outcomes was recently documented by Carlana (2019), who focuses on gender-related biased perceptions of proficiency in science and humanistic fields in a context where the composition of who becomes a teacher is considerable less selected than in my case.

Table A.8 in the Online Appendix provides suggestive evidence that the negative effects among girls of being exposed to a female head teacher tends to be weaker among those children whose teacher had stronger beliefs that instructors are able to affect a child's outcomes (the estimate for the interaction between an indicator for having a female teacher and the principal component of a teacher's belief of capacity to influence child outcomes is negative).[94] Table A.9 provides the counterpart for males, which reinforces that pure exposure to female teachers help them transition to less traditional views, irrespective of the beliefs of the teacher (the combined estimates on beliefs and its interaction with the indicator of female teacher tend to net out). It is relevant to conclude this discussion by noting that the above findings for girls are unlikely to be driven by coercive behavior on the teacher side, as I do not find differences between male and female classmates when answering the following questions: "how do you feel your teacher treats you (relative to your classmates)"; "I get frightened when I am asked a question by the teacher"; and "do you fear your teacher's punishment?". There neither are differential responses among females when the teacher is male or female.

C.5 Further Discussion on Friendship Formation

As mentioned in the main text, a reasonable counterfactual to benchmark the estimates in Table 11 is the one in which friendships are formed at random. One should note, however, that since male children are significantly more likely to form friendships with other males than with females[95] (among those students classified as friends in column 1 in Table 11, 35% are females and 65% are males), the thought experiment of removing one male peer and substituting it for a female one would mechanically increase the proportion of female peers.[96]

With this in mind, and given that I do not pursue the endeavour of modelling friendship formation, one pathway is to undertake Monte Carlo simulations to compute the average change in the proportion of female friends that would be expected, under the current gender partition of friendship networks, after replacing one male peer by a female one among the actual classroom composition. To be more specific, I proceed as follows. I first compute the proportion

[94] While we expect to be measurement error in teacher's beliefs, the fact that it is computed out of multiple correlated questions should partially alleviate this limitation.

[95] This empirical pattern of systematic tendencies of forming friendships with same-gendered individuals is often termed as "homophily" (McPherson et al., 2001).

[96] This is because of two forces: (i) there are more females to be friends with, and (ii) the *total* number of friends (which is at the denominator when computing the proportion of female friends among all friends) decreases because males form more friendships with other males than with females.

of female and male friends in the actual data that satisfy the friendship definition in column 1 of Table 11 — 0.35, as reported in the Table. Second, for every *male* child, I randomly draw a number from a uniform distribution in the 0-1 interval for each of his classmates. If the number allocated is below 0.35 for a female classmate or below 0.65 for a male one, I define that child as a friend of the reference child. Then, for every child, I replace one male classmate by a female classmate. If the random number assigned to that child is below 0.35 I define her as a friend. I then compute the proportion of female friends over the total number of friends. I repeat this exercise 1,000 times, obtaining an average female proportion of total friends of 0.393 — which is 100*(0.393-0.35) $\approx$ 4.3 percentage points higher than the empirical one.[97]

Now, because this figure is obtained by increasing the number of females in one *unit*, I need to compute how much this increase represents in percentage points of female classmates, which is how my independent variable of interest in Table 11 is measured. In a first step I compute the proportion of female peers when, taking the actual data, I substitute one male per class by a female. I find that an increase in one female in the class raises the proportion of female peers in approximately 13.05 percentage points. Since the point estimate in column 1 in Table 11 is 0.657 per every percentage point increase in the proportion of female classmates, the addition of one female, which represents an increase in 13.05 percentage points, is predicted to raise the proportion of female peers in 13.05*0.657 $\approx$ 8.57 percentage points. Therefore, under the above assumptions, there is evidence that after the swapping of a male for a female peer not only male children would form friendships with a higher number of female classmates, but also to do so disproportionally more than predicted (8.57 $\gg$ 4.3 percentage points).

