

Garcia-Brazales, Javier

Working Paper

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

Suggested Citation: Garcia-Brazales, Javier (2022) : Unlearning Traditionalism: The Long-Run Effects of Schools on Gender Attitudes, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/266544>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

Javier Garcia-Brazales[†]

First Version: January 2020. This Version: December 2022

[Click here for the most updated version](#)

Abstract

Can sustained exposure to females persistently modernize gender attitudes? I study the impact of female peers and teachers on gender roles, perceived relative gender ability, and gender-related behaviors up to nine years later. For this, I exploit two-decade longitudinal information on cognition, attitudes, aspirations, and expectations of a close-to-nationally-representative cohort of Vietnamese primary schoolers exogenously allocated to peers and teachers. I find evidence that being exposed to a higher proportion of female peers weakens traditional gender views both for males and females, and that this translates into actual behavior. Females increase their probability of enrolling at university and in male-dominated majors. A decomposition exercise shows that this is mainly mediated by increased academic aspirations, higher expected returns to education, and less traditional views on the acceptable life goals for females. Males increase both the intensive and extensive margins of home production. These results suggest that early exposure to females can shift slow-moving attitudes even in contexts of high overall cross-gender interactions.

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

JEL Codes: I24; I25; J16.

*I am grateful to Samuel Bentolila, Caterina Calsamiglia, and Monica Martinez-Bravo for their guidance. I also thank Manuel Arellano, Ghazala Azmat, Dmitry Arkhangelsky, Paula Bustos, Guillermo Caruana, Nicole De Vera, Jennifer Graves, Margarita Machelett, Pedro Mira, Mateo Montenegro and multiple participants at Princeton's 2021 YES, Warwick's 9th PhD Workshop, the 2021 EALE Annual Conference, the 6th IZA Workshop: The Economics of Education (2021), the 2021 annual meeting of the Asian and Australasian Society of Labour Economics, the 2021 European Winter Meeting of the Econometric Society, Oxford's 2021 Education Symposium, and CEMFI's seminars for useful comments. The author gratefully acknowledges the funding received from the Fundación Ramón Areces.

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.

[†]Bocconi (LEAP) and University of Exeter, javier.garcia@unibocconi.it

1 Introduction

Gender norms shape views on what is right or desirable for men and women, and have been found to be at the core of multiple gender gaps. They impact, for instance, females' academic and labor opportunities as well as their autonomy in health and family formation decisions (Bertrand et al., 2015; Jayachandran, 2015).¹ While gender norms are subject to change in certain settings, they have proven to be very persistent and little is known about their determinants and about how to influence their formation through policy (Olivetti et al., 2020; Bau, 2021). It is therefore unsurprising that interventions aiming at modernizing views (e.g., gender quotas, affirmative action) lose effectiveness over time due to their inability to fundamentally change these views (Bordalo et al., 2016; Bertrand et al., 2020).²

Recent evidence building upon Allport et al. (1954)'s contact theory suggests that exposure to outgroup members can be an effective source of attitudinal changes. Contemporaneous work from India has uncovered more inclusive views when mixing rich and poor students (Rao, 2019) and team members of different castes (Lowe, 2021). Dahl et al. (2021), the closest paper to mine, focuses on how male's gender views change upon a-few-weeks-contact with females in the Norwegian army, a male-dominated environment. It finds that such exposure decreases traditional views in the short-run but the effect fades away rapidly. This result suggests that more prolonged cross-gender interactions could lead to more sizable and long-lasting changes, particularly if they took place during childhood, a period when attitudes and skills are more malleable (Bertrand, 2019; Ashraf et al., 2020).

In this paper, I explore whether sustained early socialization — more specifically, persistent interactions with female classmates during pre-adolescence — indeed affects traditional gender views and gender-related behaviors of boys and girls over the long-run, and the mechanisms at work. While the role of peers' gender on cognitive outcomes has long been analyzed, evidence on their long-run impacts, on the mechanisms at play, and on how they affect non-cognitive outcomes is scarce (Lavy and Schlosser, 2011; Gong et al., 2019). These limitations are probably due to the high contextual and data requirements. As a matter of fact, to date, no work has placed the focus on direct measures of gender attitudes.

I fill this gap by exploiting a unique combination of widespread exogenous allocation of peers and teachers to classes during primary education in Vietnam³ and rich longitudinal information

¹These studies are rooted in Akerlof and Kranton (2000)'s identity framework, by which social categories define a person's identity and impose costs from not complying with the relevant prescribed behavior.

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

³For context, this is a country where 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 was a list of 77 jobs legally forbidden to women (World Bank, 2015). Still, Vietnam ranks around the median in the United Nation's Gender Empowerment

spanning two decades. I employ Oxford University’s Young Lives (YL), a detailed child-level dataset tracking the lives of a close-to-nationally representative cohort of 2,000 Vietnamese children from age 1 to 19. Apart from collecting academic and (test-based) cognitive information of the children, YL is unique in its emphasis on recording the evolution of their soft-skills (e.g., self-confidence, grit, interpersonal abilities) and of their expectations and aspirations (desired academic achievement and the likelihood of reaching it, labor market occupation and earnings, and family formation, among others). It additionally elicits parental attitudes and expectations for their children, investments in them, and household contextual aspects (e.g., household size, earnings). This allows me to depict a rich picture of these children’s development from birth to early adulthood. Crucially for my long-term results, an impressively-low 4% of the original cohort attrits over the twenty years of the study.

In order to obtain information on their peers and teachers during primary education, I link YL’s longitudinal information to a complementary, large-scale project aiming at providing an in-depth study of the schooling environment (peer, teachers, and schools) during Grade 5 of a subset of 1,100 YL children. Enumerators visited the 92 school sites throughout which these YL children were spread out, and extended the cross-section by sampling 20 randomly selected students in each and every Grade 5 class of the 92 schools. Socio-demographic, cognitive, and attitudinal information of all these classmates was obtained. I use this to construct detailed measures of peer characteristics. This combination of school and longitudinal individual survey data constitutes an unmatched opportunity in the peer effects and contact theory literatures to explore the long-term outcomes of schools both in terms of the range of outcomes and mechanisms observed and the time span of the study.

I then exploit a series of attractive contextual aspects of Vietnam’s primary education system in order to attain causal identification of the parameters of interest. Most notably, while endogenous peer formation is a pressing concern in the literature, I benefit from the fact that, conditional on attendance to a given school, students are allocated in an “as good as random” manner to classes at first grade. This allows me to compare outcomes of students within the same school that, by chance, get assigned to a slightly larger or smaller share of female classmates. Such strategy is supported by valuable information reported by school teachers and headmasters stating that the allocation of students into classes within schools was indeed performed in an unsystematic manner. To further confirm this, I provide multiple checks that the class composition observed ex post is consistent with an ex ante exogenous allocation. These include balancing tests across an extensive list of observable characteristics both of the children

Measure, a measure of opportunity that combines information on political participation and decision making, economic participation and decision making, and power over economic resources.

and of their households measured just a few months prior to class formation (e.g., child’s cognitive reasoning ability, objective and subjective measures of physical health, parental monetary investments in their children, and expectations about their academic and adult outcomes) and showing that the residual variation in the share of females in a class is consistent with the one arising from within-school random allocations of students in a Monte Carlo setting (Cools et al., 2019; Bietenbeck, 2020).

Building on this support for my empirical strategy, I first document that exposure to girls has sizable effects on a composite measure of gender attitudes constructed from a battery of questions eliciting views at age 15 on (i) traditional life goals (e.g., females should focus on being housewives), (ii) relative abilities (males are more intelligent and better leaders), (iii) cross-gender interactions (e.g., females should not be the ones asking or paying for a date), and (iv) intra-households decisions (e.g., females should contribute more to home production). A ten percentage point increase in the proportion of female classmates is associated with a fall in the degree of traditional views of about 20% of a standard deviation.⁴ These large effects are present for both males and females and across the various components of the composite measure.

While these results are already valuable in highlighting a novel determinant of the formation of gender attitudes among children, their relevance further increases when one acknowledges that these attitudinal changes are likely to translate into meaningful, real-world, behavior. In order to show that this is the case, I focus on two salient outcomes for adolescents, each of them expected to be more relevant for one gender. For females, I consider university attendance. This is a meaningful outcome since, in Vietnam, females’ academic aspirations are generally lower despite consistently outperforming their male counterparts in cognitive tests (Mergoupis et al., 2018). Moreover, more educated mothers tend to raise children with better outcomes (e.g., Oreopoulos et al., 2006). I estimate that being exposed to a 10 percentage points higher female share increases the probability of enrolling at university by 10 percentage points. For males, I focus on time devoted to home production, one of the most persistent forms of gender inequality (Teerawichitchainan et al., 2010). I find evidence that both the extensive and intensive margins of home production increase. For instance, the probability of contributing at least 3 hours per day to home production goes up by 7 percentage points when exposure to female peers increases in 10 percentage points.

⁴Given the novelty of the main outcome that I explore, it is not possible to make direct comparison with the existing literature. It is possible, however, to compare the magnitude to that found in the gender peer effects work when the outcome is a measure of cognition as in Gong et al. (2019). I find my estimate to be of about the same size (around 10-15% of a standard deviation). Dhar et al. (2022) also get similarly-sized treatment effects of a dummy indicating participation in their intervention consisting of in-school discussion of gender norms.

Peer effects papers are largely limited in their ability to explore the mechanisms at play. An important contribution of my work is that the unique richness of my data provides me with the opportunity to explore the mechanisms behind the increased probability in female enrollment in tertiary education in a comprehensive manner, including changes in soft-skills, academic and labor market subjective expectations, and gender attitudes. As consistent with the large shift in female’s life goals, I show that at age 15 — i.e., 3 years before attending university — girls who were exposed to a higher proportion of female classmates displayed higher aspirations to reach university. Since academic aspirations should, at least partly, be in line with professional aspirations, I further take advantage of self-reported information on desired professional careers and find that these girls were also more willing to eventually undertake jobs that are less traditionally done by females. Building on the observation that male-dominated jobs tend to be more competitive and highly-paid, I proceed to investigate the presence of a novel and related mechanism: changes in expected returns to education. Relying on information about the expected highest and lowest salary upon university completion and the probabilities attached to each event, I compute the mean expected returns under a triangular distribution (Guiso et al., 2002; Coibion et al., 2021). I find that these girls also perceived college education to have higher returns.⁵ While the respondents are currently still too young for me to observe whether their labor market outcomes actually change, I am nevertheless able to show factual evidence suggesting that this will eventually be the case: these girls did end up being more likely to enroll in male-dominated majors. Finally, exploiting scores in cognitive tests taken by all YL children independent of enrollment status, I find that increased expectations and confidence are not mediated by improved academic performance.⁶ Since, as mentioned, girls tend to outperform boys cognitively, this finding reinforces the idea that girls’ college attainment is more constrained by social norms and related attitudes than by females’ actual skills.⁷

All these mediating factors draw a consistent story behind the decrease in females’ dropout rate that is theoretically rooted in the idea that deviations from traditional gender roles among females are more costly the higher the exposure to males. This is attractive because a typical limitation of peer effects papers is their inability to provide comprehensive views on the mechanisms in place and to quantify their relative importance. In order to achieve the latter,

⁵My work contributes to the literature on beliefs about the returns to education and their impacts on enrollment decisions and preferred field of study (e.g., Stinebrickner and Stinebrickner, 2012; Attanasio and Kaufmann, 2014; Wiswall and Zafar, 2015) 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.

⁶This 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).

⁷The fact that more interactions with other females allows them to move away from traditional views is in line with the overcoming of “stereotype threats” (Marianne, 2011; Zölitz and Feld, 2021). As females are grouped with more females, the intensity of negative stereotypes that they are exposed to from their peers is expected to decline.

I implement [Gelbach \(2016\)](#)'s decomposition to quantify how much of the total variation in female enrollment explained by the above mechanisms was contributed by each mediator. I find that the above mechanisms jointly account for a large fraction (85%) of the total variation in college enrollment associated with an increase in the proportion of female peers. The biggest portion of the effects is explained by higher academic aspirations (35%), followed by increased expected earnings (25%), and increased beliefs that females in general should aim for more ambitious life goals (21%).

Moving on to the mechanisms behind males' increased home production, I employ rare information by which all sampled students report the degree of friendship and the frequency of interactions with every other classmate in the sample. This allows me to investigate the most obvious channel behind the effects: increased daily interactions with female students. I find that males in female-dominated classes are over 4 percentage points more likely to become friends with a given girl than their counterparts in male-dominated classes.⁸

Next, I consider the potential for policy to decrease overall societal traditionalism. Extending my baseline specification, I document the presence of nonlinearities in the effects of the proportion of female peers at the 45 and 55% thresholds in the proportion of female classmates. In a linear-in-means framework, a prerequisite for policies consisting in the reshuffling of students to lead to social gains is the presence of nonlinearities ([Carrell et al., 2013](#)). Otherwise, the effects of switching two students would completely offset each other. 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, 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 gender attitudes is significantly less traditional than the one arising under strict parity.

My results are robust to a rich battery of robustness checks. Most notably, I argue against the possibility that unobserved individual or school characteristics confound the results through a randomized based inference placebo test ([Imbens and Wooldridge, 2009](#)) and I provide further evidence on unconfoundedness by probing the stability of my coefficients to selection on unobservables ([Oster, 2019](#)). Moreover, I extend my main analysis by: (i) exploring teacher gender as an additional driver of gender attitudes and (ii) employing an RDD benefiting from the exogenous variation in the level of education achieved by age 15 generated by the fact that the age of compulsory enrollment into primary education is determined by the date of birth within

⁸This is in line with the predictions of contact theory, which posits that close personal interactions among members from different groups but relatively similar status may limit stereotyping and foster understanding ([Pettigrew and Tropp, 2006](#)).

a natural year. This strategy has recently been employed in my context by [Singh \(2020\)](#) to estimate the academic returns to education. Consistently with the fact that gender attitudes remain rather traditional as countries develop, I find no such return on gender attitudes.

Related Literature. My paper advances various strands of the literature. First, while the work on academic peer effects is vast in terms of cognitive outcomes, explorations of the non-cognitive impacts of peer composition are rather scarce, and limited to behavioral aspects that are directly relevant to education outcomes (e.g., in-class behavior). I instead place the focus on a novel set of outcomes — a wide range of gender attitudes — to provide the first *direct* evidence that exposure to females during primary education reduces various forms of gender traditionalism. This is crucial to understand the *mechanisms* in place behind the impact on *any* gender-related outcome (e.g., school achievement, major choice, labor force participation) and hence to inform policy.⁹ Additionally, two contextual aspects of my setting of interest are infrequent in this literature: (i) I focus on the long-run effects from a 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](#))¹⁰, and (ii) I provide evidence on developing and Southeast Asian countries, where gender roles are a fundamental source of frictions but evidence on peer effects is scarce ([Quinn and Woodruff, 2019](#); [Jayachandran, 2020](#)).

Second, my work builds upon contact theory. As nicely reviewed by [Paluck et al. \(2019\)](#), the focus of this literature has consistently been on the short-run impact of exposure to minorities, with very few studies observing outcomes after six months of exposure and typically lacking evidence on whether changes in self-reported views actually get reflected into meaningful behavior.¹¹ My paper makes two departures from this literature: (i) it considers a context where the group of interest, i.e., females, is as numerous as the male one, and (ii) it explores outcomes in the short-, middle-, and long-run. This not only highlights that short-term studies are not able to capture effects that only show up in the longer run but also allows me to provide the first evidence documenting *persistent* effects on attitudes (in my case, gender-related), which is

⁹A handful of papers have looked at specific outcomes expected to be driven by gender norms — notably the likelihood of studying a male-dominated major at university. [Anelli and Peri \(2019\)](#) 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. Opposite to my results, [Brenøe and Zölitz \(2020\)](#), find that Danish females with more female classmates are more likely to undertake girl-dominated majors. My detailed study of the mechanisms at play is key in providing a clear rationale supporting the direction of the effects in my sample.

¹⁰More generally, the scant exploration of the long-run impacts of peers is mostly focused on academic and labor market outcomes in developed countries ([Black et al., 2013](#); [Carrell et al., 2018](#); [Bietenbeck, 2020](#))

¹¹This 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. \(2021\)](#), up to 6 months later.

critical for policy to yield gains across generations (Darulich, 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. In particular, my causal estimates on attitudes are key in allowing me to rationalize how cross-gender socialization at an early age leads to the observed reduction in the leaky pipeline in female enrollment both at university and in male-dominated majors (Kabeer et al., 2005). This is important since descriptive evidence has highlighted that the presence of a 20% gender wage gap in Vietnam can partly be explained by gender norms driving women to work in lower-paid occupations in search of more flexibility in hours and better non-monetary benefits (Chowdhury et al., 2018). Finally, I explore 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).

Third, my paper connects with the studies on attitude formation and endogenous preferences. While the literature on gender norms has mostly addressed 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. This is important for policy, as interventions that only address one side are prone to be constrained by the immobility of attitudes of the other gender (Dhar et al., 2022)¹². More generally, evidence on the potential for policy to reshape culture is scarce and limited to quantifying the impacts of *large* shocks, such as a switch in the political regime (e.g., Alesina and Fuchs-Schündeln, 2007), which are prone to affect multiple dimensions of culture.¹³

With these limitations in mind, I exploit idiosyncratic 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). Indeed, my mediation analysis emphasizes that the core causes of increased female enrollment into tertiary education are the modernization of gender roles, increased aspirations, and higher expected returns to education. These results constitute a *novel and complementary* explanation for the enrollment impact of female peers that, to date, has traditionally focused

¹²Dhar et al. (2022) 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, 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 intervention. Last, I put the emphasis on schools as a whole by also looking at the role of teachers and the returns to education.

¹³Another important impediment for the effectiveness of existing policy interventions is the lack of explicit and detailed propositions — as opposed to broader prescriptions such as “increasing education” or “empowering women”.

on self-confidence as the main mechanism (e.g., [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 rising literature on female empowerment (e.g., [Bandiera et al., 2020](#); [Field et al., 2021](#))¹⁴ and on how to decrease the leaky pipeline (e.g., [Autor, 2014](#)). I further provide suggestive evidence that this policy is likely to have external validity by extending my analysis to the context of primary education in Ethiopia and showing that qualitatively similar findings seem to be present.

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 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. For instance, I find that, while manipulating peer groups is effective in changing gender attitudes, one additional year of general education is not. This is not surprising given how persistent traditional views have been despite Vietnam’s impressive educational advancement in recent years, and it serves to exemplify that more nuanced policies — such as changing the focus/framing of textbooks or adding gender discussions as part of the curriculum ([Cantoni et al., 2017](#); [Dhar et al., 2022](#)) — are likely more effective to change norms than an additional year of general learning.

Outline of the Paper. The rest of the paper is organized as follows. Section 2 introduces the context and 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 6 evaluates the potential of policies exploiting non-linear effects to reduce overall gender traditionalism. Section 7 probes the robustness of my results. Section 8 concludes.

2 Setting and Data

2.1 Setting

My focus is on primary education, which in Vietnam encompasses Grade 1 to Grade 5 (ages 6 to 10).¹⁵ Almost all schools at this level are state-managed and coeducational ([Dang and Glewwe,](#)

¹⁴This work relies on educational and training interventions. I instead highlight the role of class composition, which does not require additional logistic arrangements nor crowding out students’ and teachers’ time from other valuable activities.

