Essays on gender, development and human capital investment

Dimitria Gavalyugova

TESI DOCTORAL UPF / Year 2020

THESIS SUPERVISOR
Libertad González and Alessandro Tarozzi
Department of Economics and Business

Universitat Pompeu Fabra
Barcelona

To my parents and sisters for always believing in me.

Acknowledgements. I would like to thank my supervisors Libertad González and Alessandro Tarozzi for their guidance and time. These papers received helpful feedback from seminar participants at Pompeu Fabra University, University of New South Wales, the Reserve Bank of Australia and the Conference of the Spanish Economic Association. I am also grateful to the Australian Data Archive at the Australian National University for granting me access to the Longitudinal Survey of Australian Youth and the Australian Youth Surveys. My family, Nick and Skye made this thesis possible.

Abstract

This dissertation comprises three chapters on the intersection between gender, development and human capital investments. The first chapter explores the impact of the rapid expansion of the Cambodian garment industry on women's and children's well being. It documents a sizeable increase in schooling at early ages, but also increased secondary dropouts. It also demonstrates that the growth of garment manufacturing is associated with delays in marriage and childbearing, and potentially long-lasting improvements in girls' height. The second chapter uses a natural experiment from an Australian state and shows that relaxing compulsory mathematics and science requirements widens the gender gap in high-school STEM subject uptake. It also documents a positive externality from compulsory mathematics requirements on the uptake of science subjects, which is consistent with a setting in which it is costly to study science without any mathematics. The third chapter presents evidence that the growth of private primary schooling may have negative implications for equality in educational opportunities and learning outcomes in rural India.

Resumen

Esta disertación comprende tres capítulos sobre la intersección entre género, desarrollo e inversiones en capital humano. El primer capítulo explora el impacto de la rápida expansión de la industria de la confección de Camboya en el bienestar de las mujeres y los niños. Documenta un aumento considerable de la escolarización a edades tempranas, pero también un aumento de la deserción en la secundaria. También demuestra que el crecimiento de la fabricación de prendas de vestir está asociado con retrasos en el matrimonio y la maternidad y mejoras potencialmente duraderas en la altura de las niñas. El segundo capítulo utiliza un experimento natural de un estado australiano y muestra que la relajación de los requisitos obligatorios de matemáticas y ciencias amplía la brecha de género en la captación de materias STEM en la escuela secundaria. También documenta una externalidad positiva de los requisitos matemáticos obligatorios en la adopción de asignaturas de ciencias, lo cual es consistente con un entorno en el que es costoso estudiar ciencias sin matemáticas. El tercer capítulo presenta evidencia del crecimiento de la educación primaria privada que puede tener implicaciones negativas para la igualdad en las oportunidades educativas y los resultados del aprendizaje en la India rural.

Preface

The extensive body of research originating from Gary Becker's seminal work in the 1960s has shown that decisions about human capital formation are complex, shaped by a variety of institutional, economic and behavioral factors. Some of the most consistent findings relate to the presence of enduring gender- and income-driven inequalities in human capital investments, in both developed and developing countries. In developing economies, where initial levels of investments are low and demand for skilled labor is growing, designing effective mechanisms to reduce gaps in health and schooling can have large marginal returns. In high- and low-income countries alike, understanding how educational policies interact with gender constructs could provide a path towards eliminating wage gaps and occupational segregation. Yet, the details of these processes are far from being well understood.

This dissertation advances understanding on these issues by studying the institutional and economic drivers of gaps in human capital investments by gender and socioeconomic background in different settings. The first chapter presents evidence that the expansion of export-oriented garment manufacturing in Cambodia increased women's years of primary schooling, but also contributed to higher dropouts from secondary education. It also demonstrates that access to garment employment delayed women's marriage and childbearing, and improved investments in girls' health at early ages. The second chapter shows that a high-school curriculum reform that disincentivized the study of mathematics in Australia led to a large share of females, relative to males, dropping out of both mathematics and science. The third chapter provides evidence that the expansion of private primary schooling in India could be contributing to the persistent income-driven inequalities in learning outcomes in the country.