Additionally, given that Table 3 does not show heterogeneity in the effects on attitudes by gender, in Table A.11 I proceed similarly to Table 11 for the subsample of female children. Several aspects are worth highlighting. First, the mean of the outcome variables is increasing in the degree of "closeness". For instance, while 63% of the "at least a little close" friends in column 1 are females, this figure increases to 89% for "very close" in column 3. This, which is the opposite case to Table 11, was expected: it reinforces the idea that children are more likely to establish closer relationships with same-gendered peers. Second, the qualitative results are in line with those for males: more exposure to female peers significantly increases the proportion of "weak" friends and, in the present case, also of close ones (but not of extremely close ones). Third, the estimated coefficients are larger, as consistent with the fact that females are much more prone to form networks with other females in the first place. Overall, these findings provide further support for increased interactions with females to indeed be a likely channel

[97] In order to put bounds on the new proportion of 0.393, I report that 5% of the simulations for the proportions of female friends are below 0.387 and 5% are above 0.398.

behind my main effects on attitudes, *both* for males *and* females.

Alternative Approach: Child Fixed Effects. I complement the above analysis by creating a panel dataset in which each row is a pairwise relationship between a given child and one of his/her peers. This allows me to include individual fixed effects and to exploit within-child variation in the gender of the classmates (which is exogenous by virtue of thee quasi-experimental allocation of students to classes). I find that male children assigned to classes with more than 50% of females are around 4-5 percentage points more likely to become friends with a girl.

C.6 Other Non-cognitive Skills: Risk Attitudes

Attitudes towards risk have been shown to be an important determinant of late life outcomes (e.g. Hanaoka et al., 2018). Given the large and significant effects found with respect to gender roles one would expect that there may also be a convergence in other crucial attitudes, such as risk. In order to provide some evidence in this respect I take advantage of a special module conducted by YL in order to track children's responses to the 2020 Covid-19 pandemics (i.e. 8 years after the School Survey). This involved three rounds of phone surveys, with one of the questions asked in the first round being "suppose I offered you a lottery ticket that wins 600,000 VND ($25) with 50-50 chance. How much would you be willing to pay for such a ticket?" Given the skewness of the distribution of responses, I construct an indicator taking the value of one if the child's response is above 10,000 and zero otherwise. The results from a linear probability model displayed in Table A.12 show that while males who were exposed to more females during the School Survey academic year are significantly less likely to belong to the high-willingness-to-pay group (column 1), females were not affected (column 2).

I further exploit the context of this special module, namely Covid-19 pandemics, to explore whether these attitudes towards risk are reflected in actual behavior when it comes to protective measures in order to avoid infection. I use a battery of questions asking whether certain measured were used (e.g. whether handshakes and physical greetings are avoided or whether social distance is kept) and obtain the mean across all indicator variables created to signal the undertaking of each adaptive behavior elicited. I find that males exposed to more females indeed were more prone to implement more protective measures, both in the first (column 3) and the second (column 4) phone surveys (undertaken in June/July and August/September 2020, respectively). Given that there is extensive evidence across multiple countries that females tend to be more risk averse (e.g. Croson and Gneezy, 2009; Dohmen et al., 2011)[98], these results are in line with the overall reading of my paper in that higher exposure to female peers leads males to converge towards attitudes, views and preferences that are more in line with those displayed

[98]This also holds in my sample, as shown by the means reported in Table A.12.

by females. No effects are found for females (columns 5 and 6).