¹⁵Children at these early ages have already been documented to display traditional gender attitudes. For instance, [Bian et al. \(2017\)](#) show that 6 year old children in the U.S. already have stereotypical views favoring males as more academically able.

2018), and access to education is free — although partial financing of schools through parental contributions, known as “socialization” (xa hoi hoa), has gained preponderance. Students are legally guaranteed acceptance into schools within their catchment area and households typically exert little effort to deviate from this. For instance, 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.

The Ministry of Education and Training (MOET) is in charge of designing nation-wide policies that are then disseminated and monitored by the Bureau of Education and Training (BOET) at the district-level. A key aspect of the 2005 Education Law is its emphasis on equal access to resources and learning opportunities both across and within schools, which includes class formation being independent of observable characteristics of the students and their families. Of particular interest to my research is the legal emphasis on gender equality in access to education in the 2006 Law on Gender Equality.¹⁶ Although no formal requirement on the exact procedure to undertake such class assignment exists, this task and, more generally, the management of school activities is the responsibility of school principals, as outlined in publicly available school governance charters from MOET. Class formation takes place in Grade 1 and the general practice is to maintain it until the end of Grade 5, when children transition into middle school.¹⁷ This is attractive in that it generates a sustained treatment, i.e., maintained exposure to the same peers. Students of a same class take all subjects together in their own classroom, and frequently remain as a group for additional afternoon classes. Although interactions across classes can still happen during breaks, the bulk of the time at school is spent with classmates, arguably making this the most relevant peer group in this context.

While my data does not allow me to rule out that, in some specific cases, parents may still be able to informally influence the class to which their child is assigned, Vietnam’s outstanding accountability system operates in order to prevent this event. In particular, every primary education school is supervised by three main entities: (i) local branches of the BOET work closely with headmasters to ensure the correct technical implementation of policies; (ii) the local representative of the Communist party oversees more general matters about the provision of education at school, and (iii) a Parent Board and a Parent Committee formed by parents of the students are present in every school and class and have the legal obligation to supervise education provision and quality. This translates into them having “wide-ranging powers, including

¹⁶Articles 14 to 17.

¹⁷An exception is that a small proportion of students with high ability will be separated in advanced classes at the end of primary education as an early form of ability tracking, which is much more prevalent in middle and high school. The school survey has information that allows me to identify such students so that I can show that my results are robust with and without their inclusion.

‘the monitoring of school activities’” (McAleavy et al., 2018). Indeed, international comparisons emanating from PISA’s headmaster questionnaires position Vietnam among the countries with the largest parental accountability worldwide. Such tight supervision is complemented by cultural norms dictating that displaying an ethical behavior is the single most important trait of a good teacher. This makes deviations from proper behavior costly for headmasters, who almost invariably are experienced teachers who regard their position as “an honorable responsibility” and are expected to lead by example (McAleavy et al., 2018).

With these aspects in mind, it is not surprising that existing anecdotal qualitative evidence suggests that exogenous allocation of students to classes within schools during primary education is indeed widespread (Behr, 2005). This will be the key identification assumption in my empirical strategy and can be tested. Reassuringly, as discussed later, valuable teacher-reported information on the criteria followed by the school to form classes shows that the overwhelming majority of classes are stated to have been exogenously formed. I do not take this for granted and provide multiple checks verifying that the class composition observed ex post is indeed consistent with an ex ante exogenous allocation. A similar empirical approach for contexts where quasi-experimental allocation of students to classes is presumably present albeit not guaranteed in the absence of a randomized allocation conducted by the researcher are Carlana (2019) for Italy and Gong et al. (2018) for China.

2.2 Young Lives

Young Lives is a longitudinal study led by Oxford University collecting high-quality data on the living conditions, environments, and decisions of two cohorts of children from childhood to early adulthood across four developing countries¹⁸. The younger cohort was aged 6-18 months in 2002 (the time of the first round) while the older one was 8 years old. So far, five main data collection rounds have taken place in 2002, 2006, 2009, 2012, and 2016 (see Figure 1). A sixth round was implemented in 2020 to collect relevant pre- and post-Covid information, such as school enrollment and labor market outcomes.

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 com-

¹⁸Vietnam, India, Ethiopia, and Peru.

mune who had children aged either 1 or 8 at the time of the first round were then randomly sampled.¹⁹ I focus on the younger cohort since it is the only one for which information on peers and teachers was collected.

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

YL additionally undertook a School Survey (SS) in the 2011-2012 academic year.²⁰ Its objective was to provide an in-depth study of the academic environment (peers, teachers, and schools) of the subset of 1,138 YL children enrolled in Grade 5 during that academic year.²¹ For this, all schools located in any of the twenty sites with at least one YL student enrolled in Grade 5 were sampled. For every Grade 5 class in a school, up to twenty randomly selected students²² were also surveyed both at the beginning and at the end of the 2011-2012 academic year. Figure 1 depicts the Vietnamese education system and how the timing of the surveys relates to it.

The original 2011 School Survey dataset contains 3,284 students distributed across 176 classes in 92 school sites²³. As mentioned, 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 2011-2012 academic year.

From the School Survey, 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 ranging between 0 and 1 based on the proportion of affirmative responses to questions on 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)

¹⁹For more details, refer to [Young Lives \(2014\)](#) and Section C.1.

²⁰A second one took place in 2016-2017.

²¹The main reason for not being part of the School Survey was that the YL child was born after 31 December. This date generates a discontinuity in the legal time of enrollment in primary education hence making these children be enrolled in Grade 4 at the time of the school survey.

²²For instance, if a school had three Grade 5 classes, up to 60 children 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).

²³It 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 for the Young Cohort

Age	Academic Level	YL Survey - Survey Year
1		Round 1 -2002
2		
3	Pre-School	
4		
5	Kindergarten	Round 2 -2006
6		
7		
8	Grade 1-5	Round 3 -2009
9		
10		School Survey -2011
11		Round 4 -2012
12		
13	Middle School	
14		
15		Round 5 -2016
16		
17	High School	
18		
19	University	Round 6 -2020 (Covid)

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. 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.

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* 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, information on gender, education, wealth, experience, and locus-of-control (ability to influence children outcomes) were collected. More explicit information on the data used is provided in Section C.2.

I then link school and peer characteristics to the YL children’s longitudinal information in order to construct a unique dataset in terms of the richness of information available and the long time span covered. I now discuss the construction of my main outcome of interest as well as the teacher-provided information on the method used to form classes.

Gender Attitudes. I obtain long-term views on gender norms from a battery of twelve statements elicited exclusively in the fifth wave (age 15) of the YL longitudinal survey. 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.

Respondents were asked to describe the extent of their 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 such as being good housewives or schooling being less important for them than 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 and a more favorable split of outcomes within the household.²⁴

To gauge an overall effect, to increase statistical power, and to facilitate interpretation, in my main analysis I aggregate the twelve dimensions by first computing their simple average reflecting the level of agreement towards traditional norms — the four options are coded 1-4 with higher values indicating a more traditional view on the position of females²⁵ — and then converting it into a z-score (mean 0 and standard deviation of 1). As discussed below, my results are robust to relying instead on more sophisticated methods of data aggregation, such as constructing variance-weighted indices.

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. YL presented them with various potential allocation mechanisms, and the responses were as follows:

²⁴Employing the extent of agreement to similar statements to measure gender attitudes is widespread in the gender literature. See, for instance, [Bertrand et al. \(2015\)](#) and [Goussé et al. \(2017\)](#). Social desirability biases, which could potentially affect the results if individuals with more or less female peers were more or less likely to misreport, are unlikely to be a large concern because (i) large levels of traditionalism are still frequently reported for numerous dimensions; (ii) the gender attitudes questions are not a salient part of the survey so individuals are unlikely to pay particular attention to them; (iii) correlating my measure of traditionalism with self-reported views in the spirit of [Crowne and Marlowe \(1960\)](#) where one would also expect to find social desirability biases (own laziness, thoroughness in job performance, being moody or easily upset, or being disorganized) yields low correlations and often times in the “wrong” direction, and (iv) the most natural concern in this type of studies would be biases due to experimenter demand effect, which is irrelevant in my context. My independent variable of interest is computed predating the collection of gender attitudes, is not straightforwardly relatable to gender attitudes, and did not arise from a policy intervention aiming at changing attitudes.

²⁵I flip the ordering when necessary.

(i) randomly²⁶ (77.73% of the students); (ii) there was only one Grade 5 class²⁷ (6.68%); (iii) by location of residence²⁸ (6.62%); (iv) by general ability (4.75%); (v) by ability in mathematics (2.76%)²⁹, (vi) by age (1.47%), (vii) by ethnicity (none); (viii) by ability in Vietnamese (none), and (ix) other (none). These figures show a very attractive characteristic of my context of interest: over 90% (77.73+6.68+6.62%) of the students were exogenously allocated to classes. I provide detailed evidence on the success of the quasi-experimental allocation of students to classes in Section 3.2. In my main analysis, I focus on this subset of students but I also show the robustness of my results to alternative sample selections.

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 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)³⁰ and those with small peer groups (classes with less than eight pupils) leaves me with 937 student-level observations distributed across 74 school sites (i.e., including satellite locations³¹) and 152 classes. The number of observations is slightly reduced due to missing covariates.

2.3 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.³² The average age is 10.29, as expected for Grade 5 students.

²⁶While it is possible that teachers and headmasters do not understand the notion of “random” in the same way that academic economists do, this mechanism is expected to constitute an unsystematic allocative process of students to classes, which in academic jargon we would more usually define as “as good as random” or “exogenous”.

²⁷This is effectively an exogenous form of allocation as the absence of multiple classes renders selection of students into classes infeasible.

²⁸This form of allocation pertains to satellite schools, which are blocks of classrooms located at a distance from the main school 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. It is therefore natural to think of them as separate units from the main school, and hence as an exogenous determinant of class formation conditional on satellite fixed effects.

²⁹As mentioned, 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+2.76% of the cases in my sample).

³⁰Feld and Zölitz (2017) show how non-random selection based on ability largely biases the estimation of cognitive peer effects upwards.

³¹Given the geographical separation of satellites 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.

³²This 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

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 is around 0.5. Extreme proportions (e.g., below 0.3 or above 0.6) are very infrequent³³, as consistent with exogenous allocation of students to classes³⁴.

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 traditional view indicated by the row variable. I additionally document the distributions of responses separately for males and females and including the four original possible responses in Online Appendix Figures B.3 and B.4.³⁵

Overall, gender norms are rather regressive among 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, there is significantly less support for traditional norms in other dimensions. 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, which was 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 and suggests that gender attitudes might already be well-developed by age 15.

in Vietnam — males and females do not differ in their enrollment rates (Nguyen, 2012). Importantly, this figure (and others such as the share of female teachers) closely match nation-wide averages at primary education computed using administrative data (Mai and Brundrett, 2020).

³³There 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 infrequent, as consistent with exogenous assignment of students to classes. Finally, it is worth remarking that co-ed classes are universal in the sample.

³⁴The 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).

³⁵Additionally, 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 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 — 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 is 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 countries but much more traditional than American or European (unreported) ones.

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

	(1)	(2)	(3)	(4)	(5)
	Mean	Standard Deviation	Min	Max	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	0.27	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. Health problems are self-reported by the children in relation to any condition that affects their performance at school (for context, national statistics show that 32% of Vietnamese children are stunted, [Young Lives, 2014](#)).

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 (and zero if (s)he disagrees or strongly disagrees). The exact statements are reported under “Views on Gender Norms” in Appendix C. Mean Score is computed as the average score (on a 1-4 scale) across the twelve dimensions at the individual level.

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:

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

where y refers to an outcome of interest (primarily views on gender norms at YL’s 2016 round), and ics denotes student i enrolled in class c at school s during primary education. Adding a “-” sign in front is short-hand notation for indicating that a variable is computed as the leave-out-mean among the class peers of individual i . ϵ is the error term.

PF measures the proportion of female students in individual i ’s Grade 5 class, excluding the person of reference.³⁶ β_1 , our coefficient of interest, captures the average effect of sustained exposure during primary education to a higher proportion of female classmates (when children progress to lower secondary education in Grade 6 they invariably transition to different schools, often times significantly far from their primary education centers, and change peers).³⁷ β_1 captures an “exogenous” peer effect in Manski (1993)’s terminology, as it arises from background characteristics of the students, and not from malleable dimensions such as their achievement or their behavior.

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.³⁸ The standard way of dealing with this

³⁶Recall that classmates are a natural reference group. 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).

³⁷The headmasters from all 74 school sites in my sample report that the school exclusively teaches grades 1 to 5. It is worth noting 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.

³⁸Recall that school choice on the students’ side is not a very sophisticated process in practice. 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,

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).³⁹

The addition of school fixed effects cannot, however, account for 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 on 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. The variation that I exploit therefore comes from the fact that, by chance, one class within a school has a slightly higher share of female classmates than another.⁴⁰ My main findings will prove robust to multiple checks reported in Section 7, including randomization based inference (Imbens and Wooldridge, 2009), coefficient stability to selection on unobservables (Oster, 2019), and removal of influential observations (Broderick et al., 2020), as well as to different sample selection criteria.

Moreover, one central concern in long-run studies is selection out of the sample. Young Lives made an impressive effort of participant tracking (the attrition rate is only 2.5% over 15 years), which minimizes such concerns, as further addressed in Section 7.⁴¹

Another 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

36.54% of them 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%)

³⁹The 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.

⁴⁰To visualize this, one can think of a situation where a school has a total of 60 students (30 girls) allocated into two equally-sized classes. I will be relying on natural variation by which one class might end up having 16 girls and another one 14. In Section 3.2, I verify that there is still sufficient within school variation in PF across classes to identify β_1 .

⁴¹The 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.

more than 20 students⁴²) is classical, this will bias the estimates towards zero. Under random assignment to classes within schools and if students are missing at random, which both hold given the design of the school survey (i.e., random inclusion of 20 students per class), [Sojourner \(2013\)](#) shows how 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 (I can compute it because I know the size of each class). I take this correction as my preferred specification.

With these aspects in mind, identification is achieved from variation in the percentage of female classmates within schools across classes.⁴³ 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 \bar{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. I cluster the errors at the class level (the level of randomization) in order to account for possible correlation of the outcomes among classmates.⁴⁴

3.2 Validity of Exogenous Class Formation

The key identification assumption is that the exogenous assignment to peers within schools implemented by the principals was successful. The richness of my data allows me to perform a wide range of tests that consistently point in that direction.

3.2.1 Balanced Characteristics

I take advantage of the longitudinal information from YL, which provides detailed individual and household information measured in 2006, hence *just predating* enrollment in the first year of primary education. In [Figure 2](#) I show that, conditional on school fixed effects, observable characteristics that one would expect to influence the incentives of parents and/or headmasters to select certain children into specific classes are not correlated with my treatment — the share of females classmates in Grade 5. In particular, I find no evidence in favor of systematic differences along a set of dimensions pertaining to child development (objective and subjective physical health, quantitative cognitive ability), parental investments (measured by log food and non-food consumption), parental locus of control (e.g., whether they believe that they can help

⁴²Both the mean and the median class size among the students in my estimating sample are 28.5 and the 75th percentile is 33.

⁴³On average, there are 22 students per school who belong to the YL survey and so for whom I can observe their long-run views on gender roles.

⁴⁴My main results hold under the more conservative choice of clustering errors at the school or the commune (third-degree administrative unit) level.

their child if sick, which could correlate both with the incentives and the ability to affect class assignment) nor in terms of parental expectations about adult outcomes of their child (e.g., ideal age at which their child should start earning income).

Appendix Figure B.7 abounds on this conditional exogeneity by showing that the parent's desired level of education for their child does not differ either, nor does their perceived probability that such level is eventually reached. There are no differences either in the reasons that parents allude to for having children, such as to provide additional economic or old-age care support. 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 measure of traditionalism, and find no systematic relationship with class gender composition. Finally, female peer composition is not related either to a set of observable teacher characteristics such as their gender, tenure, type of contract, their beliefs that they can succeed at helping even the worst students, and, notably, their pedagogical ability⁴⁵ (Tables A.6 and A.7).

While, unfortunately, I do not have measures of gender attitudes prior to Grade 1 enrollment, as a complementary check, I take advantage of the gender attitudes elicited for the subset of young cohort students that *did not* participate in the school survey to obtain estimates of the impact of relevant observables⁴⁶ on gender attitudes. I then use these estimates to predict gender attitudes for those children that *did* participate in the school survey. Regressing these predicted attitudes on the proportion of female peers yields a point estimate of -0.042 with a p-value of 0.401. Hence, as much as a wide range of observables can predict gender views, there was no difference between those children that were assigned to more or less females. Following Altonji et al. (2005), the lack of selection on observables suggests that selection on unobservables is unlikely to contaminate the exogeneity of the female share.

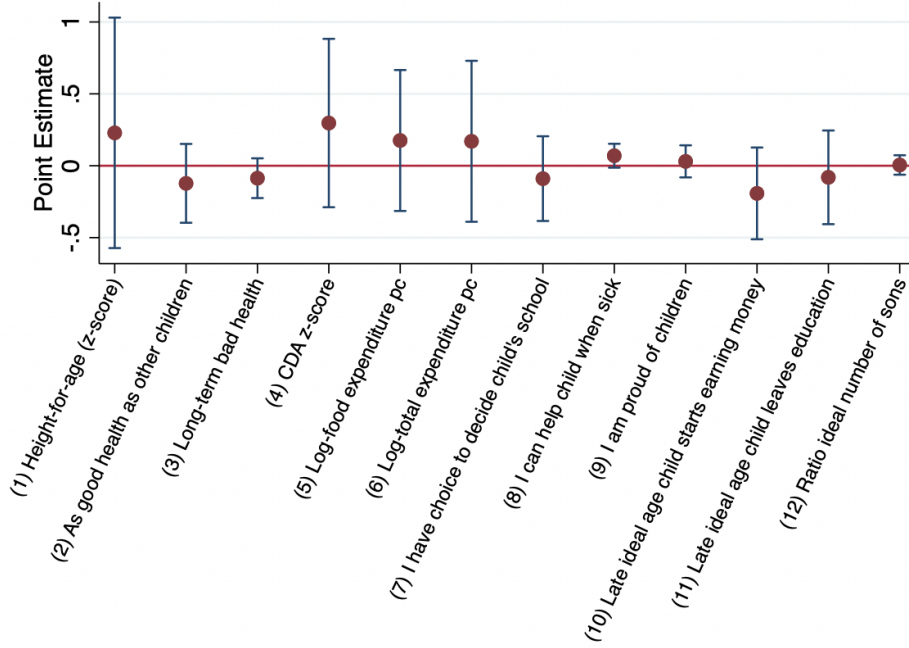
For the final balance check, I build upon the fact that there generally are multiple YL children per class (on average there are four). Because the YL children sampled in the first wave of the longitudinal study were selected as a random sample of the individuals in their cohort, I can verify the non-systematic allocation of students to classes by testing whether the distribution of YL children across classes within schools is consistent with an exogenous assignment.⁴⁷ I first replicate the analysis in Figure 2 where the dependent variable takes the

⁴⁵This is a unique piece of information from the School Survey 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.

⁴⁶These 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.