Contents

List of Figures

List of Tables

Chapter 1

AS YOU SEW, SO SHALL YOU REAP? THE ROLE OF GARMENT MANUFACTURING IN THE LIVES OF CAMBODIAN WOMEN

1.1. Introduction

More than 40 million workers worldwide were employed in the garment industry in 2014, the majority of whom were in developing Asia, and three quarters of whom were women (ILO, 2015). Beginning in China, India and Bangladesh in the 1980s, the garment industry's presence in Asia expanded further southeast in the 1990s. Cambodia is one of the countries most affected by this expansion. Since 1995, employment in Cambodia's garment sector has grown from less than 50,000 in 1995 to 700,000 in 2017 (ILO, 2017). It now accounts for more than 40% of women's wage employment in the country and is a significant contributor to its threefold increase between 2000 and 2015 (World Bank, 2018). These drastic shifts from agricultural labour to formal salary employment opportunities are bound to affect numerous aspects of women's lives. This paper presents a comprehensive analysis of the role the expansion of garment manufacturing played in women's employment, education, marriage, fertility, and health.

Manufacturing expansion has often been associated with increases in female labor force participation (Atkin, 2009; Heath and Mobarak, 2015). Rigorous evaluations of the welfare effects of garment manufacturing on women, however, are few (Atkin, 2009; Sivasankaran, 2014; Heath and Mobarak, 2015), and the majority of the existing analyses focus on India and Bangladesh (Sivasankaran, 2014;

Heath and Mobarak, 2015, Kabeer, 2002; Hewett and Amin, 2000). With Vietnam on its way of surpassing Bangladesh by value of garment exports and Cambodia having the highest per-capita garment export value in the world (World Trade Organization, 2018), it is becoming increasingly important to provide a rigorous analysis of the various welfare implications that garment factory jobs have in these recently affected regions. With the current paper I attempt to address this gap in the literature.

This study fits among the group of works examining the effects of export-oriented manufacturing on development outcomes (Verhoogen, 2008; Atkin, 2016; Brambilla, 2013). It is also closely related to the literature examining the impact of wage employment opportunities on women and children's lives and well-being (Dasgupta, 2000; Strauss and Thomas, 1995; Atkin, 2009; Heath and Mobarak, 2016). The results in this paper indicate that access to garment jobs has a sizeable positive effect on women's educational attainment, both in absolute terms and relative to boys. However, it also leads to women's increased dropout from post-primary stages of education, when garment job access raises the opportunity cost of schooling. These findings hold a significant relevance to policy in developing contexts, and especially in Cambodia, where most students stop attending after primary school and policy-makers are actively seeking ways to bolster secondary enrolments (World Bank, 2019). The literature studying the determinants of schooling in developing countries commonly focuses on supply-side interventions, such as building schools and providing cash transfers to increase attendance (Kremer and Holla, 2009; Duflo, Dupas, and Kremer, 2011; Filmer and Schady, 2008). A better understanding of demand is crucial, as it could imply that the expansion of low-skilled export-oriented manufacturing jobs in much of the developing world may be an effective way to improve outcomes for students who would otherwise obtain no schooling, but may also limit the returns to education above a certain educational threshold.

I exploit the variation in timing and location of garment factories between 2000 and 2014 – a period of rapid geographic expansion of garment manufacturing in Cambodia. For the major part of my analysis I use data on district-level factory openings merged with the four waves of the Cambodian Demographic and Health Survey (Cambodia DHS). The district-level panel nature of the data, combined with the fact that garment exporters chose their locations to minimize costs, rather than based on population characteristics, allow me to alleviate concerns linked with the non-random allocation of factories to districts. In all specifications I control for district level trends in outcomes, including by gender. I also rule out the possibility that local changes in infrastructure resulting from the opening of a factory could be driving the results. Following the removal of garment and textile quotas in 2005[1], garment manufacturing in the country rapidly intensified and spread from six percent of districts in 2000 to 25 percent in 2014 (1.2). This rapid expansion, combined with evidence that most work commutes take place within districts (Cambodian Population Census, 2008) and the large average size of garment factories relative to district population, allows me to explore the transformative effects that the arrival of garment manufacturing has on women's lives.