C.7 Additional Attitudinal Outcomes: Fertility Preferences

In Table A.18 I explore whether the *ideal* self-reported age of first childbearing, as well as the total number of children and their gender composition, were affected by female exposure. This is an outcome for which there is ample cross-country evidence pointing at systematic differences in males' and females' ideal number of children, with females systematically desiring less (Ashraf et al., 2014).[99] In my sample this is also the case and the differences, although statistically significant at the 10% level, are not large (2 for males and 1.96 for females). Panel A and B report the results for males and females, respectively. Unlike for views on gender norms, here there is significant heterogeneity across genders. Males who experienced more contact with female peers wish to have significantly less children, and this is true equally for sons and daughters. On the other hand, females did not experiment any change: the point estimates for the desired number of children being close to zero so that they remain at a desired level just slightly below the natural replacement rate (1.95)[100] and, if anything, there are some signs that they wish to somewhat postpone the first childbearing. This was expected given the increases in the ambition for a professional career uncovered before, and it leads to a convergence in the stated preferences across genders. This finding is consistent with males becoming more aware of the quantity-quality trade-off and past work emphasizing the role that an increase in female empowerment has in decreasing the desired number of children (Becker, 1960; Westoff et al., 2010), and contributes one of the first pieces of evidence of the role of female peers on fertility decisions after Brenøe and Zölitz (2020) for Denmark.

C.8 Further Evidence on Mechanisms

C.8.1 Heterogeneity

Heterogeneous Effects. Ex ante, one could expect the effects on gender norms to be heterogeneous along multiple dimensions. For example, perhaps individuals from less wealthy families have more deeply-rooted views that are harder to change. However, the converse could very much be true. It therefore remains an empirical question whether the effects uncovered are heterogeneous and, if so, in which directions.

[99] For a review on theories aiming at explaining these differences (e.g. cultural, biological), see Alger and Cox (2013). For a discussion on macroeconomic models emphasizing decisions on fertility and growth with two parent-families, refer to Doepke and Tertilt (2016).

[100] It is worth noting that this figure is consistent with other nationally representative datasets for Vietnam. For instance, using the Demographic Health Survey (DHS), Westoff et al. (2010) finds that 92% of Vietnamese women with two children do not wish to have any more. This is the highest level of opposition to a third children among a set of 60 countries where the DHS was undertaken by the year 2008.

In Table A.19's column 1 I interact the proportion of female peers with an indicator, computed using information from the second round of YL (collected in 2006), that takes the value of 1 if the child's mother was reported to have a saying in at least 50% of a set of important family decisions (enumerated in Section D), and zero otherwise. By focusing on the subset of male children, this analysis can provide important insights on how the effects of female exposure may differ across children with varying levels of initial acceptance of traditional gender roles, as proxied by the maternal decision power within the household experienced by the child prior to class formation. This is a relevant measure of traditionalism as, in patriarchal societies like the Vietnamese, husbands are expected to manage the household finances (as well as providing them in the first place). The findings are the following.[101]

First, as expected, male children who grew up in households where the mother had more power in the family decisions display less traditional gender norms (-0.30, on average). Second, being exposed to more female peers continues to strongly reduce agreement with traditional norms (-0.69). These effects differ, however, based on the children's maternal decision power: the effects of female exposure are significantly smaller for those children whose mother had more saying in household decisions, as shown by the positive and significant coefficient on the interaction term. These results suggest that higher exposure to female peers operated mainly through individuals who would otherwise have been likely to display more traditional views on gender norms — in a sense, affecting the "extensive margin" of agreement with more equal gender roles — rather than through increasing the equality views of children who were more prone to have them to begin with based on their family upbringing. Indeed, while both children whose mothers were and were not important decision makers benefit from female exposure, the absolute value of the effect of the increase in the proportion of female peers (from 0 to 1) is much larger for those whose mother was not (-0.694) than for those whose mother was (-0.694 + 0.643). This points to a converge towards more equal gender views across children with varying levels of baseline views, rather than to a polarization.[102]

In the remaining of the table I consider an extensive set of other potential sources of heterogeneity, pooling male and female children, including: (1) an indicator of any of the parents being illiterate; (2) an indicator for belonging to an ethnic minority; (3) an indicator for

[101] Note that this exercise provides an additional robustness check, as it explicitly estimates regressions that control for household-level gender norms. It is relevant to point that, while I do not find heterogeneity in treatment effects for females based on decision power, the level effects of the proportion of female peers remains stable among the female population.