⁴⁷It is relevant to note that there is very little variation in class size (number of students) 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

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



Notes. The specification estimated is: $y_{ics} = \alpha + \beta_1 PF_{-ics} + \beta_2 Female_{ics} + \lambda_s + \epsilon_{ics}$, where y represents the various observable characteristics for which I conduct the balance tests. The rest of the notation follows that of Equation 1. 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 the child has some long-term health problems, respectively. (4) uses standardized scores in the Cognitive Development Assessment – Quantitative test (CDA-Q). Outcomes (5) and (6) measure the household’s (log) monetary expenses (per capita) on food and on any other dimension, respectively (with a two-week recall). (7)-(9) are indicators for whether parents agree or strongly agree with the following three statements: (a) “I can have a choice about which school to send my child to”, (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”. (10) and (11) are indicators for whether the parents believe that their child should fulfill the respective statements after ages 25 and 23, respectively. The sample includes all YL children that were also surveyed in the School Survey and whose class-assignment satisfies the selection criteria outlined in Section 2.2. 90% confidence intervals are reported.

value of one if the child is a participant in YL and zero otherwise. I obtain a point estimate of 0.104 with a p-value of 0.387. I additionally follow the literature in performing Fisher’s exact test for whether students’ observables (i.e., being a YL child) and class assignment are statistically independent (e.g., Lavy and Sand, 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.⁴⁸

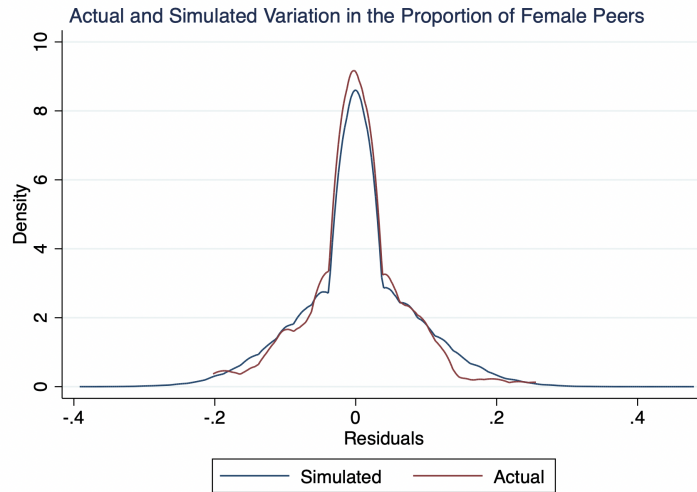
standard deviation being 0.05.

⁴⁸A 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.

3.2.2 The Empirical Variation in Female Share is Consistent with a Randomized Setting

In order to show that my identifying variation is as good as random, 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. 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 employed is indeed quasi-random.⁴⁹ It is relevant to note that, in this regression, $1 - R^2 \approx 0.41$, which highlights that there is enough identifying variation left after the inclusion of our maintained controls (see also footnote 33). Further evidence in favor of the exogenous assignment of students to classes is provided in Appendix C.3.2.

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



Notes. Distribution of residuals from a regression of the share of female peers on gender, main controls (age, parental education, ethnicity, educational resources and wealth indices, and number of books at home), and school fixed effects. The red curve depicts the residuals generated from the actual data, while the blue one is obtained from randomly allocating existing students within a school across classes (1,000 repetitions).

⁴⁹In a similar vein, I follow [Lavy and Schlosser \(2011\)](#) in obtaining the empirical distribution of the standard deviation of the residuals from a regression of the proportion of female peers on school fixed effects. For this, the gender of the children in each simulation is drawn from a binomial distribution where the success probability is the proportion of female students in the school in the actual sample. I find that, for 97% of my schools, the observed standard deviation of the residuals in the true data is contained within the 90% empirical confidence interval.

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).

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 (the outcome is a z-score). 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 it should be under exogenous assignment. Both of these theoretical predictions hold.

Peers’ Gender or Correlated Characteristics? Given that females may systematically differ from males along dimensions other than gender, it is relevant to explore the stability of our coefficient to the inclusion of other peer dimensions such as household-level characteristics and, arguably more importantly, personality traits. 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 additionally feature a 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) accounting for usually-unobservable traits likely correlated with the outcome and that tend to differ across genders.⁵⁰

Finally, in column 5 I explore whether the effects are heterogeneous for males and females. 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 a quarter of a standard deviation, and significant at the

⁵⁰In any case, one would argue that, as long as potentially omitted variables are systematically (and consistently) related to gender, the estimated coefficient provides the policy estimate of interest.

1% level.⁵¹

While the various specifications in Panel B are already demanding, I further verify the stability of this result in multiple ways. In Table A.8, I show that it holds under alternative sample selections based on class formation (e.g., only including students under a fully “random” assignment as reported by the teachers). In Figure B.12, I document the stability of the 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 controls⁵² 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) when also contemplating employing as covariates the dimensions analyzed in the balance tests (Figures 2 and B.7) and teacher characteristics (Table A.6) only suggests to control for gender and the wealth index (which was already done in Table 3’s columns 2 and 7). Furthermore, I allow for an even more flexible model specification through double-selection lasso linear regressions including all possible interactions among the controls included in column 7 (as well as their squared terms), which leads to a point estimate of -2.491, very similar to the -2.512 reported in Table 3.⁵³

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. I follow Anderson (2008) in constructing variance-weighted indices for the four attitudinal categories after dichotomizing each of their items (as I did in Table 2). Table 4’s Panel A shows negative estimates over the board. Females are less expected to have traditional life goals, are less often regarded as possessing lower cognitive and leadership abilities, and are expected to interact with males in a more equal manner. The estimate on intra-household decisions is also negative, but it is smaller in size and non-significant.

Heterogeneity by Gender. 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

⁵¹Results when first standardizing each component into a z-score and then standardizing again are provided in Appendix Table A.9.

⁵²I 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. In unreported results I do feature a binary version of it with no effect on the results. 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.

⁵³Given that Belloni et al. (2014)’s algorithm highlights the predictive power of the wealth index, I also experiment by allowing for up to quintic polynomials of this variable with no noticeable change.

Table 3: Effects of the Proportion of Female Peers on Long-term Views on Gender Attitudes (R5, Age 15)

Panel A: Without Correction for Partial Observability of Peers					
	(1)	(2)	(3)	(4)	(5)
	Mean Agreement with Traditional Gender Norms at Age 15				
Proportion Female Peers	-0.972** (0.433)	-0.865** (0.398)	-0.835** (0.399)	-0.958** (0.443)	-0.911* (0.530)
Female	-0.547*** (0.065)	-0.552*** (0.064)	-0.543*** (0.066)	-0.451*** (0.064)	-0.596* (0.334)
Female*Prop. Female Peers					0.093 (0.668)
Additional Controls	None	Indiv.	Indiv.+Peers	I.+P+Non-cognitive	Indiv.
Observations	880	880	880	792	880
R-squared	0.228	0.243	0.253	0.310	0.243
Panel B: With Correction for Partial Observability of Peers					
	(6)	(7)	(8)	(9)	(10)
	Mean Agreement with Traditional Gender Norms at Age 15				
Proportion Female Peers	-2.719*** (0.519)	-2.512*** (0.491)	-2.204*** (0.747)	-1.532** (0.700)	-2.476*** (0.542)
Female	-0.596*** (0.066)	-0.597*** (0.065)	-0.581*** (0.068)	-0.461*** (0.064)	-0.553** (0.236)
Female*Prop. Female Peers					-0.129 (0.661)
Additional Controls	None	Indiv.	Indiv.+Peers	I.+P+Non-cognitive	Indiv.
Observations	880	880	880	792	880
R-squared	0.276	0.291	0.300	0.310	0.291

Notes. The outcome is the z-scored version of a composite attitudinal measure (theoretically ranging from 1 to 4) described at the bottom of Table 2. 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. Panel B proceeds similarly but accounting for the partial observability of peers by including the interaction between school fixed effects and the proportion of peers observed in the class and multiplying the share of female classmates by the proportion of peers observed (Sojourner, 2013). 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 2.2 and who are also observed in round 5 of the YL longitudinal dataset. Standard errors clustered at the class level (151 clusters) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

are large, and extend across all dimensions — the effect on intra-household decisions is still non-significant, but it becomes apparent that the negative sign for the pooled sample was driven by females. 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, the children and the husband’s family, as well as displaying modest manners and proper speech, and these roles have proven hard to change (Jiang, 2009).⁵⁴ Their views on females’ overall ability remains not statistically significant.

4.3 The Role of Teacher Gender

Given existing evidence that teachers are important determinants of students’ outcomes (Dee, 2007; Muralidharan and Sheth, 2016), I complement this literature by considering the possibility that being exposed to female teachers can be an additional driver of attitudinal shifts.⁵⁵ In my sample, 95% of the homeroom teachers are responsible for instructing both mathematics and Vietnamese. This creates intense and prolonged interactions with the students (the total weekly time of instruction of these two subjects is, on average, 10 hours.) Teachers are assigned to schools by public authorities and headmasters have little-to-no control over their hiring, firing, and salaries. However, they do have the ability to assign teachers to classes. While lacking information on the exact procedure for assignment of teachers to classes makes identification of these effects less clean than those from the peers’, the large homogeneity of classes within schools documented above makes systematic patterns of teacher-student characteristics unlikely. This is supported by the rich balance tests in Tables A.6 and A.7, as discussed in Section 3.2.

In Table 5, I proceed by augmenting my baseline specifications with a dummy taking the value of one if the homeroom teacher is a female — while also controlling for the highest education achieved by the teacher, years of tenure, indicators for having a temporary contract or a secondary job, and a wealth index. I find that, while exposure to a female teacher constitutes an additional force in decreasing the degree of traditionalism among male children (Panel A), for female children (Panel B) the effect is, if anything, the opposite: female teachers show signs of fostering traditionalism. In both cases, the two dimensions most affected are life purpose and intra-household decisions, as one would expect from teachers’ acting as role models.

It is worth further exploring the somewhat detrimental effects of female teachers on female

⁵⁴More generally, under the set of moral principles known as “Three Obediences and Four Virtues”, women are expected to subordinate to males at all 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). Moreover, women’s virtues consist of working diligently, and displaying modest manners, proper speech, and morality.

⁵⁵Past 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).

students' gender views. While at first sight this may seem counterintuitive, one should note that teaching primary students is a traditionally female-dominated occupation (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).

In Panel C, I provide suggestive evidence that this is — at least partly — the reason behind the above findings. In particular, the negative sign in the interaction term indicates that, among female teachers, those who have a higher perceived locus-of-control for affecting students' outcomes are less detrimental for gender attitudes. This is in line with Carlana (2019)'s findings that teachers with weaker gender stereotypes help make attitudes more gender progressive, and highlights the potential of policies that (a) foster self-confidence and more modern views among teachers and/or (b) decrease students' vulnerability to traditional attitudes.

4.4 Nonlinearities

So far, I have estimated an average marginal effect common to all increases in the proportion of females, as standard in the literature. However, conceptually, one would expect that a given increase in the proportion of female peers will not have the same effect at different points in the distribution (e.g., when starting from 30% instead of from 60%). This is also relevant from a policy perspective, as it will lead to different estimates of the potential benefits of an intervention.

Table 6 builds upon the baseline analyses and first replaces 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 female peers reduce their traditionalism significantly more than the omitted category (i.e., individuals in classes where females are the minority). 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 estimated effect for group (iii) is -0.266 points — larger than that for any of the other two categories. The p-value of the F-test of equality of the estimates for categories (i) and (iii) is 0.007.

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

	Life Purpose	Abilities	Cross-Gender Interactions	Intra-household Decisions
Panel A: Full Sample				
	(1)	(2)	(3)	(4)
Proportion Female Peers	-1.442*** (0.460)	-1.322** (0.559)	-2.429*** (0.626)	-0.787 (0.633)
Observations	865	865	865	865
R-squared	0.253	0.226	0.209	0.162
Panel B: Females				
	(9)	(10)	(11)	(12)
Proportion Female Peers	-1.753** (0.807)	-2.373** (1.085)	-2.504*** (0.915)	-1.696 (1.239)
Observations	418	418	418	418
R-squared	0.321	0.304	0.241	0.295
Panel C: Males				
	(5)	(6)	(7)	(8)
Proportion Female Peers	-1.383** (0.573)	-0.287 (0.841)	-2.450** (0.938)	0.572 (1.275)
Observations	430	430	430	430
R-squared	0.292	0.305	0.306	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 a given attitudinal category from the inverse covariance matrix of all relevant items. The final index is again normalized to have a mean of 0 and a standard deviation of 1. Unlike in Table 3, I restrict the sample to individuals with available information on *all* attitudinal items, which explains the difference in sample size. Standard errors clustered at the class level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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 within the male and female subsamples (columns 5 and 6).

Overall, the findings of this exercise indicate that female-dominated classes are particularly conducive to attitudinal changes. While my ability to make claims about the marginal impacts at the extreme of the distribution is limited by the fact that class compositions with a very

Table 5: The Impact of Teacher Sex on Broad Categories of Gender Attitudes (Age 15)

	Life Purpose	Abilities	Cross-Gender Interactions	Intra-household Decisions
Panel A: Males				
	(1)	(2)	(3)	(4)
Female Teacher	-0.553*** (0.203)	-0.155 (0.198)	0.116 (0.245)	-0.701*** (0.199)
Observations	420	420	420	420
R-squared	0.310	0.326	0.312	0.223
Panel B: Females				
	(5)	(6)	(7)	(8)
Female Teacher	0.341 (0.210)	0.126 (0.288)	0.451 (0.300)	0.467* (0.239)
Observations	408	408	408	408
R-squared	0.339	0.324	0.273	0.324
Panel C: Females' Heterogeneity by Teacher's Locus-of-control				
	(9)	(10)	(11)	(12)
Female Teacher	0.479** (0.224)	0.243 (0.326)	0.765** (0.381)	0.723** (0.293)
Teacher believes his/her job can influence kids' lives	0.183 (0.141)	0.205 (0.188)	0.304 (0.219)	0.290 (0.178)
Female Teacher*Job Can Influence	-0.277** (0.134)	-0.156 (0.182)	-0.510** (0.216)	-0.363** (0.180)
Observations	401	401	401	401
R-squared	0.350	0.326	0.290	0.326

Notes. Replication of the specifications in Table 4 with the inclusion of teacher controls (highest education, tenure, indicators for having a temporary contract or a secondary job, and wealth index). “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 estimates for the proportion of female peers remain significant and are omitted for brevity. Standard errors clustered at the class level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

high or very low proportion of females are infrequent empirically due to the as good as random allocation of students into classes, it is worth noting that these results are consistent with existing evidence arguing that female-dominated and fully-female environments are most effective in improving girls’ outcomes (e.g., [Shan, 2021](#)).

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 for the proportion of female students 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 limits 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 on gender norms is indeed class-driven and not cohort-driven. In Section 6, I offer a more detailed discussion on how, given my main results, one could think of feasible policy interventions that may successfully impact outcomes.

Table 6: Nonlinear Effects of the Proportion of Female Peers on Views on Gender Norms

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean Agreement with Traditional Gender Norms (Age 15)					
Proportion Female Peers (> 50%)	-0.291*** (0.077)					
Proportion Female Peers (0 – 45%)		0.055 (0.088)	0.002 (0.095)	0.005 (0.095)	-0.031 (0.149)	0.033 (0.138)
Proportion Female Peers (> 55%)		-0.266** (0.106)	-0.254** (0.105)	-0.239** (0.109)	-0.484*** (0.148)	-0.370* (0.217)
Sample	Full	Full	35% ≤ Prop. Fem.	35% ≤ Prop. Fem. ≤ 65%	Males	Females
Observations	880	880	791	738	433	431
R-squared	0.289	0.286	0.305	0.310	0.334	0.285

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 2-6 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 also 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. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.5 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, at least partly, decide upon. This might not happen if, for instance, deviating from traditional gender norms is too costly even after attitudinal changes have taken place. 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.⁵⁶

4.5.1 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 7 shows that there is none by age 15, a time when, while education is no longer compulsory (it finishes at Grade 9, i.e., age 14), the dropout rate is still relatively low (about 20%). A key strength of my data is its long time series component, which allows me to explore whether effects do show up in the long-run, when dropout rates increase and children are likely to make more decisions on their own.

For this, I take advantage of an ad hoc round of YL conducted in 2020 with the aim of collecting information on how the Covid-19 pandemic disrupted the lives of the participants. While this round mostly elicits information on Covid-related aspects, it also has information on pre-pandemic 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 pandemic is over — at the time that the children are already 19 years of age. Column 3 shows that girls who were more exposed to other females are significantly more likely to remain at school at age 19. In particular, a 10 percentage point increase in the proportion of female classmates is associated with an increase in 10 percentage points in the probability of remaining enrolled (in a context where 52% of the children are still enrolled, generally at university⁵⁷), while no effects are found for males (column 4).⁵⁸ What it is more, in column 5 I explore whether these girls are less likely to undertake a major in which females are traditionally more inclined to enrol. This is indeed the case.⁵⁹ These findings are relevant given

⁵⁶Other 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.

⁵⁷Official statistics from the Ministry of Education and Training confirm that there is close to parity in access to university across genders.

⁵⁸Because YL asks retrospective questions about school enrollment status by year, it is actually possible to explore the exact timing when 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.

⁵⁹I consider male-dominated majors to be: mathematics and statistics; computer and communications sciences; construc-

i) the inequalities in access to tertiary education across genders of otherwise observationally-equivalent individuals (Mergoupis et al., 2018); ii) that female minority status at traditionally male-dominated majors has been shown to be a cause of higher female dropout (Shan, 2021), and iii) that selection into certain sectors is behind a significant fraction of the gender wage gap in Vietnam (Chowdhury et al., 2018).⁶⁰

Table 7: Is Long-run School Enrollment Affected by Peers' Female Composition?

	(1)	(2)	(3)	(4)	(5)
	Currently Enrolled Age 12	Currently Enrolled Age 15	Currently Enrolled Age 19	Currently Enrolled Age 19	Non-male-dominated Major (Age 19)
Proportion Female Peers	-0.012 (0.016)	-0.120 (0.300)	0.990** (0.390)	-0.398 (0.613)	-0.522** (0.214)
Sample	All	All	Females	Males	Females at University
Mean (sd) Dep. Var.	0.997 (0.056)	0.82 (0.39)	0.52 (0.50)	0.50 (0.50)	0.92 (0.28)
Observations	878	878	408	378	192
R-squared	0.139	0.219	0.366	0.388	0.384

Notes. Replication of the specification in Table 3's column 7 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. Column 5 only employs the subset of the 408 females who are enrolled at university and for whom we observe the academic major. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.5.2 Time Use: Males' Home Production

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

In Table 8, I focus on the male subsample, whose reduced traditionalism has the potential to make them more willing to contribute time to chores⁶¹. Column 1 shows that being exposed to a higher fraction of female peers in Grade 5 leads to a larger amount of hours being 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,

tion and architecture engineering; other technologies; agriculture, forestry and aquaculture; transports; food processing; and armed force and police.

⁶⁰I 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 and low ability females benefit from increased exposure, the effect may be somewhat larger among those with a lower initial cognitive score (results available upon request).

⁶¹A 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 household head.