Similarly to Heath and Mobarak (2015), I use a binary variable denoting increased access to garment work, and confirm that the arrival of a garment factory in a district is associated with a six percentage point increase in women's garment employment and a five percentage point increase in their wage employment relative to men. I also demonstrate that women in garment districts work close to two more hours more per week than women in districts without a factory. I show that these increases are driven by women between ages 17 and 30, and are strongest among those below age 21. These results confirm that the arrival of garment manufacturing in a district has sizeable implications for young women's earning potential and work lives.

Increasing women's earning capacity through introducing garment work opportunities could increase the return to investment in girls' schooling and health relative to boys (Atkin, 2009; Munshi and Rosenzweig, 2006; Heath and Mobarak, 2015; Jensen, 2012). On the other hand, it could also increase the opportunity cost of schooling at ages when factory work is accessible (Atkin, 2016). My findings on educational outcomes confirm the presence of both mechanisms. I show that there is a sizeable increase in girls' educational attainment in affected districts, both in absolute terms and relative to boys. I find that these gains are driven predominantly by a significant increase in enrollments in primary schooling, while attendance in secondary schooling, when the opportunity costs rise (Atkin, 2016), actually decreases.

The next results suggest that garment manufacturing jobs have contributed to the decrease in fertility and early marriage of Cambodian women (National Institute of Statistics, 2015). I use retrospective marriage and fertility data to present evidence that garment exposure delayed marriage and childbearing for women (but not for men) in garment districts. The fact that these effects are strongest within the age range 17 to 21, which coincides with the ages when women are most likely to begin garment employment, along with the finding that schooling increases only up to age 10, shows that it is garment work, and not the decision to remain longer

[1]The Multi Fiber Agreement, which governed international trade of textiles and garments since 1974, expired at the end of 2004, removing quotas for developing countries.

in school, that drives these results. These findings are in accordance with previous works analyzing the effects of female employment on family formation (Becker 1973; Caucutt, Guner and Knowles, 2002).

Next, I focus on the impact of garment arrival on girls' and women's health. I show that an additional year of exposure to a garment factory in their district is associated with gains in height for girls relative to boys (ages 0-5) and adult women (ages 15-49) with such exposure in early childhood. To my knowledge, with Atkin (2009) this is the only paper to provide a rigorous analysis on the effect of garment employment opportunities on health outcomes. While Cambodia has achieved notable success in reducing the prevalence of stunting by nearly half between 2000 and 2014, 32 percent of children aged 0-5 remain stunted (DHS Cambodia 2000-2014). In line with Atkin's (2009) findings of positive height effects for girls in Mexico, I show that among children aged 0-5, girls' height-for-age increases by between a sixth and a fifth of a standard deviation. Limiting the sample to children born before the factory opening, I show that these results are not driven by changes in marriage or fertility preferences following a factory opening. I also present evidence that these early gains likely persist through adulthood, thus contributing of the literature on the importance of early investments on adult outcomes (Case and Paxson, 2008a; Smith, 2009; Schoellman, 2016).

Lastly, I examine a number of potential non-unitary household channels which, in conjunction with higher returns to investing in girls, could be driving the effects on children's health and early schooling investments. In line with previous work on women's employment and empowerment (Sivasankaran, 2014; Kabeer, 2002; Hewett and Amin, 2000), I find that access to garment employment opportunities results in an increase in women's decision-making power over household spending and gives them control over a larger share of household income. A number of studies have shown that women have a stronger preference for spending on child goods than men (Thomas, 1997; Behrman, 1997), which means that expanding their access to resources or strengthening their say on spending may increase children's schooling or health outcomes. Additionally, women have been shown to have a preference for spending on female children (Duflo, 2003), which would explain why I find increases in young girls' but not boys' height. I find that these empowerment gains are concentrated among women that were already married at the time of the factory opening, which rules out the possibility that they are driven by the increases in women's education (Strauss and Thomas, 1995) and changes in marriage matches. In fact, I show that changes in the marriage market following garment exposure appear to match women with relatively more educated men who earn at least as much as them. The welfare implications of this for both women and children could go in either direction (increased consumption versus less equitable spending), which is why it is important for future research to bring improved understanding of the long-run consequences of garment manufacturing on marriage markets.