[102] One may note that, in the extreme case of undertaking the thought experiment of shifting a male student whose mother was not the main decision maker from a class with 0% female peers to another one with 100%, his predicted agreement with traditional gender norms would be lower than for someone whose mother was an important decision maker (-0.694 vs. -0.694 - 0.304 + 0.643). However, for more plausible scenarios that do not require extreme extrapolation — e.g. moving from the 10^{th} (0.333) to the 90^{th} (0.611) percentile of the distribution of female peers, the fact that children with mothers who make more decisions have lower values of agreement with traditional norms (-0.304) would prevail over the marginal effects of increases in the proportion of female peers, hence making children born to mothers with more decision power still more likely to report less traditional gender views.

being located above the median in terms of the wealth index; (4) a continuous measure of self-perception of not being exceptionally good at academics; (5) a continuous measure of low effort at school, and (6) a continuous measure of having high interest about school. The results show that heterogeneous effects along these dimensions are not present: the level effects of the proportion of female peers variable continue being negative, strongly significant and of about the same initial magnitude, and its interaction terms are never significant.[103]

Heterogeneity by Regional Traditionalism. The above results suggest that it is children belonging to relatively more traditional households who experience larger shifts towards less traditional views. I complement that analysis at the micro level by showing that communities (the sampling units in the first stage of the YL's sampling procedure) with more traditional values at the aggregate level also experience larger shifts on average. For this I make use of the views on gender roles reported by the older cohort (Table A.3) in order to compute the average degree of traditionalism at the community level without using measures that have been influenced by my treatment of interest. I then split the sample into traditional and non-traditional communities (they are above and below the median level of traditionalism, respectively). Conducting separate regressions yields negative and significant effects for both subsamples but, once again, they are larger for the more traditional areas (the point estimates are -1.479 and -0.444). The null hypothesis of equality is rejected at the 5% confidence level.[104]

The Role of Female Peers' Academic Ability. The above analyses have pointed to both increased exposure and direct interactions with female peers as drivers of the changes towards less traditional roles. The study of the impact on males suggests that they revise their negative biases about female abilities. If this is indeed the case, one would expect that exposure to more able females should have a separate effect on traditionalism. In Figure B.16 I undertake the same analysis as in Figure 4 with the additional feature of including an indicator taking the value of 1 if the average score among the females in the class is high (I define this threshold as 30% of a standard deviation higher than the sample mean).

The results for the male subsample, presented on the left panel, show that those dimensions intimately connected with academic abilities (parental encouragement for college attendance, the importance of good academics, and intelligence) are shifted towards less traditional values for those children exposed to more academically able female peers.[105] These changes in males'

[103] I do not report the levels of the interacting variables for the sake of parsimony.

[104] A complement of this analysis is to explore whether the results differ in rural and urban locations. I find the main effect to be -0.549 for rural areas and -0.229 in urban areas. The former is significant at the 1% level while the latter is not at conventional levels. This is likely due to lack of power — the urban sample is reduced to 197 observations. I then replicate the exercise allowing for nonlinearities through the inclusion of an indicator for the proportion of female peers to be over 50% and find significant effects both for urban and rural areas at the 2% confidence level. The point estimates are -0.077 and -0.125 for rural and urban locations, respectively.

[105] A natural extension is to interact both variables of interest. The interaction is negative a significant even after controlling for own and male peers' ability when using "intelligence" as outcome.

views are expected to be key in leading to more egalitarian outcomes both at home and in the labor market. Importantly, female peers' ability also impacts the views on intra-household decisions, which so far had remained unaltered. This suggests that males are willing to give more power in the household to wives that are regarded as more able. Interestingly, the responses for swearing instead shift towards more traditional views, which could be rationalized by the fact that more able individuals are less expected to rely on bad manners and poor etiquette. Turning to the results for the female subsample on the right panel we find weaker effects. This suggests that the main effects were more likely driven by reduced exposure to males than to a positive updating of their perceptions of the skills of other females.