Table 8: Effects of the Proportion of Female Peers on the Contribution to Home Production

	Age 12			Age 15		
	(1)	(2)	(3)	(4)	(5)	(6)
	Total Home Hours	Any Home Hours	Over 3 Home Hours	Total Home Hours	Any Home Hours	Over 3 Home Hours
Proportion Female Peers	1.088 (1.373)	3.701** (1.649)	0.894* (0.454)	0.846** (0.347)	0.162 (0.433)	0.719** (0.335)
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	0.353	0.257	0.261	0.334	0.284

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 (fetching 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

respectively.

Returning to the same outcomes in round 5 (when children likely have more saying in this decision and provide more home production on average) confirms the higher contributions to home production along the intensive margin.⁶² The economic magnitude is both reasonable and nontrivial. For instance, column 6 implies that a ten percentage point increase in the proportion of female peers leads to an increase of seven percentage points in the probability of the child spending at least three hours per day on home production (about a fifth of the standard deviation). In terms of nonlinearities, Table A.12 shows that the effects are significantly larger among individuals exposed to more than 50% of female peers.

⁶²A regression of time allocation in round 3, i.e., prior to the School Survey, using the same specification as in the Table 8 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 0.17 (p-value of 0.37) when the outcome is an indicator for providing a positive number of home or care hours and it is -0.05 (p-value 0.88) when the outcome is an indicator for contributing over 2 hours per day.

4.5.3 Cognitive Outcomes

In contrast to 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.⁶³ Table 9 explores the effects on cognition and on related non-cognitive outcomes. The latter are interesting both in themselves and also to better understand potential mechanisms.⁶⁴ Given that in this exercise I am purely interested in those changes taking place in the short-run (i.e., during the academic year of the School Survey), I observe all the necessary information for every student present in the School Survey so that I no longer need to impose that the child belongs to the YL's longitudinal study. This allows me to gain statistical power (results are similar when using exclusively YL children, as in my main analysis). 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 interpret the results as *gains* since 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 school⁶⁵, has previously been found in the literature (e.g., [Gong et al., 2019](#)).⁶⁶

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 survey round. This is an attractive feature since it provides direct measures of actual cognitive ability, rather than having to resort to proxying for the effects on cognition through grade completion or school enrollment, which are noisy measures, particularly in de-

⁶³In Appendix C.5 I document the effects on fertility preferences.

⁶⁴Having said this, it is an almost impossible endeavour to attempt to explore all possible channels. Indeed, an important candidate, namely improved class atmosphere, cannot be studied as measures for this dimension are unavailable to me.

⁶⁵In 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.

⁶⁶Replacing 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.) 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). 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.

Table 9: 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	216.831** (100.750)	122.178 (127.243)	-1.785*** (0.670)	-1.851** (0.934)	-0.126 (0.966)
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	0.447	0.179	0.171	
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	145.043 (99.908)	42.530 (118.404)	-0.716 (0.649)	-0.416 (1.198)	0.265 (0.946)
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	0.417	0.498	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, which explains the increase in sample size. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

veloping countries (Singh, 2020). Crucially, these cognitive scores are available for every YL student, *irrespective* of her school enrollment status.

Table 10 shows that there is little evidence in favor of increased cognitive performance in the middle-run, and the point estimates even turn negative in some cases — albeit they are not significant. This is consistent with ample 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 Rate

Given YL’s aim to track children’s development, 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’ lower dropout rate in a comprehensive manner.

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.13 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.⁶⁷

Perceived Returns to Education. Another potential — and complementary — channel is that this increased academic ambition is closely related with a shift in the expected returns to education. If children believe — accurately or not — that returns to education are higher, they will be more likely to remain at school (Cunha and Heckman, 2007; Jensen, 2010; Attanasio and Kaufmann, 2014).⁶⁸

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

⁶⁷Finding 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).

⁶⁸Indeed, the connection between expected returns and enrollment among the Young Lives’ Peruvian young cohort has been recently shown by Favara et al. (2021).

at least [midpoint between max and min level]? Under the standard assumption of a perceived triangular distribution of wages, 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.14, I show that females who interact with a higher share 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 the effects are smaller 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. The estimated coefficient for the variance is small and not significant. This can explain why, while both the expected mean and variance increased in column 1 (albeit the latter not significantly), 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, also provide a mechanism rationalizing both why females that are more exposed to female peers are more prone to still be enrolled by age 19.⁶⁹

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 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 traditionally female-dominated — i.e., it takes value 1 if the desired profession belonged to (ii) or (iii).⁷⁰ Note that this variable is different from that in Table 7, where I used information on *actual* college major choices. Table A.15 shows suggestive evidence that while girls that were majority in the class are more likely to wish to have less traditional female jobs (using the binary indicator in column 3⁷¹; 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 *occupation* to which female students

⁶⁹It 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).

⁷⁰My 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.

⁷¹Unreported, I find that this effect is mostly driven by category (ii) jobs — females become more likely to belong to this category.

aspire.

Increased Academic Self-confidence. In Table A.16’s column 2, I show that, while short-term gains in self-confidence were absent among girls, by round 5 this is no longer the case and the differences are strongly significant. This not only highlights the importance of tracking outcomes over time to better understand their dynamics and eventual effects, but also provides another potential channel for why females increased their school enrollment in the long-run.⁷²

Mediation Analysis. The above discussion has identified multiple potential channels at work: (i) fall in the degree of acceptance of traditional career goals for females; (ii) reduced perceptions of females being less able than males; (iii) increased aspirations to reach university; (iv) shifts in expected returns to university: higher mean and lower variance of expected earnings; (v) increased confidence. All of these mechanisms are consistent both with female peers’ being more encouraging towards less traditional careers and male peers exerting less stereotypical views.⁷³ One would be interested in understanding how the estimate of the proportion of female peers in the base model (i.e., column 3 in Table 7) changes as all these mediating factors are introduced linearly as controls (the “full model”, Equation 2). I call this difference θ .

$$Enrolled_{ics2020} = \gamma_0 + \alpha_M PF_{-ics} + \sum_J \alpha^j p_{ics}^j + \gamma'_1 X_{ics} + \lambda_s + \epsilon_{ics}, \quad (2)$$

where I denote each of the mechanisms by “ p ”, and $j = \{\text{life goals, perceived relative abilities, aspirations for university, mean expected income from university, variance of expected income from university, confidence}\}$.

Gelbach (2016) provides a decomposition of θ into the relative importance of each mechanism 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

⁷²Why 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.

⁷³As 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, this result is not surprising because the three mechanisms were little impacted 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 for 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 10: Effects of the Proportion of Female Peers on Long-term Academic Outcomes

Panel A: Males				
	(1)	(2)	(3)	(4)
	Vietnamese	Math	Vietnamese	Math
	Age 12	Age 12	Age 15	Age 15
Proportion Female Peers	46.938 (82.079)	92.004 (72.433)	-47.670 (84.863)	-27.486 (88.577)
Observations	428	428	429	428
R-squared	0.424	0.509	0.403	0.441
Panel B: Females				
	(1)	(2)	(3)	(4)
	Vietnamese	Math	Vietnamese	Math
	Age 12	Age 12	Age 15	Age 15
Proportion Female Peers	49.206 (116.823)	-13.226 (88.496)	-24.757 (94.713)	92.539 (95.292)
Observations	419	419	429	427
R-squared	0.360	0.507	0.367	0.457

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

a time, the six main mechanisms found above (I consider the mean and the variance in (iv) separately) on the leave-out mean of the proportion of female students in the class, together with the baseline controls. Such regressions take the following form:

$$p_{ics}^j = \beta_0 + \beta_j PF_{-ics} + \beta'_2 X_{ics} + \lambda_s + \epsilon_{ics}. \quad (3)$$

This yields, for each p^j , an estimate ($\hat{\beta}_j$) of how the proportion of female classmates shifts each of these mediators.

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 between the base and the full model ($\hat{\theta} \approx 0.86$) by computing, separately for each variable j , the following ratio: $RI_j = \frac{\hat{\alpha}^j \hat{\beta}_j}{\hat{\theta}}$, i.e., the fraction of the explained treatment effect of female peers that is driven by each mediator.

Table 11 shows that, based on the estimated $\hat{\theta}$, the six channels considered are able to explain about 90% of the full effect of female peers on enrollment.⁷⁴ While estimates from a mediation analysis should generally be interpreted with caution⁷⁵, increased enrollment seems to be mostly driven by higher aspirations for university (35% of the total effect) and expected earnings (25%). These are self-centered characteristics. Additionally, the reduction in traditionalism with respect to acceptable life goals for females in the society also explains 21% of the total effect.⁷⁶ It is noteworthy that self-confidence — the common explanation alluded to explain females' reduced enrollment or cognitive results upon exposure to high proportion of males — seems to be less relevant in accounting for the gap in enrollment rates when all the potential mechanisms are jointly considered. 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).

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 on 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 genders. To validate this claim, I exploit the school survey's friendship nominations for every

⁷⁴0.859/1.016, where 1.016 is the estimate of the proportion of females in the base model (equation 3).

⁷⁵Estimates would be biased if one were to think that the mechanisms included in the decomposition 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 about 90% of the total effect should reduce this concern.

⁷⁶The coefficients reported in the table should be interpreted as the relative importance of a mechanism, holding all other mediators fixed. One could therefore have modelled our framework in a way that some of the mechanisms interact with or determine another.

Table 11: Mediation Analysis: Breakdown of Relative Importance of Channels in Decreased Female Dropout

Mediator	(1) Coefficient	(2) Standard Error
Aspires to University	0.347***	0.104
Expected Earnings	0.241*	0.146
Traditional Life Goals	0.214*	0.127
Self-confidence	0.078	0.095
Traditional Female Abilities	-0.003	0.038
Variance Earnings	-0.018	0.043
Total Explained ($\hat{\theta}$)	0.859***	0.224

Notes. Gelbach (2016)'s decomposition of the role of the various mechanisms for explaining the gap in enrollment rates across female children exposed to higher and lower proportions of female peers. "Total Explained" represents the difference in the estimate of the variable "proportion of female peers" between the full and the base models. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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 disproportionately more likely to form friendship groups with a higher fraction of females?

In Table 12, 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 close friends; (c) close friends; (d) very close friends, and construct three indicators that are increasing in the strength of the friendship.⁷⁷ 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 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 a value of 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').

⁷⁷For 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 “beyond-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.4⁷⁸ I argue that, under reasonable assumptions, these estimates suggest that increasing the fraction of females classmates raises the share 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.⁷⁹

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 interactions 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 12 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.6, I offer additional evidence consistent with the above discussion. In particular, I: (i) show that being surrounded by objectively more able female peers’ — as

⁷⁸This section also provides a parallel analysis for females, yielding similar qualitative results.

⁷⁹This is confirmed through an alternative specification in which an indicator for a given peer being classified as a friend is regressed on the gender of that peer and on individual fixed effects of the child of reference. Note that for this exercise I create a pseudo-panel where each child of reference is observed as many times as peers were sampled in his/her class. I find that males in classes with more than 50% of females are over 4 percentage points more likely to become friends with girls.

Table 12: 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	0.930** (0.387)	0.025 (0.262)	-0.250 (0.337)
Mean (sd) Dep. Var.	0.35 (0.19)	0.21 (0.21)	0.12 (0.23)
Observations	428	423	368
R-squared	0.495	0.455	0.258
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	0.745*** (0.268)	0.002 (0.300)	-0.139 (0.283)
Mean (sd) Dep. Var.	0.33 (0.21)	0.20 (0.24)	0.15 (0.26)
Observations	420	413	343
R-squared	0.427	0.394	0.309
Panel C: Combined Measures			
	(7)	(8)	(9)
Female Friends/Total Friends defined as...	1+4	2+5	3+6
Proportion Female Peers	0.764** (0.310)	-0.149 (0.208)	-0.163 (0.280)
Mean (sd) Dep. Var.	0.30 (0.22)	0.16 (0.23)	0.10 (0.22)
Observations	420	405	296
R-squared	0.443	0.343	0.290

Notes. All regressions replicate the specification in Table 3's column 7 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 jointly satisfied. 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 for a given category, regardless of the gender, 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 of males’ shifts towards less traditional views on female abilities. This is in line with a revision upon exposure of pre-conceived negative views on female ability, and (ii) investigate in more detail *who* experiences larger treatment effects. I find that the main source of heterogeneity is pre-exposure traditionalism at an aggregate level (measured as the overall traditionalism in the community of residence as reported by the old cohort). I find that the largest effects are found in more traditional areas. A more nuanced exercise through the estimation of quantile regressions in Table A.21 uncovers that the impact is, however, not uniformly distributed, and the more traditional part of the support of the distribution is not affected by the share of female classmates. This suggests that, while there is room to change attitudes in more traditional areas, impacting extremely traditional individual views might be harder to achieve through mere classroom exposure.

6 Room for Policy

Is there room for effective policy interventions aiming at changing gender attitudes? Absent the results from above, one may be tempted to think that the answer is no: although the status quo is the universal presence of coeducation in primary education and only 2% of male children report not having any female friends, gender attitudes are still strongly traditional. The presence of nonlinearities in Table 6 is therefore an encouraging finding to inform class formation policies. In particular, it suggests that it is possible to reduce *overall* traditionalism in the society by reallocating female students from male-dominated and/or gender parity classes to female dominated ones. Given that such action would lead to losers (i.e., students in the class from which the female student gets removed), policy makers might instead be more interested in exploring the possibility of yielding *Pareto improvements*. For this to be feasible, it would be necessary to find “flat” regions from which a girl could be removed without any harm to the students that remain in the class, and be placed in a class where an extra girl is particularly effective in changing attitudes. In Figure B.13, I estimate linear splines to show evidence suggesting the presence of such flat part in gender parity classes (those with 40-55% of females).⁸⁰

With these considerations in mind, the objective of this section is to quantify whether potential Pareto gains would be sizable in practice based on my non-linear estimates. To carry out such exercise, I impose the following requirements for a potential policy revolving around class composition to be operational: (i) it should *only* rely on the *actual* school gender composition, i.e., it should not involve attracting new students nor forcefully displacing them

⁸⁰A more detailed exploration with larger sample size would be crucial to better inform this discussion.

into a school, (ii) it should only rely on shifting girls whose original class composition lied in the flat portion found in Figure B.13, 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).⁸¹

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 6 to shift female students from the flat part to female-dominated classes. 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 allocated to 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, equalize the number of girls across all sections.

2. Prediction of counterfactual gender norms

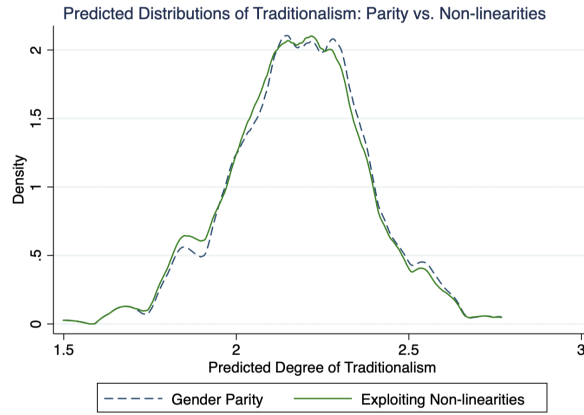
- Based on the estimates of my nonlinear model (with three categories for the proportion of females) in Table 6, use the proportion of female classmates generated

⁸¹For instance, an interesting avenue to consider is whether exposure to more able females could have a larger impact than peers’ gender per se. However, this is beyond the scope of my study, as academic ability could be an endogenous outcome itself.

under the two counterfactual scenarios above to obtain the corresponding two values 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 4 that, indeed, the main gains under Counterfactual 2 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.⁸² Kolmogorov-Smirnov’s test of equality of the two distributions rejects the null at the 4% confidence level. These effects are particularly meaningful given that recent evidence suggests that convergence towards 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 4: Distribution of Predicted Outcomes Under the Two Counterfactual Scenarios



Notes. Distribution of predicted gender norms using the non-linear estimates from Table 6 under two scenarios: attaining gender parity across classes (blue, discontinuous curve) and exploiting non-linearities (green curve).

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 Law dictates that children should

⁸²One way of thinking about the validity of the model proposed is by looking at the R^2 in the baseline nonlinear specification. It can be seen from Table 6 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.

enroll in primary education in the calendar year that they turn 6 years of age. This 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.⁸³ In [Figure 5](#), one can appreciate that, while we do have the expected discontinuity in educational achievement (individuals to the right of the threshold, who are born after 1 January, have about 0.7 years of education less), this does not translate into different support towards traditional gender roles.⁸⁴ It is important for interpretation to note that the returns that I am estimating are local average treatment effects⁸⁵ of one extra year of education between grades 8 and 9, which could be different from those at other stages in life. 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). This is consistent with the documented persistence of traditional gender views even as Vietnam rapidly progressed along the economic development path in recent years.

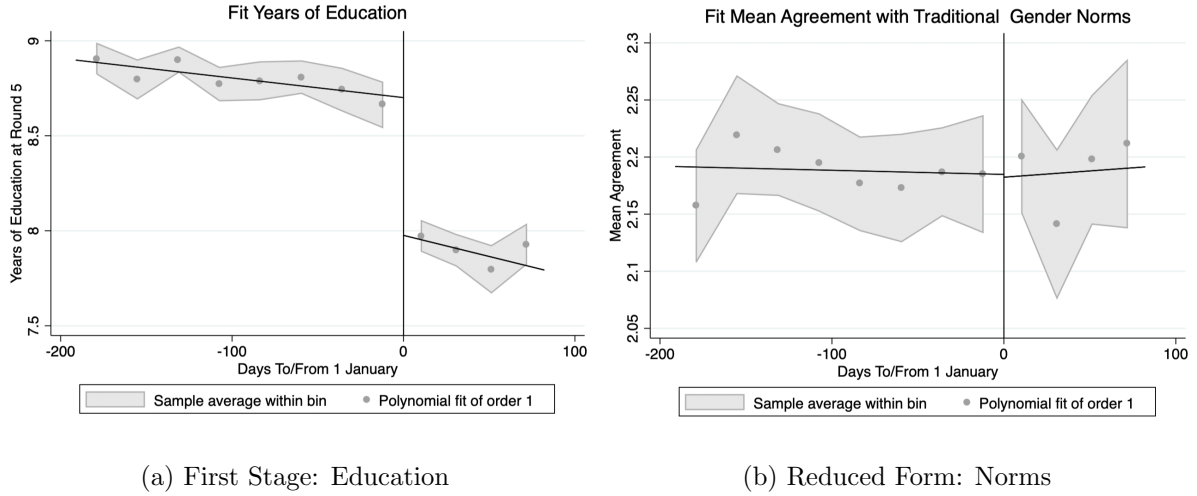
Comparison with the Role of Female Siblings. Apart from peers, another source of early socialization with females is through sisters. In order to provide an even more complete picture of the relative size of my main estimates, I compare them with the impact of female siblings. To deal with the potential endogeneity arising from unobserved time-invariant heterogeneity in son preference present in Confucianism-related cultures, I obtain 2SLS estimates of the effect of having at least one sister on gender norms where I instrument the presence of a sister 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, and given the preference towards sons prevailing in Vietnam and affecting fertility decisions ([Haughton and Haughton, 1995](#); [Pham et al., 2012](#)), the likelihood that individuals have a female sibling is lower as the parental incentives for having more children fall. I find a precisely

⁸³He provides local density plots for the date of birth around 1 January ([McCrary, 2008](#)) and formally tests for a break in the density ([Cattaneo et al., 2018](#)). I replicate these analyses in my sample and further show that other observables do not jump discontinuously at the threshold (available upon request).