The rest of the paper proceeds as follows. Section 1.2 provides background on the garment industry in Cambodia and the specific ways in which it affects work, schooling, and marriage decisions. In section 1.3 I describe the data and in section 1.4 I outline the empirical strategy. Section 1.5 presents the results and section 1.6 concludes.

1.2. Background of the garment industry in Cambodia

The garment industry in Cambodia originated in the mid 1990s with a cluster of factories around the capital Phnom Penh. The explosive growth of the industry and the geographic expansion of production across the country took off in the early 2000s following trade liberalization (Younus and Yamagata, 2012). Starting with less than 50,000 workers in 1995, Cambodian garment factories employed more than 700,000 workers in 2017 (see figure 1.1). Data show that, by 2016, more than 80 percent of the garment work force was outside of Phnom Penh, and nearly 50 percent was outside of its neighbouring provinces (ILO, 2018). Figure 1.2 displays the geographic expansion of registered garment factories between the four waves of the Cambodian DHS survey.

The abundance of cheap low-skilled labour force makes developing countries like Cambodia particularly attractive outsourcing destinations for the labour-intensive final assembly of textiles based on imported materials (also referred to "cut, make and trim" production) (ILO, 2015). More than 80 percent of workers in Cambodian garment factories are women (ILO, 2017). Women's lower wages, perceived advantage in performing repeated manual tasks, and lower propensity to unionize makes them the preferred choice for garment employers (Standing, 1999; Fontana, 2003).

The rapid growth of the garment industry, combined with the overwhelming employment of female workers has the potential to affect women's lives in a number of ways. Firstly, the garment industry predominantly hires young women, who would otherwise be beginning marriage and childbearing or finishing their education. Figure 1.5 shows how much younger the garment labour force is compared to workers in other industries. More than 70 percent of Cambodian garment workers are younger than 30 and more than 50 percent are younger than 25.

Figure 1.6 shows that garment workers are a lot more likely to have at least some schooling compared to workers in other industries. Nearly 50 percent have between five (primary completion is at six) and eight (lower secondary completion is at nine) years of schooling. Workers with a few years of schooling are easier to train and more productive, which is commonly rewarded within factories (Salinger, 2006). At the same time, the garment industry could also affect school enrollment decisions through the opportunity of better (supervisory) jobs within the factory. These positions usually require some to complete secondary education (Salinger, 2006). Nonetheless, supervisory positions are few and more than 95 percent of jobs in the sector are in manufacturing (ILO, 2018).

Wages in the garment sector are competitive on the Cambodian labor market, and provide women who would otherwise work in agriculture with 50 percent higher hourly earnings (figure 1.7). In fact, the garment sector is the only one in Cambodia with a mandatory minimum working wage, which is adjusted upwards periodically (ILO, 2018). Having started work in a factory, women also have an incentive to remain there, as they begin receiving a monthly seniority bonus in their second year of employment, which increases linearly with every subsequent year (for up to 11 years) (ILO, 2018). Therefore, once established, access to garment jobs is expected to have long-lasting impacts on young women's wage employment. This is likely to affect their timing of marriage and childbirth. Through this delay or through access to wage employment itself, garment jobs could increase women's decision-making power within the household directly, by giving them control over a higher share of household resources, or indirectly - by providing them with an outside option (Atkin, 2009). Mothers' bargaining power and increased schooling, combined with the increased return to investment in girls have the potential to positively affect children's, and especially girls' (Duflo, 2003), nutrition and long-term well-being.

1.3. Data

The main analysis in the paper uses establishment year and district data on all 517 garment factories registered with the Cambodia Ministry of Commerce by 2014. In Cambodia, in order to export, companies need to be registered with the Ministry, which means that the factory data consists of all garment exporters. As I discuss in the next section, unregistered subcontracting factories are common, but locate close to registered exporters.