C.9 Further Robustness Checks

Excluding Schools with Potential non-Exogenous Allocation of Students. In Section 3.2 I discussed that only for three schools in my sample I cannot not reject the null of no systematic allocation of YL children to classes. In Table A.27 I show that excluding students from these schools in my analysis does not alter the point estimate nor its statistical significance.

Placebo Analysis: Looking at Long-term Outcomes not Expected to Have Changed. As as additional form of placebo test, I look at a series of outcomes measured in round 4 that are highly correlated with household characteristics and that are unlikely to have been affected by peers. This is useful because, while my baseline balance check from Figures 2 and B.7 did not detect statistically significant differences across students exposed to different proportions of female students, one could worry if these students might have nevertheless been in different trajectories that just had not been captured that early in children's lives. Table A.28 shows that various measures of health and human capital (e.g. height, which is largely determined around the time of entry to school[106]) or family resources (wealth index and access to internet) still do not show any statistical differences at round 4 (nor are they present at round 5 - unreported).

Placebo Analysis: Siblings' Outcomes. In the fourth and fifth rounds of the YL survey, cognition and time use of one sibling per child belonging to the longitudinal study was recorded. While it is possible that there are some spillovers from a child's class composition into his/her brother/sister due to contact between siblings at home, I think of sibling's behavior as an additional opportunity to check the validity of my identification assumptions. In particular, we do not expect to find systematic differences in cognition and time use across the siblings of the School Survey respondents based on exposure to female classmates. If differences did emerge, this would be suggestive that children who were exposed to more females might have been on different trajectories that would end up in systematic differences in rounds 4 and 5 even

[106] Martorell and Habicht (1986) discuss how adult height is largely determined by age 7.

if I did not find evidence for this in my balance check for round 2. Table A.29 show that indeed there are no differences in the intensive nor the extensive margin of home production in rounds 4 and 5 nor on cognitive abilities in round 4. What is more, YL's Round 3 also contains sufficient information to allow me to compute home hours provision across siblings. Reassuringly, I do not find statistical differences (results available upon request) at that point in time, which reinforces my interpretation of siblings' outcomes as a placebo test. This lack of systematic differences also among siblings during round 3 further reinforces the lack of different levels/diverging pre-trends for home production across children who were exposed to more or less female peers, as discussed in footnote 53.

No Changes in Educational Investments. It could be the case that parental investments and views on their child's education react to the composition of the class. While, unfortunately, the second round of the school survey does not feature information that could be useful in this sense, I can exploit related information from the fourth wave of the YL survey.

Table A.24 shows that the proportion of female peers is not correlated with the extensive margin of participation in extra tutorials outside school (column 1), nor on the total number of hours or money spent on such activity (columns 2 and 3, respectively). Moreover, there was no change in the ideal education level that the parents wish their child achieves nor in whether they claim to know the names of their child's friends (which I assume to be informative of the degree of parental involvement and time investment on the child).[107] These findings, coupled with the lack of changes in classroom composition years later mentioned above, suggest that parental endogenous reactions to class composition are not present.

Restricting the Within-school Variation Exploited. Although Figure B.2 demonstrates that, while there is sufficient within school variation in the proportion of female peers across classes, the differences are not large enough to suggest that there may have been systematic allocations of children's of a given gender to certain classes. A further robustness check that I perform is to restrict the estimating sample to those schools where there is sufficient — but not "excessive" — variation in the proportion of females. In Table A.30 in the Online Appendix I show that restricting the sample to schools where the gap between the highest and the lowest proportion of female peers that a child is exposed to is less than 0.25 (column 1) or between 0.1 and 0.25 (column 2) does not alter the results.[108] Column 3 shows that this is robust to consid-

[107] There were no changes either along the following three dimensions closely related to parental practices and proximity with the child: (i) parents claim to know what their child does outside school; (ii) parents claim to know the parents of their child's friends, and (iii) parents claim to know their child's teacher.