⁸⁴More generally, this analysis, together with the reduced form study of the impact on males' home production and on the perceptions about 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 disproportionately more likely to leave formal education in order to care for others, and (c) unfair treatment in the labor market.

⁸⁵Plotting the CDFs of years of education achieved by round 5 separately for children born before and after 1 January shows a clear jump in years of schooling between 8 and 9, while the functions do not cross. This supports the monotonicity assumption behind the interpretation of the estimates as local average treatment effects.

Figure 5: No Returns to Education on Gender Norms



Notes. Panel (a) depicts the first stage of distance of date of birth to/from the 1 January cutoff on years of education achieved by the fifth round of YL. Panel (b) documents the reduced form effect of the running variable on gender attitudes (composite measure in the 1–4 scale).

estimated zero impact of female siblings.⁸⁶ While the exclusion restriction might not hold, this is a useful scaling exercise that 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 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.

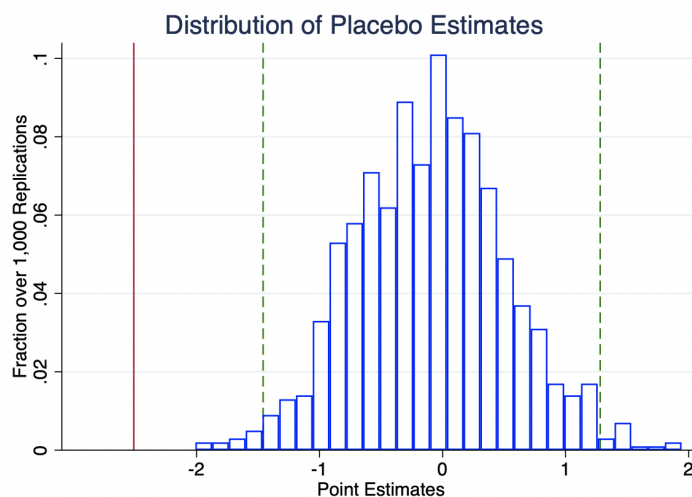
Randomization Based Inference: Placebo Peer Allocation. In order to verify that the effects uncovered are not driven by unobserved student or school characteristics, I conduct a falsification test based on simulating random allocations of students to classes within schools, without replacement, and maintaining the original size of the classes (Olivetti et al., 2020). I then run the specification in Table 3’s column 7 (with the remaining controls fixed at their true values) and replicate this exercise 1,000 times. A strength of this exercise is that it allows the researcher to be agnostic about asymptotic distributions and instead empirically construct the nonparametric distributions of test statistics via replication of the randomization procedure (Imbens and Wooldridge, 2009).

⁸⁶As expected, the first stage is very strong. Having a first-born male is associated with a large fall in the probability of having a daughter, with an F-statistic of 1224.66 and a p-value less than 0.0001.

Reassuringly, the distribution of estimated treatment effects reported in Figure 6 shows that the baseline estimate of -2.512 (signaled by the red line) lies outside the 95% confidence interval obtained from the simulations, which is [-1.459, 1.281] — as indicated by the dashed green lines — and centered around 0.⁸⁷ Conducting similar exercises for other major outcomes explored in the paper (e.g., male’s home production) provides further support to my main results and are available upon request.

Additionally, in Figure B.14 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. Results under this scenario mirror those in Figure 6.

Figure 6: Distribution of Placebo-Generated Estimates of the Proportion of Female Peers on Gender Norms (empirical 95% CI indicated by green lines)



Notes. Distribution of point estimates from the share of female peers in a regression of the form of Table 3’s column 7, where the class composition is generated randomly among the actual students in the school (1,000 repetitions). The green dashed lines indicate the empirical 95% confidence interval while the red line indicates the actual point estimate obtained in Table 3’s column 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

⁸⁷Complementary, 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.

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).⁸⁸ 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 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 $[-2.512, -1.283]$, which does not contain zero.⁸⁹

Accounting for Influential Observations. Contemporaneous work by [Broderick et al. \(2020\)](#) highlights the risk that estimates obtained even in ex ante very clean empirical setups (e.g., successful RCTs) are driven by as few as one observation. I apply their proposed metric to find the set of observations that is 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 result⁹⁰). Hence, this constitutes strong supportive evidence in favor of my findings⁹¹.

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 inbetween the school survey and the time when I measure gender-related views (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

⁸⁸She 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.

⁸⁹Note also that the bounds are contained within 2.8 times the standard error of the controlled estimate (i.e., 0.492), which is an alternative measure of coefficient stability proposed by [Oster \(2019\)](#).

⁹⁰Indeed, 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.

⁹¹In particular, [Broderick et al. \(2020\)](#) emphasize the strength of their approach in acting as a robustness check for inference (i.e., for the use/treatment of standard errors and their ability to account for non-random sampling).

round, and only 1 out of the 17 was part of my estimating sample for the main analyses.

Table A.20 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 is 0.631).⁹²

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 than other closer countries such as Thailand or Malaysia. This would make one think that if attitudes can be changed in Vietnam, there should also be room for changes in these other countries. 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.15 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 different cultures and institutional systems.

Additional Robustness Checks. In Online Appendix section C.7, I show that the main 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 schools for which the Fisher's exact test rejected independence of observable characteristics across

⁹²Another 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'Haultfoeuille et al. (2018) yields consistent results to those in the baseline model.

classes; (iii) further accounting for potential selection into schools (beyond school fixed effects) through (a) re-estimating the baseline equation for the subsample of schools with and without other primary schools located nearby that could potentially attract the families; (b) through the inclusion of spatially more aggregate fixed effects (community-level); and (c) through accounting for the very small subset of children that ever transfer schools during primary education.

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 socialization literature that sustained exposure to females in contexts of already frequent cross-gender interactions is capable of leading to more modern gender views in the long-run, and I investigate the mechanisms at place.

My focus is on Vietnam, a country with marked gender gaps and strong traditional gender roles relative to Western standards, 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. A rich battery of tests lends 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 20% 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 a relevant 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. As

more data becomes available in the next years, I will enrich the analysis by exploring family formation outcomes, including marriage matching and timing, childbearing, and intra-household decisions.⁹³

Overall, my findings provide a rich picture of the role of cross-gender interactions and integration, both in the scope of outcomes explored and in the extended time-frame along which they are observed. These two dimensions are needed in order to better inform policy, not only in developing but also in developed countries, where inequality in access to opportunities by gender and the appropriateness of single-sex schools is still hotly debated. For instance, going beyond exclusively accounting for academic spillovers highlights the pressing limitations that disregarding socio-emotional gains arising from certain class configurations has. 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). The fact that improved female conditions does not have side effects among males (e.g., in terms of cognitive outcomes, self-confidence, etc.) is an important reason helping to justify the manipulation of class composition as a plausible policy intervention.

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 is 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.

⁹³Indeed, regressions for the old cohort controlling for gender, wealth index, parental and own education, ethnicity, household size, caregiver's gender and age, and community fixed effects show that individuals with more traditional gender norms are significantly more likely to be married and to have children by age 21.

References

- Akerlof, George A and Rachel E Kranton**, “Economics and identity,” *The Quarterly Journal of Economics*, 2000, *115* (3), 715–753. (return to page [1](#))
- 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 - 41](#))
- Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew**, “The nature of prejudice,” 1954. (return to page [1](#))
- 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 [21](#), [51](#))
- 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 [25](#))
- 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 [27](#))
- 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 - 41](#))
- , **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 [1](#), [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 [50](#))
- 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 [4](#), [37](#))
- Autor, David H**, “Skills, education, and the rise of earnings inequality among the “other 99 percent”,” *Science*, 2014, *344* (6186), 843–851. (return to page [8](#))
- 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 [1](#), [54](#))
- Beaman, Lori, Raghavendra 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 [1](#))
- 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 - 42](#))
- Behr, Helen**, “Comparing rural and urban primary education in the Mekong Delta,” 2005. (return to page [10](#))
- 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 [25](#))
- Bertrand, Marianne**, “The gender socialization of children growing up in nontraditional families,” in “AEA Papers and Proceedings,” Vol. 109 2019, pp. 115–21. (return to page [1](#))
- , **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 [1](#), [13](#))
- , **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 [1](#))
- 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 [27](#))

- 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 8)
- 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 3, 6, 23, 37)
- 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 - 36)
- Bordalo, Pedro, Katherine Coffman, Nicola Genaioli, and Andrei Shleifer**, “Stereotypes,” *The Quarterly Journal of Economics*, 2016, 131 (4), 1753–1794. (return to page 1)
- 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 6, 37, Appendix - 42)
- 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 19, 51)
- 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 28)
- Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang**, “Curriculum and ideology,” *Journal of Political Economy*, 2017, 125 (2), 338–392. (return to page 8)
- Carlana, Michela**, “Implicit stereotypes: Evidence from teachers’ gender bias,” *The Quarterly Journal of Economics*, 2019, 134 (3), 1163–1224. (return to page 10, 28, Appendix - 45)
- 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, 46)
- , **Mark Hoekstra, and Elira Kuka**, “The long-run effects of disruptive peers,” *American Economic Review*, 2018, 108 (11), 3377–3415. (return to page 6, 37)
- Cattaneo, Matias D, Michael Jansson, and Xinwei Ma**, “Manipulation testing based on density discontinuity,” *The Stata Journal*, 2018, 18 (1), 234–261. (return to page 48)
- 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 - 39)
- 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)
- Chowdhury, Iffat, Elizaveta Perova, Aneesh Mananna, and Hillary C Johnson**, “Gender gap in earnings in Vietnam: why do Vietnamese women work in lower paid occupations?,” *World Bank Policy Research Working Paper*, 2018, (8433). (return to page 7, 33)
- Chung, Bobby W**, “Peers’ parents and educational attainment: The exposure effect,” *Labour Economics*, 2020, 64, 101812. (return to page Appendix - 46)
- 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 4)
- Cools, Angela, Raquel Fernández, and Eleonora Patacchini**, “Girls, boys, and high achievers,” *Technical Report, National Bureau of Economic Research* 2019. (return to page 3, 8)
- Crowne, Douglas P and David Marlowe**, “A new scale of social desirability independent of psychopathology.,” *Journal of consulting psychology*, 1960, 24 (4), 349. (return to page 13)
- Cunha, Flavio and James J Heckman**, “Identifying and estimating the distributions of ex post and ex ante returns to schooling,” *Labour Economics*, 2007, 14 (6), 870–893. (return to page 37)
- Dahl, Gordon B, Andreas Kotsadam, and Dan-Olof Rooth**, “Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams,” *The Quarterly Journal of Economics*, 2021, 136 (2), 987–1030. (return to page 1, 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 8, 18)
- 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 [27](#))
- Dhar, Diva, Tarun Jain, and Seema Jayachandran**, “Reshaping adolescents’ gender attitudes: Evidence from a school-based experiment in India,” Technical Report 3 2022. (return to page [3](#), [7](#), [8](#))
- Doepke, Matthias and Michèle Tertilt**, “Families in macroeconomics,” in “Handbook of Macroeconomics,” Vol. 2, Elsevier, 2016, pp. 1789–1891. (return to page [Appendix - 41](#))
- , **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 [54](#))
- 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 [54](#))
- D’Haultfoeuille, 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 [52](#))
- Favara, Marta, Paul Glewwe, Catherine Porter, and Alan Sanchez**, “Expecting Better? How Young People Form Their Earnings Expectations,” 2021. (return to page [37](#))
- 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 [14](#), [Appendix - 39](#))
- 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 [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 [51](#))
- Fruehwirth, Jane Cooley and Jessica Gagete-Miranda**, “Your peers’ parents: Spillovers from parental education,” *Economics of Education Review*, 2019, 73, 101910. (return to page [22](#), [Appendix - 36](#))
- 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 [27](#))
- 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](#), [39](#), [42](#))
- 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 [10](#))
- , —, and —, “Gender peer effects on students’ academic and noncognitive outcomes: Evidence and mechanisms,” *Journal of Human Resources*, 2019, pp. 0918–9736R2. (return to page [1](#), [3](#), [35](#))
- Goussé, Marion, Nicolas Jacquemet, and Jean-Marc Robin**, “Marriage, labor supply, and home production,” *Econometrica*, 2017, 85 (6), 1873–1919. (return to page [13](#))
- Guilmoto, Christophe**, “Recent change in the sex ratio at birth in Viet Nam: a review of evidence,” 2009. (return to page [14](#))
- 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 [4](#))
- 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 - 39](#))
- Haughton, Jonathan and Dominique Haughton**, “Son preference in Vietnam,” *Studies in Family Planning*, 1995, pp. 325–337. (return to page [48](#))
- 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 [54](#))
- 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 [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 [5](#), [19](#), [49](#))

- International Labour Organization**, “Gender Equality in Recruitment and Promotion Practices in Viet Nam,” ILO’s Policy Brief 2015. (return to page [54](#))
- ISDS**, “Social determinants of gender inequality in Vietnam,” 2016. (return to page [47](#))
- James, Zoe**, “Young Lives Vietnam school survey round 1 (2011-2012): Data user guide,” 2013. (return to page [Appendix - 36](#))
- Jayachandran, Seema**, “The roots of gender inequality in developing countries,” *economics*, 2015, 7 (1), 63–88. (return to page [1](#))
- , “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 [37](#))
- Jiang, Xinyan**, “Confucianism, women, and social contexts,” *Journal of Chinese philosophy*, 2009, 36 (2), 228–242. (return to page [27](#))
- Kabeer, Naila, Thi Van Anh Tran, and ML Vu**, “Preparation for the future: priority strategies to promote gender equality in Viet Nam,” 2005. (return to page [7](#))
- 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 [1](#), [23](#), [37](#))
- **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 [22](#))
- 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 - 28](#))
- 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 [48](#))
- Lim, Jaegum 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 [28](#))
- Liu, Amy YC**, “Sectoral gender wage gap in Vietnam,” *Oxford Development Studies*, 2004, 32 (2), 225–239. (return to page [Appendix - 35](#))
- 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 - 36](#))
- Lowe, Matt**, “Types of contact: A field experiment on collaborative and adversarial caste integration,” *American Economic Review*, 2021, 111 (6), 1807–44. (return to page [1](#))
- Mai, Dung Thi Thuy and Mark Brundrett**, “The beliefs and attitudes of teachers and school leaders on gender equality in Vietnamese primary schools: case studies of four schools,” *Education 3-13*, 2020, 48 (4), 392–404. (return to page [15](#))
- Manski, Charles F**, “Identification of endogenous social effects: The reflection problem,” *The Review of Economic Studies*, 1993, 60 (3), 531–542. (return to page [18](#))
- Marianne, Bertrand**, “New perspectives on gender,” in “Handbook of labor economics,” Vol. 4, Elsevier, 2011, pp. 1543–1590. (return to page [4](#))
- 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 - 44](#))
- McAleavy, Tony, Tran Thai Ha, and Rachael Fitzpatrick**, *Promising Practice: Government Schools in Vietnam.*, ERIC, 2018. (return to page [10](#))
- McCrary, Justin**, “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 2008, 142 (2), 698–714. (return to page [48](#))
- 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 - 40](#))
- 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 [3](#), [33](#))
- 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 [27](#))

- Nguyen, Cuong**, “Gender Equality in Education, Health Care, and Employment: Evidence from Vietnam,” 2012. (return to page [15](#))
- Nguyen, Ngoc P**, “An assessment of the Young Lives sampling approach in Vietnam,” 2008. (return to page [Appendix - 35](#))
- 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 [18](#))
- 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 - 35](#))
- Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou**, “Mothers, peers, and gender-role identity,” *Journal of the European Economic Association*, 2020, 18 (1), 266–301. (return to page [1, 49](#))
- 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, 47](#))
- , **Marianne E Page, and Ann Huff Stevens**, “The intergenerational effects of compulsory schooling,” *Journal of Labor Economics*, 2006, 24 (4), 729–760. (return to page [3](#))
- Oster, Emily**, “Unobservable selection and coefficient stability: Theory and evidence,” *Journal of Business & Economic Statistics*, 2019, 37 (2), 187–204. (return to page [5, 19, 50, 51](#))
- 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](#))
- Pettigrew, Thomas F and Linda R Tropp**, “A meta-analytic test of intergroup contact theory,” *Journal of Personality and Social Psychology*, 2006, 90 (5), 751. (return to page [5](#))
- Pham, Bang Nguyen, Timothy Adair, Peter S Hill, and Chalapati Rao**, “The impact of the stopping rule on sex ratio of last births in Vietnam,” *Journal of Biosocial Science*, 2012, 44 (2), 181–196. (return to page [48](#))
- 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 [1](#))
- 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 [28](#))
- 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 [14](#))
- 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 [4](#))
- Shan, Xiaoyue**, “Does Minority Status Drive Women Out of Male-Dominated Fields?,” *Unpublished working paper*, 2021. (return to page [30, 33](#))
- 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 [6, 37, 48](#))
- 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 [20, 26](#))
- 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](#))
- Teerawichitchainan, Bussarawan, John Knodel, Vu Manh Loi, and Vu Tuan Huy**, “The gender division of household labor in Vietnam: Cohort trends and regional variations,” *Journal of Comparative Family Studies*, 2010, 41 (1), 57–85. (return to page [3, Appendix - 43](#))
- 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 - 19](#))
- van Lent, Max**, “Fathering Daughters and Personality,” *Available at SSRN 3742024*, 2020. (return to page [48](#))
- Westoff, Charles F et al.**, “Desired number of children: 2000–2008,” *DHS Comparative Reports*, 2010, (25). (return to page [Appendix - 42](#))

- 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 [1](#))
- Young Lives**, “Young Lives Survey Design and Sampling in Vietnam,” 2014. (return to page [11](#), [16](#), [Appendix - 35](#))
- , “Young Lives Survey Design and Sampling (Round 5): Viet Nam,” 2017. (return to page [51](#))
- Zölitz, Ulf and Jan Feld**, “The effect of peer gender on major choice in business school,” *Management Science*, 2021, 67 (11), 6963–6979. (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 (0.003)	0.012 (0.025)	0.009 (0.009)	-0.010 (0.015)	0.257*** (0.020)	0.007 (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 (0.155)	0.136 (0.291)	0.363 (0.283)	-0.044 (0.076)	-0.015 (0.017)	0.007 (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 during 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 (0.020)	-0.011 (0.010)	-0.012 (0.009)	0.012 (0.019)	-0.006 (0.010)	-0.003 (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 Rein- forcement Class	Enjoy School	Worry about Exams
YL Child	-0.046 (0.042)	-0.024 (0.032)	-0.002 (0.017)	-0.004 (0.011)	-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 who were present in the School Survey and whose class-assignment satisfied the sample selection criteria outlined in Section 2.2. 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)	(2)	(3)	(4)	(5)
	Mean	Standard Deviation	Min	Max	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)					
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 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. Mean Score is computed as the average score (on a 1-4 scale) across the twelve dimensions at the individual level. Respondents are from the Old Cohort and are interviewed in the fifth round of the YL survey (2016, age 21).