[108] It is worth emphasizing the following property of my exercise. For classes with 20 sampled students, substituting one male for a female increases the ratio of female peers (leave-out mean) by 0.052632 — both for males and females. This means that, for a school with two classes, one with a 10-10 female-male gender divide and one with a 11-9 one, the gap between the highest and the lowest proportion of female peers is 0.10526 (i.e. 2*0.052632). Indeed, given the linearity of the problem, such gap will always be given by (1 + highest number of females in a class - lowest number of females in a class)*0.052632. Therefore, a school with two classes, one with 11 females and one with 8 already has a gap of 0.2105. A difference of one more females would already set

ering, for instance, only classes where the proportion of females is between 0.35 and 0.65.

Effects on Within-class Dispersion in Gender Norms. An additional piece of suggestive evidence relies on exploring the degree of class-level variation in gender norms within schools. Because a subset of YL children are enrolled in the same class, I can compute a measure of dispersion of gender norms among YL students in a same class. I can then compare the variation in gender norms across classes within a same school that have different proportion of female peers. If female peers do matter, we would likely expect a fall in the dispersion of gender norms. This is an exercise somewhat reminiscent of Carlana (2019), who explores whether the academic gender gap is reduced in classes with a less stereotyping teachers relative to classes with a more stereotyping one, within the same school. I find that, using an indicator for the class having at least 50% of females yields a negative and significant effect on the class level standard deviation in gender norms of -0.098 (significant at the 10% level).

Accounting for Extreme Responses. In order to verify that the results are not driven by those individuals who display the highest and lowest degrees of traditionalism, in Table A.31 I repeat the estimation of column 7 in Table 3 both using as outcome the logarithmic transformation of my gender norms measure (column 1) and winsorizing its values at the 1^{st} and 99^{th} percentiles (column 2). The lack of significant changes with respect to the baseline estimation was to be expected given the results on the quantile regressions.

Further Accounting for Selection into Schools. The inclusion of school fixed effects, together with the observation that the single most important reason for school choice for almost 90% of the parents is school proximity to home, helps building confidence in the ability of my empirical approach to control for selection of students into schools. I now provide further support for this claim by re-estimating the specification in Table 3's column 7 for the subsample of children whose school headmaster reports that there is no other school in the same commune that can attract students from the same overall pool (sample size is 394). I obtain a point estimate of -0.445 — very similar to the baseline result — and significant at the 1% level.[109] This means, in passing, that the complementary subsample, that is, bigger communities that have multiple schools — and where children are more likely to be disconnected from other children outside their own schools — also yield similar findings. Indeed, I obtain a point estimate of -0.552, significant at the 5% level, for the subsample of communities with at least one other school.

it above the 0.25 threshold that I am using.

[109] Given that the variation in the proportion of females that I exploit comes from exogenous allocations at the school level, I verify that my baseline results, which rely on a pooled OLS estimation, hold when I estimate separate effects for each school and account for the noise embedded in this process by taking their average. This is a direct approach to exploiting within-school variation, as opposed to the inclusion of school fixed effects which, while also exploiting within-school variation in the proportion of female peers, still imposes, for instance, that the impact of the covariates is the same across schools.

An additional approach to account for potential selection into schools is to control for more aggregate spatial characteristics beyond the inclusion of school fixed effects (Chung, 2020). In Table A.32 I replicate column 7 in Table 3 while controlling for region[110] and rural/urban location in column 1 and for community fixed effects in column 2. The point estimates remain virtually unchanged, which is consistent with the inclusion of school fixed effects in my baseline specification being sufficient to account for selection into schools.

No Selection into Subsequent Peer Groups. The key benefit of observing my outcome of interest years later is that I am able to explore the long-run effects, but it also allows for the possibility that students and families endogenously react to the initial female peer composition by selecting into future classrooms within or across schools. Although I am not able to observe the class composition for each subsequent academic year for the YL children after Grade 5, I provide two pieces of suggestive evidence that selective classroom formation is unlikely to explain the main findings.