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	-0.017 (0.039)	-0.021 (0.034)	0.082 (0.105)	-0.054 (0.043)	0.192 (0.374)
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	2,886	2,883	2,894	2,889	2,877
R-squared	0.674	0.048	0.102	0.427	0.267
Panel B					
	(6)	(7)	(8)	(9)	
	Wealth Index	Home Educational Resources	Takes Private Classes	No Rein- forcement Class	
Proportion Female Peers	0.086* (0.050)	0.047 (0.055)	-0.097 (0.211)	0.070 (0.209)	
Mean (sd) Dep. Var.	0.52 (0.18)	0.71 (0.27)	0.38 (0.48)	0.63 (0.48)	
Observations	2,894	2,894	2,887	2,894	
R-squared	0.387	0.316	0.347	0.348	

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 2.2. 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)	(2)
	Home Academic Support	Academic Ability
Proportion Female Peers	0.016 (0.499)	0.040 (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 (0.253)	0.108 (0.538)	-0.066 (0.146)	-0.005 (0.255)	-4.498 (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 (6.254)	0.125 (0.081)	0.246 (0.218)	-0.106 (0.762)	-0.005 (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 characteristic (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 2.2. Standard errors clustered at the class level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.7: Exogeneity of Assignment of Students to Homeroom Teachers of Different Gender

	Female Teacher
Female	0.006 (0.007)
Father Can Read	0.014 (0.019)
Home Educational Resources Index	0.030 (0.034)
Wealth Index	0.006 (0.058)
1-5 Books at Home	-0.006 (0.021)
6-10 Books at Home	-0.016 (0.027)
More than 10 Books at Home	-0.027 (0.025)
Ethnic Minority (non-Kinh)	-0.045** (0.018)
If I Work Hard I Can Go To College	0.008 (0.012)
Worry About Exams	0.008 (0.011)
Mathematics Score First Test	-0.000 (0.000)
Low School Effort	-0.004 (0.005)
Age	0.238 (0.169)
Age ²	-0.010 (0.008)
p-value F-test	0.364
Observations	2,711
R-squared	0.565

Notes. Regressions of an indicator of the homeroom teacher being a female on his/her students' characteristics and school fixed effects. The p-value of the F-test of joint significance of the included potential predictors of teacher's gender is 0.364. 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 2.2. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Robustness to Different Sample Selections

	(1)	(2)	(3)	(4)
	Mean Agreement with Traditional Gender Norms at R5 (Age 15)			
Proportion Female Peers	-2.512*** (0.491)	-2.582*** (0.490)	-2.110*** (0.738)	-2.802*** (0.504)
Sample	Random+location+single	Random+location	Random	Full
Observations	880	843	728	1,027
R-squared	0.291	0.287	0.306	0.291

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 0.438, 0.279, and 0.404 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.9: Robustness of Effects on Gender Norms to First Standardizing Each Component

	(1)	(2)
	Standardized Norms (Age 15)	
Proportion Female Peers	-0.849** (0.399)	-2.509*** (0.488)
Peers Correction	No	Yes
Observations	880	880
R-squared	0.246	0.293

Notes. Replication of the specification in Table 3’s columns 2 and 7 where the outcome variable is constructed first standardizing each of its components, then obtaining the mean across all the components, and finally standardizing this value again. 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	-2.352*** (0.519)	-2.576*** (0.571)
Observations	548	248
R-squared	0.283	0.304

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	1.567*** (0.297)	0.111 (0.178)	0.007 (0.163)
Mean (sd) Dep. Var.	0.63 (0.22)	0.81 (0.23)	0.89 (0.23)
Observations	429	423	356
R-squared	0.475	0.479	0.37
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	1.525*** (0.304)	0.792*** (0.254)	0.479 (0.341)
Mean (sd) Dep. Var.	0.67 (0.23)	0.80 (0.25)	0.87 (0.26)
Observations	420	396	318
R-squared	0.438	0.389	0.336
Panel C: Combined Measures			
	(7)	(8)	(9)
Female Friends/Total Friends defined as...	1+4	2+5	3+6
Proportion Female Peers	1.566*** (0.330)	0.391 (0.236)	0.087 (0.235)
Mean (sd) Dep. Var.	0.71 (0.24)	0.85 (0.23)	0.91 (0.22)
Observations	418	381	269
R-squared	0.460	0.444	0.426

Notes. All regressions replicate the specification in Table 12 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: Nonlinear Effects of the Proportion of Female Peers on the Contribution to Home Production

	(1)	(2)	(3)	(4)	(5)	(6)
	R4 (Age 12) Total Home Hours	R5 (Age 15) Total Home Hours	R4 Over 2 Home Hours	R5 Over 2 Home Hours	R4 Any Home Hours	R5 Any Home Hours
Proportion Female Peers (> 50%)	0.272 (0.189)	0.427** (0.197)	0.078 (0.056)	0.099** (0.047)	0.029 (0.058)	0.123** (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 8 with the replacement of the proportion of female peers with 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.13: Effects of the Proportion of Female Peers on Academic Aspirations

	(1)	(2)
	Aspires to Reach University (YL R5, Age 15)	
Proportion Female Peers	-0.313 (0.426)	0.504** (0.212)
Sample	Males	Females
Observations	428	431
R-squared	0.385	0.366

Notes. These regressions replicate the specification in Table 3's column 7 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.14: 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	1.349*** (0.348)	0.549* (0.309)	2.314 (1.460)	1.449 (1.151)	
Mean Expected Returns					0.166*** (0.056)
Variance Expected Returns					-0.001 (0.016)
Sample	Females	Males	Females	Males	All
Observations	424	439	425	439	776
R-squared	0.287	0.296	0.243	0.230	0.294

Notes. Under the standard assumption of a perceived triangular distribution of wages, 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]? The outcome in column 5 is measured in 2020 — the Covid round — and takes 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.15: 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 (0.174)	-0.204 (0.141)	0.345** (0.174)	0.014 (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 traditional female jobs (coded as 1), gender neutral jobs (coded 2), and traditional 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.16: Effects of the Proportion of Female Peers on Self-confidence in Abilities

	(1)	(2)
	Confidence in Self (YL R5, Age 15)	
Proportion Female Peers	0.325* (0.184)	0.548*** (0.165)
Sample	Males	Females
Observations	433	431
R-squared	0.315	0.287

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.17: 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	-0.034 (2.515)	-1.471* (0.798)	-0.337 (0.433)	-0.536** (0.262)
Mean (sd) Dep. Var.	27.28 (3.02)	1.99 (0.39)	1.03 (0.25)	0.97 (0.28)
Observations	423	433	433	433
R-squared	0.389	0.271	0.207	0.251
Panel B: Females				
	(5)	(6)	(7)	(8)
	Age First Child	Total Number of Children	Number of Sons	Number of Daughters
Proportion Female Peers	0.835 (1.982)	-0.750* (0.380)	-0.293 (0.202)	-0.320 (0.244)
Mean (sd) Dep. Var.	26.72 (2.96)	1.95 (0.40)	0.97 (0.25)	1.00 (0.29)
Observations	408	431	430	431
R-squared	0.411	0.322	0.314	0.241

Notes. All regressions replicate the specification in Table 3's column 7, 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.18: Heterogeneous Effects of the Proportion of Female Peers on Long-term Views on Gender Norms

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean Agreement with Traditional Gender Norms (YL R5, Age 15)					
Proportion Female Peers	-2.520***	-2.497***	-2.537***	-2.606***	-2.500***	-2.433***
	(0.483)	(0.504)	(0.544)	(0.512)	(0.552)	(0.482)
Any Parent Illiterate*Proportion Female Peers	-1.265					
	(1.263)					
Ethnic Minority*Proportion Female Peers		0.351				
		(0.513)				
High Wealth Index*Proportion Female Peers			0.039			
			(0.574)			
Not Better Student than Other*Proportion Female Peers				0.273		
				(0.312)		
Low School Effort*Proportion Female Peers					-0.363	
					(0.225)	
High Interest School*Proportion Female Peers						0.297
						(0.265)
Sample	All	All	All	All	All	All
Observations	880	880	880	849	844	851
R-squared	0.296	0.291	0.291	0.298	0.308	0.301

Notes. All regressions replicate the specification in Table 3's column 7. A description of the interacting variables used is provided in Section D. Level effects are omitted for brevity since no heterogeneous effects are present. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.19: Effects by Gender of the Proportion of Female Peers and Exposure to High-Ability Female Peers On Views on Gender Norms

	(1)	(2)	(3)	(4)
	Life	Abilities	Cross-Gender	Intra-household
	Purpose		Interactions	Decisions
Panel A: Males				
Proportion Female Peers	-1.399**	-0.335	-2.412***	0.587
	(0.580)	(0.822)	(0.834)	(1.198)
High Ability Female Peers	-0.144	-0.437***	0.339**	0.132
	(0.144)	(0.144)	(0.158)	(0.272)
Observations	430	430	430	430
R-squared	0.292	0.309	0.308	0.192
Panel B: Females				
Proportion Female Peers	-1.773**	-2.327**	-2.528***	-1.722
	(0.817)	(1.087)	(0.950)	(1.258)
High Ability Female Peers	0.079	-0.152	0.079	0.085
	(0.188)	(0.184)	(0.228)	(0.193)
Observations	417	417	417	417
R-squared	0.320	0.305	0.240	0.294

Notes. Replication of Table 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 (defined as being 30% of a standard deviation higher than the sample mean). Standard errors clustered at the class level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.20: Non-selective Attrition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Matched with YL's Round 5 (Age 15)						
Female	-0.001 (0.008)						-0.002 (0.007)
Ethnic Minority		0.004 (0.009)					-0.001 (0.012)
Home Educational Resources			-0.018 (0.016)				-0.013 (0.017)
Literate Mother				-0.026 (0.021)			-0.026 (0.024)
Low School Effort					-0.001 (0.003)		-0.002 (0.003)
Wealth Index						-0.042 (0.037)	-0.035 (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 “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.21: 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	-0.992** (0.428)	-0.841** (0.408)	-0.633 (0.552)	-0.665 (0.652)	-0.697 (0.582)
Observations	880	880	880	880	880
R-squared	0.183	0.219	0.225	0.220	0.17

Notes. Quantile regressions using the same controls as in Table 3's column 7. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.22: 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	-2.509*** (0.444)	-2.428*** (0.467)
Observations	629	562
R-squared	0.267	0.272

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 (3 schools) and 0.1 (6 schools) in columns 1 and 2, respectively. Schools with only one class are also excluded since multiple classes are needed to conduct Fisher's test. Standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.23: 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	-4.300 (3.870)	0.057 (0.107)	-0.041 (0.412)	-0.082 (0.268)	0.030 (0.054)	-0.210 (0.233)
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 exposure to female peers. Normal BMI is defined based on the z-score for the child's 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.24: Robustness Check: Effects on Sibling's Outcomes

	(1)	(2)	(3)	(4)	(5)
	R4 (Age 12) Total Home Hours	R4 Any Home Hours	R5 (Age 15) Total Home Hours	R5 Any Home Hours	R4 PPVT Raw Score
Proportion Female Peers	1.293 (2.096)	0.046 (0.414)	0.494 (1.212)	0.125 (0.359)	-3.837 (7.635)
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	0.423	0.485	0.502	0.421	0.577

Notes. All outcomes are obtained from 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.25: 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	0.313 (0.242)	0.395 (2.813)	478.426 (490.965)	-0.664 (0.577)	0.166 (0.270)
Observations	877	876	871	859	879
R-squared	0.401	0.398	0.259	0.209	0.191

Notes. Replication of the specification in Table 3's column 7 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.26: 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	-2.251* (1.173)	-1.795 (1.105)	-2.526*** (0.811)
Sample	< 0.25	0.1 < difference < 0.25	0.35 < level < 0.65
Observations	600	395	738
R-squared	0.302	0.306	0.312

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.27: Robustness Check: Accounting for Extreme Responses

	(1)	(2)
	Log Norms	Winsorized Norms
Proportion Female Peers	-0.389*** (0.082)	-2.418*** (0.495)
Observations	880	880
R-squared	0.283	0.292

Notes. Replication of column 7 in Table 3 when using as outcomes the logarithm of the 1–4 measure of traditionalism (column 1) and when winsorizing the z-scored version for the top and bottom 1 percent of the distribution (column 2). Standard errors clustered at the class level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.28: 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	-2.587*** (0.496)	-2.467*** (0.510)
Aggregate Fixed Effects	Region	Community
Observations	880	859
R-squared	0.297	0.318

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

Table A.29: Is There Selection into Subsequent Academic Peer Groups?

	(1)	(2)
	Proportion Female Peers 2017 (Age 16)	Ever Transfers Schools
Proportion Female Peers 2011	0.207 (0.403)	-0.022 (0.090)
Observations	208	878
R-squared	0.527	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.30: 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	-3.044*** (0.550)	-2.832*** (0.939)
Female	-0.636*** (0.087)	-0.501*** (0.145)
Observations	509	182
R-squared	0.260	0.314

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

B Additional Figures (For Online Publication)

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

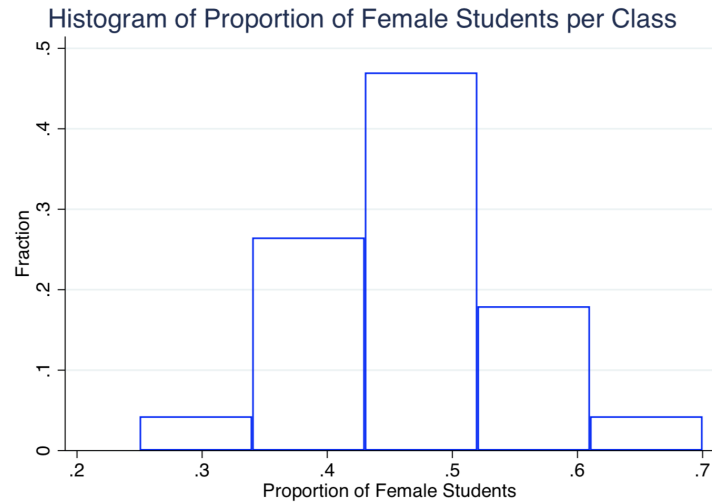
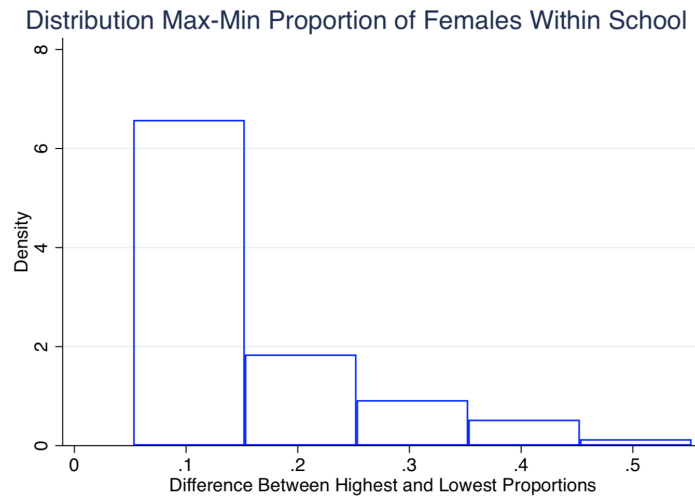


Figure B.2: Distribution of the Within-School Variation in the Proportion of Female Peers



⁹⁴The horizontal axis labels are as follows: 1: strongly disagree; 2: disagree; 3: agree; 4: strongly agree. 95% confidence intervals are also reported.

⁹⁵Ethiopia, 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 of Views on the Different Dimensions of Traditionalism by Gender⁹⁴

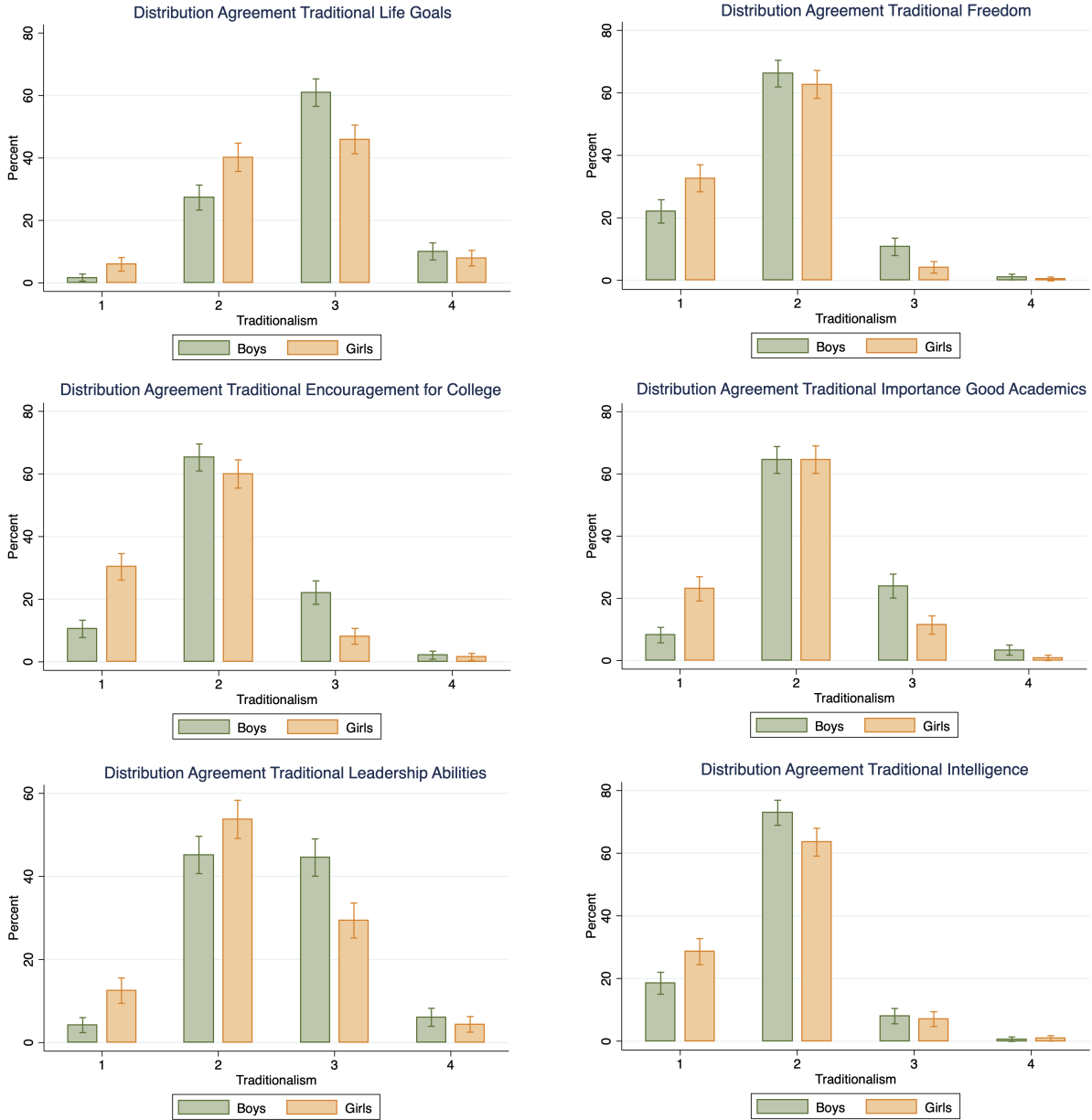


Figure B.4: Distribution of Views on the Different Dimensions of Traditionalism by Gender (Continued)

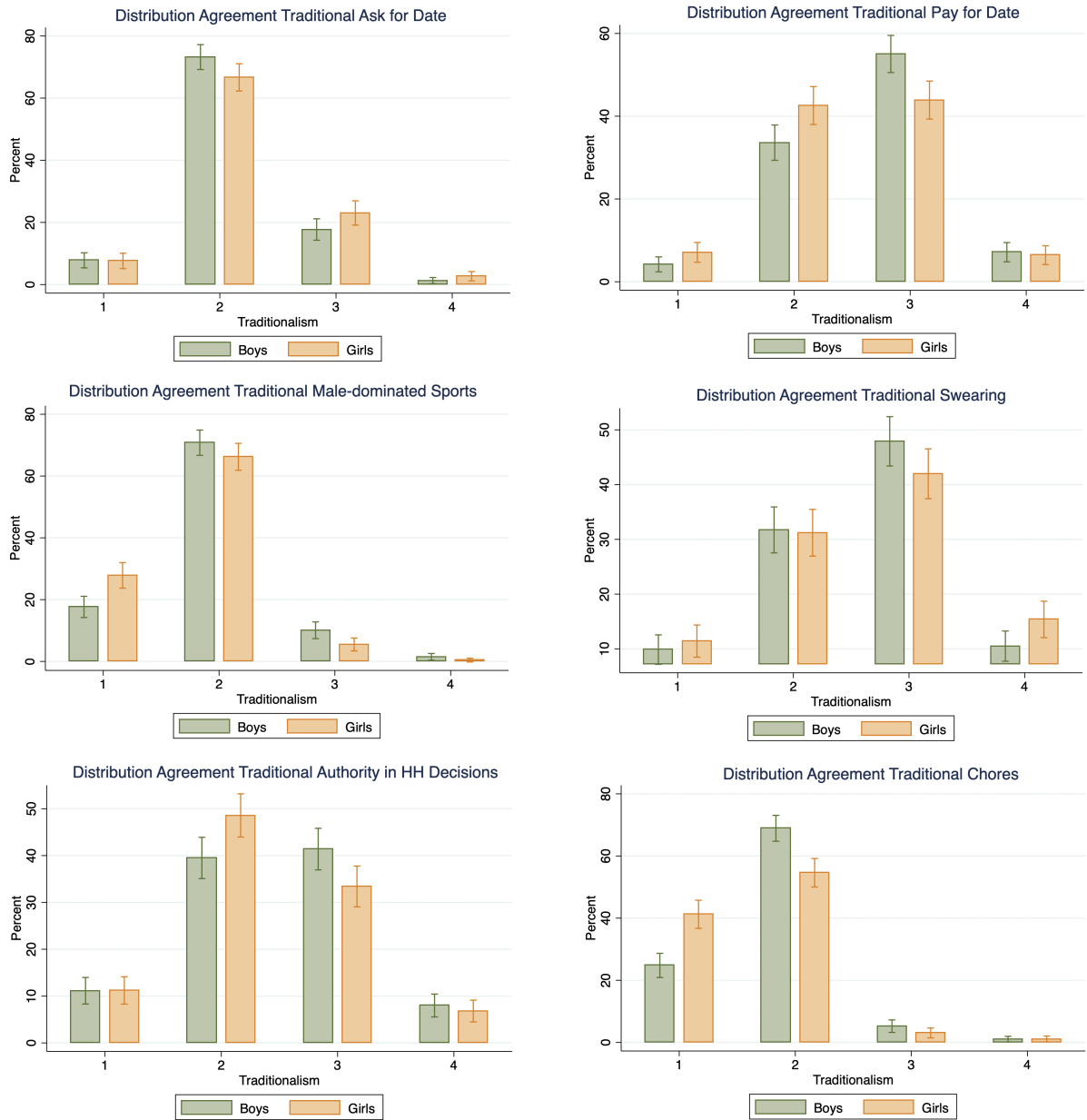


Figure B.5: Distribution of Traditional Gender Views Across Countries (Source: World Values Survey, Wave 5)⁹⁵

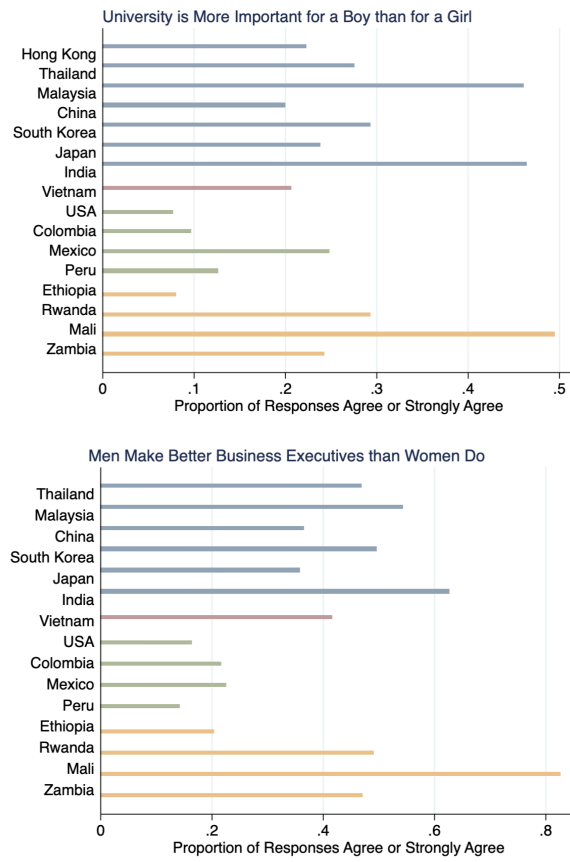


Figure B.6: Distribution of the Difference Between Actual and Simulated Proportion of Female Peers for One Randomly-selected Class per School

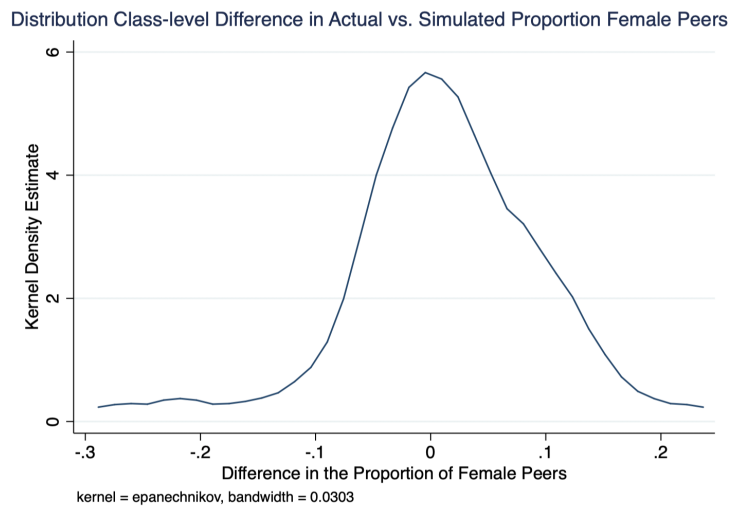
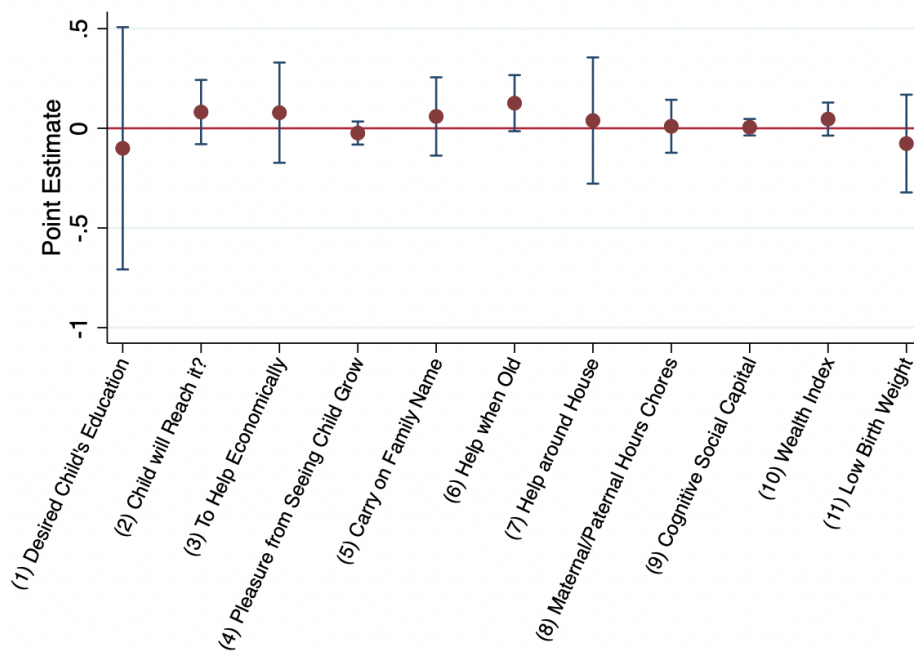
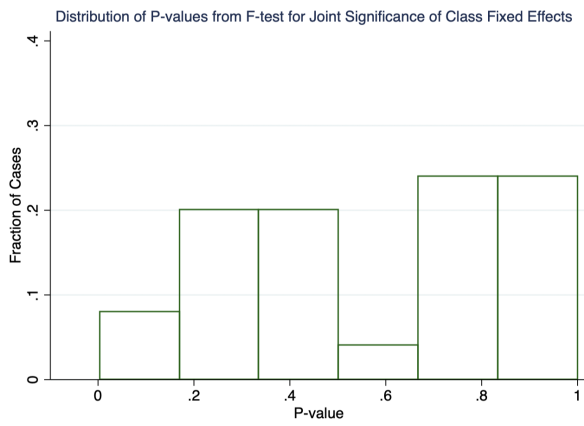


Figure B.7: Balance of Pre-determined Covariates: (i) Parental Expectations and Reasons for Having Children (Age 5); (ii) YL Basic Characteristics in Round 1 (Age 1)

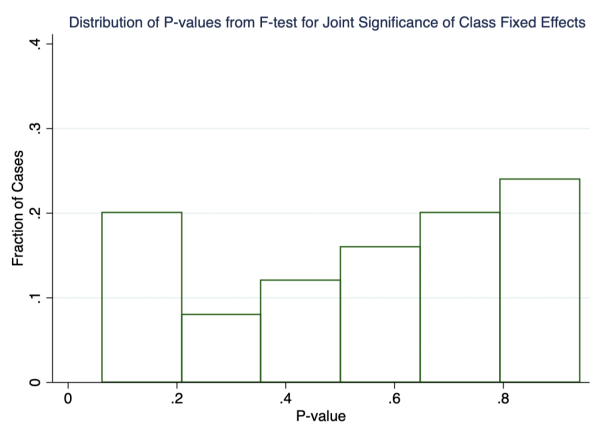


Notes. Regressions replicate those from Table 2 with the inclusion of a gender indicator as control (the qualitative and quantitative results are virtually identical if the gender indicator is not included). 1 uses as outcome the years of education that the parents wish their child to achieve (I code 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 give birth to a 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 based on caregiver's responses to 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". 90% confidence intervals are reported.

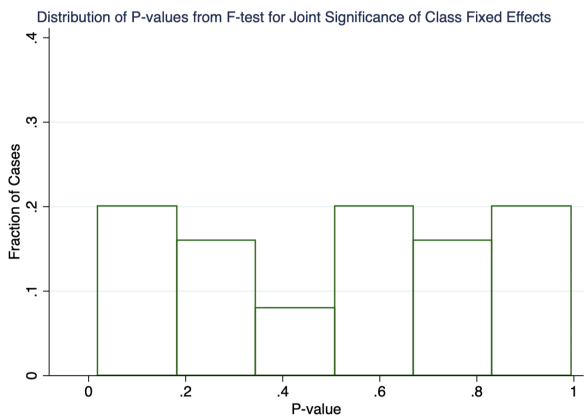
Figure B.8: Distribution P-value Joint Significance F-Tests for Class Fixed Effects: Parental Expectations Measured at Age 5 (I)



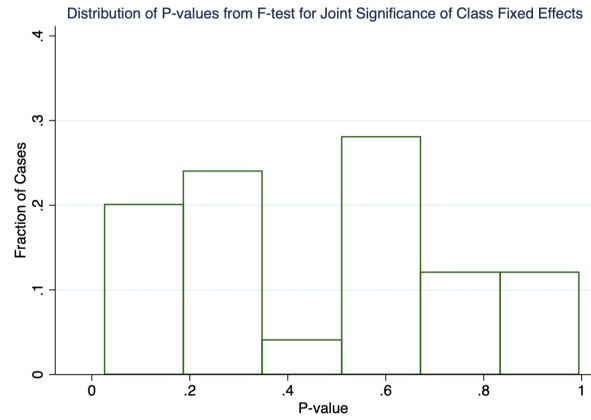
(a) Age Start Earning Money



(b) Age Leave Education

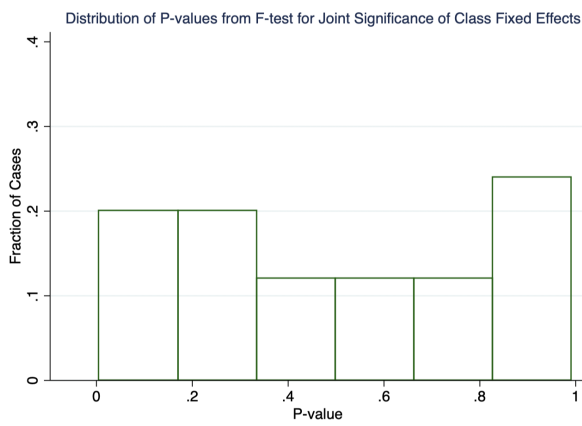


(c) Age Financially Independent

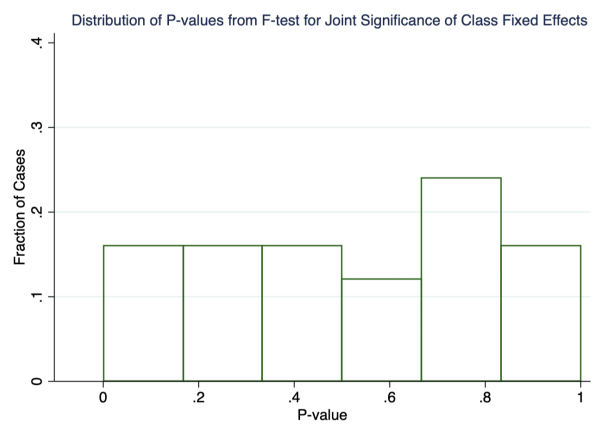


(d) Age Form Own Household

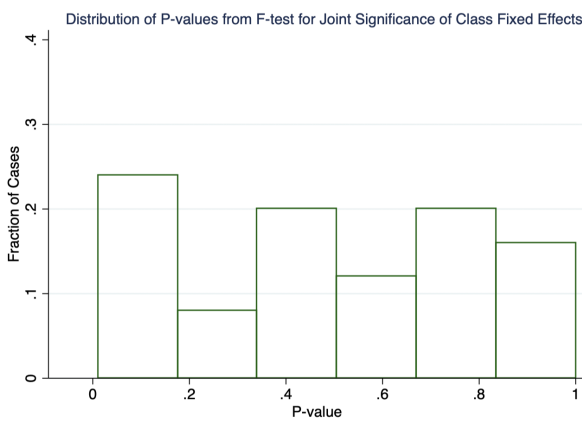
Figure B.9: Distribution P-value Joint Significance F-Tests for Class Fixed Effects: Parental Expectations Measured at Age 5 (II)



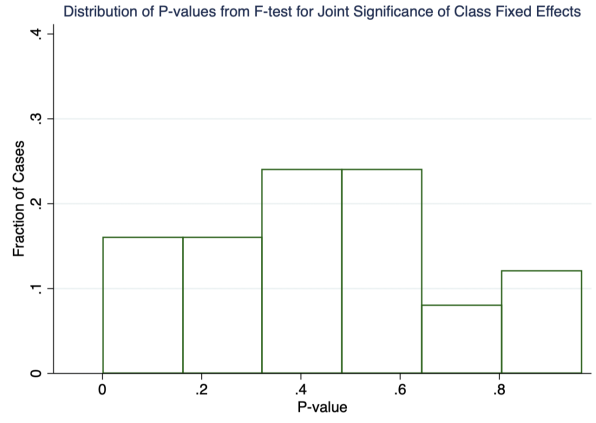
(a) Age Marriage



(b) Age Have Children

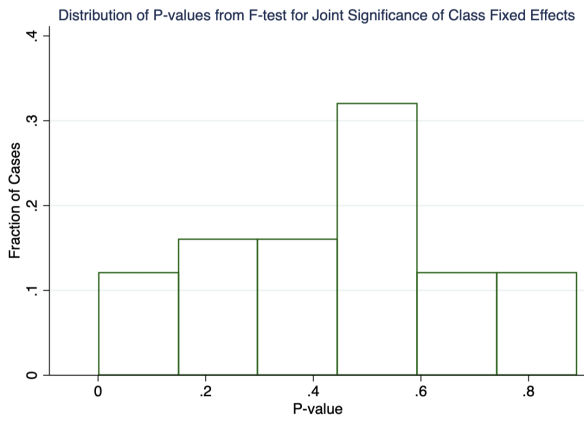


(c) Should Live Close

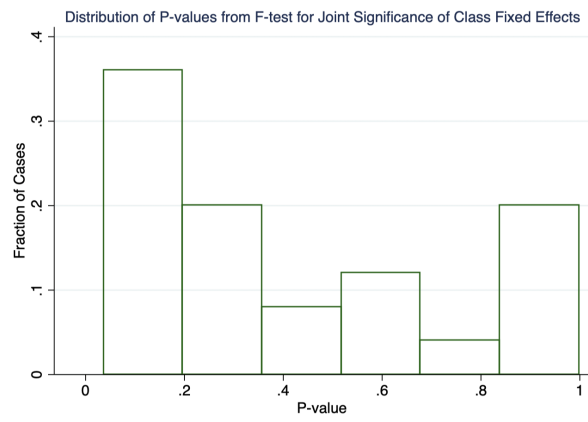


(d) Should Help Siblings Financially

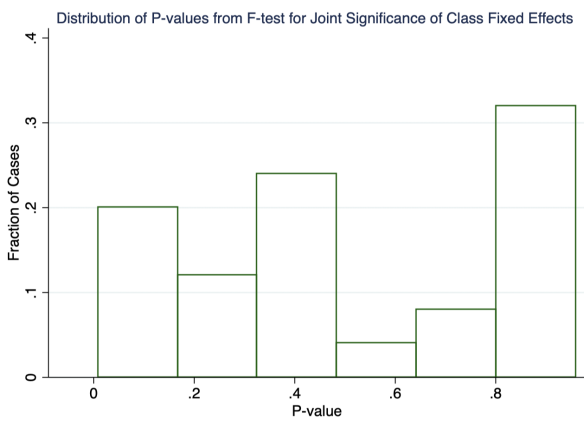
Figure B.10: Distribution P-value Joint Significance F-Tests for Class Fixed Effects: Parental Expectations Measured at Age 5 (III)