First, Young Lives conducted a second School Survey in 2017 when students were already in Grade 10. While the objective was to sample all YL children who participated in the 2011 School Survey that were enrolled in Grade 10, budgetary reasons led to only a (random) subset of them (N=325) to be sampled (together with *all* of their classmates). I find that only 7% of the YL children who shared the classroom with another YL child in 2011 remain being classmates in 2017. This is an overestimation of the actual class persistence given that all YL children in a given school were sampled in the 2011 SS, but not all the non-YL classmates. Moreover, I correlate the proportion of female peers that each child was exposed to in each of the two academic years surveyed in the school surveys to see if there's persistence in the classroom gender composition faced by students. Column 1 in Table A.22 does not show an statistically significant relationship.

Second, I make use of YL's academic history module to obtain information on whether the child continued being enrolled in the same school or transferred to a different one between Grade 5 and Grade 9[111]. Column 2 in Table A.22 shows no correlation between the female peer composition and whether the child changes schools at some point until Grade 9, inclusive. This is not surprising as only 3% of the children in my sample changed schools.

No Contemporaneous Selection into Peers. A related concern is that children and families may switch schools over the course of primary education after observing an unsatisfactory (for them) class composition in their first school. If that were the case, the final allocation that I

[110] The six regions in my sample are: Northern Uplands, Red River Delta, Phu Yen, Da Nang, South Eastern, and Mekong River Delta.

[111] There is an almost universal school transfer in Grade 10, since students finish middle school and promote to high school. Only a small proportion of schools (about 10% in YL's 2017 School Survey) teach both lower secondary education (from Grade 6 to Grade 9) and upper secondary education (Grade 10 and above).

observe in round 5 could be the product of self-selection. This is, however, an unlikely concern for two reasons. First, we have formally confirmed the quasi-experimental allocation of children to classes in the schools employed in the estimation. Second, as stated in footnote 29, students changing schools throughout primary education is a very infrequent event. In any case, I use detailed information on each YL child yearly enrollment from grades 1 to 5 to identify those individuals that do move schools during that period. Replicating Table 3's column 7 excluding movers yields a point estimate of -0.475 with a standard error of 0.105 (sample size is 855).

Effects not Driven by Persistent Peer Matching. One potential concern about interpreting the effects on attitudes as "long-term" ones is that female exposure could be persistent over the years in-between the school survey and the fifth round of YL. As mentioned, the fact that class composition is observed in Grade 5, the last course of primary education, and that students need to transition to a different school for Grade 6 makes persistence very unlikely — particularly because admissions to lower secondary schools is often competitive, often times based on entrance examinations. Moreover, a similar phenomenon takes place as children transition into high school at Grade 10. This implies that another source of changes in peer composition takes place one year before the fifth round of the YL survey is fielded. which means that the effects uncovered cannot be purely driven by sustained exposure.[112]

As an additional check, one would expect that peer group composition is less persistent in those schools with more classes per grade, as there is more room for new peer group configurations over time. In Online Appendix Table A.23 I reproduce my main analysis restricting my sample to children who attended schools with at least 3 classes (column 1) or at least four classes (column 2). Both the size of the point estimate and its significance are preserved.

[112]It is worth noting that high schools are larger and more dispersed than primary schools (Iyer et al., 2017).

D Variables Description (For Online Publication)

In this appendix I provide further details on the construction of certain variables that were not described in the main text (in alphabetical order).

- "Any parent illiterate": indicator taking the value of 1 if any of the parents were reported to be illiterate, and zero otherwise.

- "Confidence in Self": I use the following set of questions collected in round 5 about the degree of confidence in oneself (both in general and in unfamiliar situations) to create separate binary variables taking the value 1 if the child agrees or strongly agrees with the statement and 0 otherwise (disagree and strongly disagree were the other options) and then obtain their mean at the individual level:

  - I am as good as most people
  - When I am confronted with a problem, I can usually find the means and ways to get what I want
  - I am confident that I could deal efficiently with unexpected events
  - I can do things as well as most people
  - I can always manage to solve difficult problems if I try hard enough
  - Thanks to my resourcefulness, I know how to handle unforeseen situations
  - I can solve most problems if I invest the necessary effort

- "Difficult to complete homework?": a three-category (always; sometimes; rarely/never) variable for the question "Do you find it difficult to complete your homework?"

- "Enjoy school": a three-category (always; sometimes; rarely/never) variable for the question "Do you enjoy school?"

- "Feel pressure at school?": a three-category (always; sometimes; rarely/never) variable for the question "Do you feel under pressure to perform well at school?"

- "High interest in school" is the principal component of the following five items (each has four categories: strongly agree; agree; disagree; strongly disagree):

  - I daydream a lot in class
  - I often do my homework without thinking
  - I am usually interested in my schoolwork (I invert it)

  - I often feel like quitting school
  - I am always waiting for the lessons to end

- "Hours extra classes": number of hours responded to the question "During an average week, how many hours has your child attended extra classes?"
- "Ideal education level reached": categorical level of education answered by the child's caretaker when asked: "Ideally, what level of formal education would you like your child to complete?"
- "If I work hard I can go to college": a four-category (strongly agree; agree; disagree; strongly disagree) variable for the following statement "If I work hard I think I can go to college or university."
- "Knows names of child's friends": an indicator taking the value of 1 if the caretaker answered yes to the question "Do you know the names of your child's friends?"
- "Long-term Health Issue": an indicator taking the value 1 if the caretaker answered yes to the question "Does the child have any long term health problems that affect his/her daily life?"
- "Low-BMI-for-Age" an indicator taking the value 1 if the child is reported to be "moderately thin" or "thin".
- "Low school effort" is the principal component of the following five items (each has four categories: strongly agree; agree; disagree; strongly disagree):
  - I pay attention to the teachers during lessons
  - I study hard for my tests
  - I am willing to do my best to pass all the subjects
  - I do not give up easily when I am faced with a difficult question in my schoolwork
  - I am not willing to put in more effort in my school work (I invert it)
- "Money extra classes": rupees amount responded to the question "During the school months of the last academic year, approximately how much money has the household paid on average per months for this child's extra classes?"
- "Mother is a decision maker"[113]: an indicator taking the value of 1 if the proportion of

[113] The mean in my estimating sample is 0.68 with a standard deviation of 0.47.

affirmative responses to whether the child's mother is the main decision maker (or decisions are taken jointly by the household) is at least 50% across the following dimensions:

1. Who is the main person responsible for making the key decisions about most of the land?
2. Who mainly controls the use of the earnings from the sale of goods or rent from most of the land?
3. Who is the main person that can make decisions about most of the animals?
4. Who is the main person who controls the use of the earnings from the sale of goods or rent from these animals?
5. Who is the main person responsible for making the key decisions about the sale of livestock or livestock products?
6. Who is the main person responsible for controlling the earnings from the sale of livestock or livestock products?
7. Who is the main person responsible for making the key decisions about work for wages activities?
8. Who is the main person responsible for controlling the earnings from work for wages activities?
9. Who is the main person responsible for making the key decisions about business and self-employment activities?
10. Who is the main person responsible for controlling the earnings from business and self-employment activities?

- "No health problems": an indicator taking the value of 1 if the child stated not to have any health problem that affects him or her in school.
- "No reinforcement class": an indicator taking the value of 1 if the child stated not to be attending extra classes in any subject.
- "Own Popularity": I use the following set of questions collected in round 4 about the degree of own popularity to create separate binary variables taking the value 1 if the child agrees or strongly agrees with the statement and 0 otherwise (disagree and strongly disagree were the other options) and then obtain their mean at the individual level:
  - I make friends easily

  - I am popular with kids of my own age
  - Most other kids like me
  - Other kids want me to be their friend
  - I have lots of friends
  - I am easy to like
  - I get along with other kids easily

- "Takes private classes": indicator taking the value of 1 if the child takes extra private (paid) classes and zero otherwise in the last academic year.

- "Worry about exams": a three-category (always; sometimes; rarely/never) variable for the question "Do you worry about exams/tests?"