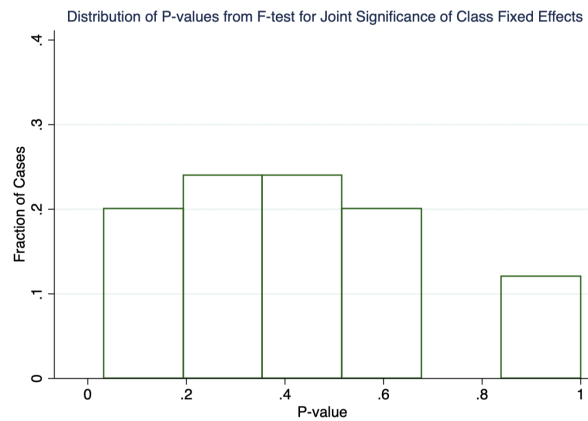
(a) Should Help Housework



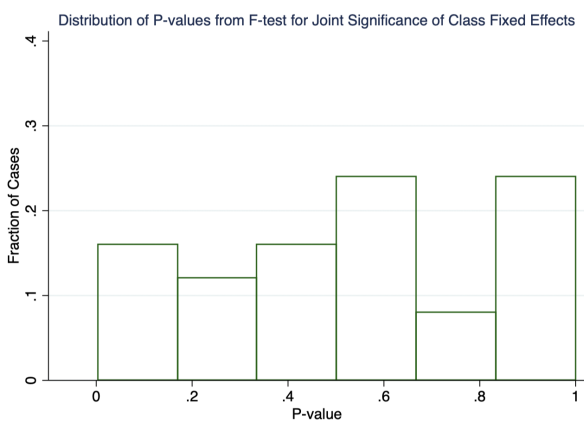
(b) Should Help Parents Financially



(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)

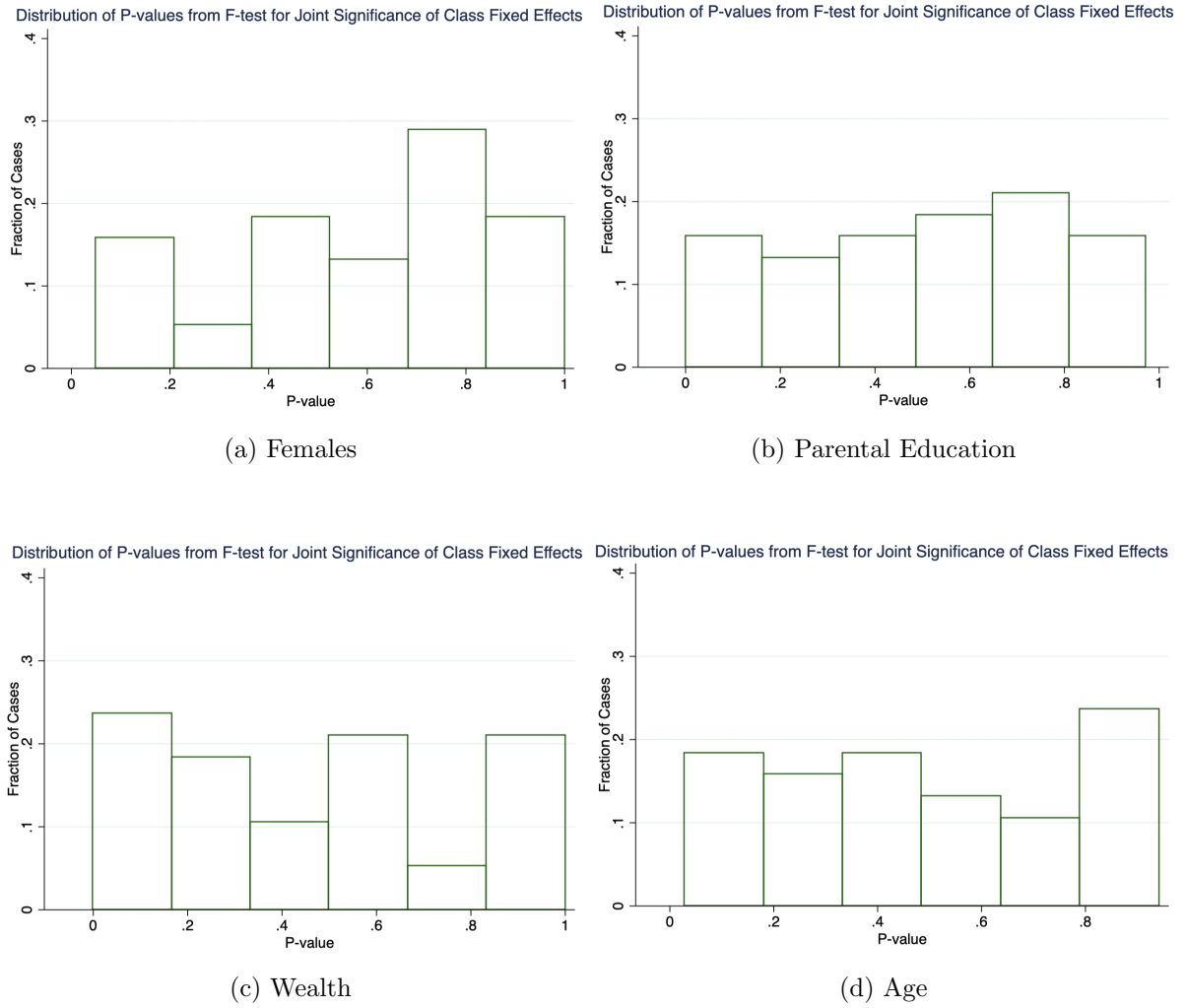


Figure B.12: Estimates from All Combinations of Main Controls

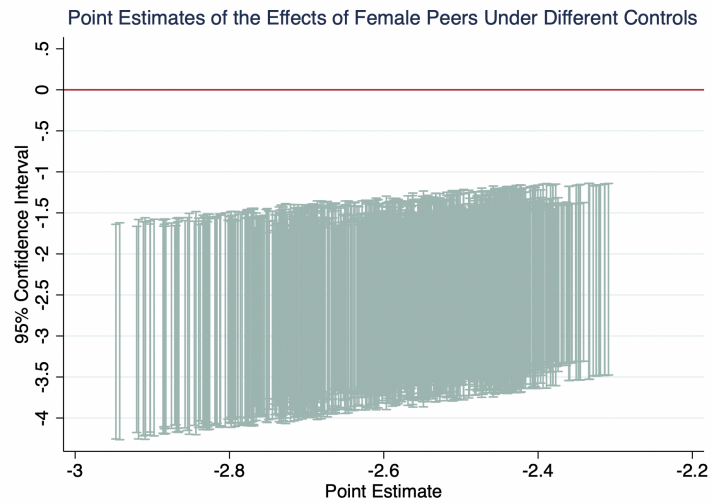


Figure B.13: Allowing for Different Slopes: Predicted Values from Model with Linear Splines

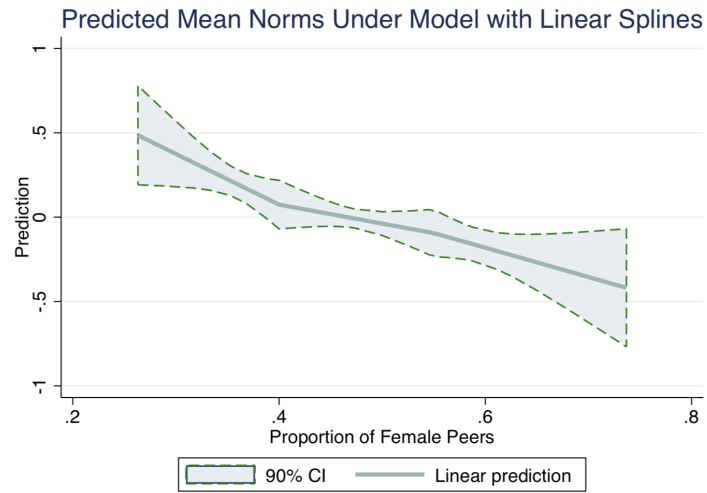


Figure B.14: Distribution of Placebo-Generated Estimates of the Proportion of Female Peers on Gender Norms (empirical 95% CI indicated by green lines) When Reshuffling Attitudes Within Schools

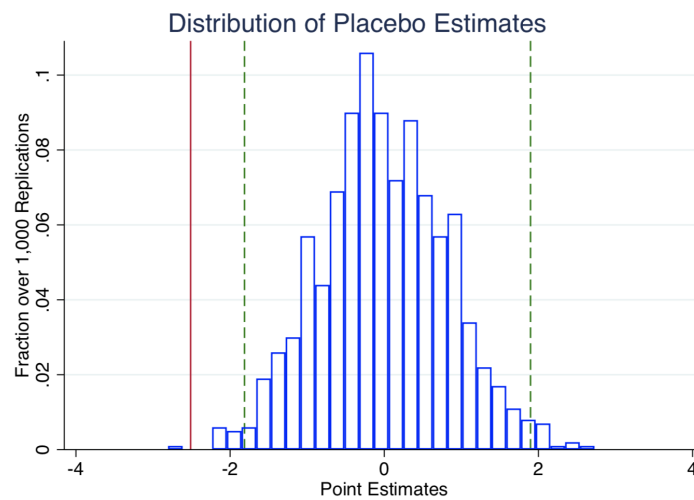
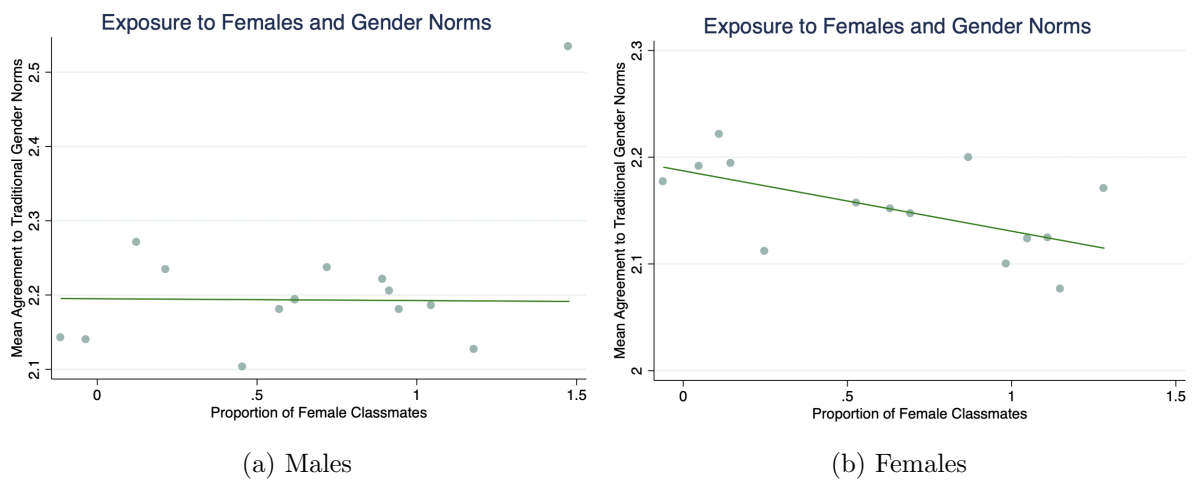


Figure B.15: Correlations of Proportion of Female Peers and Long-term Views on Gender Norms by Student Gender: Ethiopia



C Extended Analyses (For Online Publication)

C.1 Representativity of the Sample

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 sample 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 target population (Nguyen, 2008). Indeed, I highlight that those key outcomes of interest that can be contrasted with nationally representative studies (e.g., attitudes towards several gender norms and fertility preferences) yield very close figures.⁹⁶

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 wave⁹⁷ 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 more likely to be enrolled in school, since this was a requirement to be eligible for 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 suggest that YL children are systematically different from their peers. Table A.2 shows that this is not the case.⁹⁸

C.2 Additional Details on Available Survey Information

Sociodemographics (School Survey). 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 index⁹⁹. This is useful because wealth has been shown to be important for addressing

⁹⁶Another 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).

⁹⁷The original cross-sectional dimension (i.e., at the first round) of this cohort was 2,000.

⁹⁸See Section D for more details on the variables used.

⁹⁹I 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; car.

omitted 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 (School Survey). 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).¹⁰⁰ 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 (School Survey). The School Survey collects rich information on the subset of randomly-selected peers — I provide evidence on the successful randomization in Table A.2.¹⁰¹ 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 classmate¹⁰²: (1) how would you describe your friendship with this classmate?, and (2) how much do you do things with this classmate outside of school?¹⁰³ I use this information

¹⁰⁰Importantly, on the date of the second round (an unannounced one), only 2% of the children did not attend school.

¹⁰¹More details on the process followed and its success are provided in James (2013).

¹⁰²This 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.

¹⁰³The options available for question (1) were: not close friends; a little/sometimes friends; close friends; very close friends.

to explore friendship formation and cross-gender interactions.

Teacher Characteristics (School Survey). 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 (School Survey). 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. 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

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.¹⁰⁴

¹⁰⁴To check whether these very low mean values of the distance variable are a result of averaging out larger differences across classrooms within schools, in Figure B.6 I plot the 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 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 are four)¹⁰⁵ — 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, in Table A.4 I show that the proportion of female classmates is not correlated with a wide range of individual observables like ethnicity, parental education, child’s health conditions, and educational resources at home.¹⁰⁶ 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 characteristic) 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 0.95.¹⁰⁷

Finally, performing Fisher’s exact tests 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

¹⁰⁵As consistent with randomization, the proportion at the school level has the same mean and standard deviation as the one at the class level.

¹⁰⁶I 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 conventional levels of significance (p-value of 0.111).

¹⁰⁷Additionally, I again follow Feld and Zölitz (2017) in running separate regressions for each school for the main predetermined characteristics (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.

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 Further Discussion on Friendship Formation

A reasonable counterfactual to benchmark the estimates in Table 12 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 females¹⁰⁸ (among those students classified as friends in column 1 in Table 12, 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 friends.¹⁰⁹

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 in the actual classroom composition. To be more specific, I proceed as follows. I first compute the proportion of female and male friends in the actual data that satisfy the friendship definition in column 1 of Table 12 — 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 $100 \cdot (0.393 - 0.35) \approx 4.3$ percentage points higher than the empirical one.¹¹⁰

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 12 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 12 is

¹⁰⁸This empirical pattern of systematic tendencies of forming friendships with same-gendered individuals is often termed as “homophily” (McPherson et al., 2001).

¹⁰⁹This 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.

¹¹⁰In 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.

0.930 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 $13.05 \times 0.930 \approx 12.14$ 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 disproportionately more than predicted ($12.14 \gg 4.3$ 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 12 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 12, was expected: it reinforces the idea that children are more likely to establish closer relationships with same-gendered peers (homophily). 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 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 the 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.5 Additional Attitudinal Outcomes: Fertility Preferences

In Table A.17, 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 desiring less (Ashraf et al., 2014).¹¹¹ 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

¹¹¹For 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).

for males and females, respectively. Males that experienced more contact with female peers wish to have significantly less children, and this is particularly true for daughters. On the other hand, females experience a somewhat smaller decline (not driven by a particular child gender) and yielding an average desired number of children slightly below the natural replacement rate (1.95)¹¹². If anything, there are some signs that females wish to somewhat postpone the first childbearing. This was expected given the increases in the ambition for a professional career uncovered in the main text, 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.6 Further Evidence on Mechanisms

C.6.1 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.

In Table A.18 I explore heterogeneity based on: (1) an indicator of any of the parents being illiterate; (2) an indicator for belonging to an ethnic minority; (3) an indicator for 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 not significant.¹¹³

Heterogeneity by Regional Traditionalism. I complement the above analysis at the micro level by exploring whether communities (the sampling units in the first stage of the YL’s sampling procedure) with more traditional values at the aggregate level experience larger attitudinal shifts on average. For this, I make use of the views on gender roles reported by

¹¹² 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.

¹¹³ I do not report the levels of the interacting variables for the sake of parsimony.

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 they are larger for the more traditional areas (p-value of the null hypothesis of equality is 0.056).¹¹⁴ Qualitatively similar results are found if I instead split the sample between northern and southern regions, the latter being more economically developed and less traditional (Teerawichitchainan et al., 2010), based on the 17th parallel.

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 Table A.19 I undertake the same analysis as in Table 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 are presented in the top panel. They show that, while the estimates of the effects of the female share barely change, being exposed to more able females helps in making males' views on females' overall cognitive and leadership abilities more progressive — as expected.¹¹⁵ Interestingly, the opposite is found for cross-gender interactions, which could be rationalized by the fact that more academically able females are expected to devote less time to sports and to rely less on poor etiquette and improper language. Turning to the results for the female subsample, we do not find a significant role of female peers' ability. This suggests that the main effects of the role of peers were more likely driven by reduced exposure to traditional norms (because of having a lower share of males in the class) than to a positive updating of their perceptions of the skills of other females.

C.7 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 at the 5% level. In Table A.22, I show that

¹¹⁴A complement of this analysis is to explore whether the results differ in rural and urban locations. I find the main effect to be -2.505 for rural areas and -2.526 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.

¹¹⁵A natural extension is to interact both variables of interest. The interaction is negative and significant even after controlling for own and male peers' ability.

excluding students from these schools (and for those for which I cannot reject it at the 10% level) 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 an 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 that these students might have nevertheless been in different trajectories that just had not been captured that early in children's lives. Table A.23 shows that various measures of health and human capital (e.g., height, which is largely determined around the time of entry to school¹¹⁶) 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 if I did not find evidence for this in my balance check for round 2. Table A.24 shows that indeed there are no differences in the siblings' intensive nor the extensive margin of home production in rounds 4 and 5 nor in their 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.

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.25 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

¹¹⁶Martorell and Habicht (1986) discuss how adult height is largely determined by age 7.

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).¹¹⁷ 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. 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.26 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.¹¹⁸ Column 3 shows that this is robust to considering, 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 emerging across classes within a school that had 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 teacher 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.27 I repeat the estimation of column 7 in Table 3 both using as outcome the logarithmic transformation of my (pre-standardized) gender norms measure (column 1) and winsorizing the standardized values at the 1st and 99th percentiles (column 2). The lack of significant changes

¹¹⁷There 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.

¹¹⁸It 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 female would already set it above the 0.25 threshold that I am using.

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 -2.293 — very similar to the baseline result — and significant at the 1% level.¹¹⁹ 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 -2.871, significant at the 5% level, for the subsample of communities with at least one other school.

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.28 I replicate column 7 in Table 3 while controlling for region¹²⁰ 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

¹¹⁹Given 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.

¹²⁰The six regions in my sample are: Northern Uplands, Red River Delta, Phu Yen, Da Nang, South Eastern, and Mekong River Delta.

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, while not all the non-YL classmates were. 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.29 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 9¹²¹. Column 2 in Table A.29 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 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 37, students changing schools throughout primary education is a very infrequent event (likely because there is no variation in school ownership that parents may want to exploit and because the main criterion for school choice is proximity to residence), and for this to cause a bias they should be able to affect their class allocation in the new school. 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 Smaller Schools. In Online Appendix Table A.30, I reproduce my main analysis restricting the 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.

¹²¹There 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).

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”¹²²: an indicator taking the value of 1 if the proportion of

¹²²The 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.
 - "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?"