I link the factory data with the four available waves of the Cambodian Demographic and Health Survey, which contain information on educational attainment, school attendance for more than 270,000 individuals, marriage, fertility and labor supply for nearly 100,000 women and men aged 15-49, and health outcomes for 16,000 children below age 5. Crucially for my analysis, while there is random displacement of coordinates on the primary sampling unit level of the Cambodia DHS, this displacement is confined within district limits, which means that I am still able to link factories with the correct district (Burgert, Colston, Roy, and Zachary, 2013). This leaves me with a district-level panel over the 2000, 2005, 2010 and 2014 waves of the Cambodian DHS. I complement my analysis by linking the factory locations with more detailed labor supply and wage data from the 2010-2016 annual waves of the Cambodian Socio-Economic Survey (CSES), which is another nationally-representative survey with a stronger focus on employment.

Table 1.1 contains summary statistics of individual and village-level[2] characteristics for garment (prior to factory arrival) and non-garment districts. Panel A shows that educational attainment for both men and women was higher in districts where garment factories opened. Age at first birth is slightly higher for women in garment districts as well, while there is no significant difference between height in centimeters and age at marriage. Turning to village characteristics, there is no significant difference between distance to national borders or population density (even if the difference in population density is sizeable). The most notable difference between villages in garment and non-garment districts is in their proximity to cities with population above 100,000 people [3]. As discussed in the next section, this is due to the fact that minimizing transportation and electricity costs is a major determinant of garment factory locations. The closer proximity to large cities could also explain why educational attainment was higher for both men and women in garment districts, as there were likely more non-agricultural jobs available to begin with.

[2]Village and primary sampling unit are used interchangeably.

[3]In the 2008 Census, there were eight of them: Phnom Penh, Ta Khmau (Kandal), Serei Saophoan (Beantey Meanchey), Battambang, Siem Riep and Kampong Cham, all located near the major trade routes shown in figure 1.3

1.4. Empirical strategy

1.4.1. Overview

I exploit the variation in timing and location of garment manufacturing across Cambodia (figure 1.2). Since garment districts are different from non-garment ones in terms of fixed characteristics and in ways that could trend differently by gender (table 1.1), I allow for time-invariant district-level differences, and include gender-specific district time trends, which control for different trends in key outcomes, including by gender, between districts with and without garment openings. I also include gender-province-year[4] fixed effects, which flexibly control for the fact that these outcomes may trend differently by gender across Cambodian provinces over time. With these steps I alleviate the concern that my results are driven by selection of factories based on time-invariant or time-varying differences in characteristics between garment and non-garment locations. Wherever possible, I use additional control groups, including males in the same district, and individuals within the same district who were not exposed at key ages. To account for linear growth (or decline) in enrollments, years of schooling, health outcomes and marriage patterns across age-cohorts and genders within districts, I incorporate district-gender-birth-year trends.

Akin to other major garment exporting countries (Heath and Mobarak, 2015), in Cambodia garment factory location choices are driven by minimizing transportation and electricity costs, which makes reverse causality unlikely. The uneven access and high costs of electricity (World Bank, 2013), along with the reliance on truck transport to major ports (World Bank, 2014), makes it highly improbable that factories would choose where to locate based on population characteristics. Indeed, there are clear agglomerations of factories near roads and major cities over time in Cambodia (World Bank, 2014; figure 1.2). Figure 1.2 displays the expansion of factories over the four waves of the Cambodia DHS and figure 1.3 contains the main land transport corridors to deep sea ports passing through the country. They confirm that garment growth to the south, east and north of Phnom Penh was taking place in proximity to the country's major export routes. As previously discussed, I control everywhere for fixed baseline differences associated with these non-random location choices, as well as for time-varying trends across districts by gender. Furthermore, I perform placebo analyses and show that my results are not driven by different pre-trends across garment and non-garment districts (table 1.12).

[4]After districts, the 25 provinces are the next largest administrative units in Cambodia.

The findings of this study rely on the assumption that proximity to a factory is a key determinant of garment job opportunities. While garment commutes are common, 95 percent of Cambodian workers commute within the same district (Cambodian Population Census, 2008). Indeed, figure 1.8 shows that women in garment districts are approximately six times more likely to be employed in a garment factory, compared to women in untreated districts. To further corroborate this, information about garment jobs is likely less accessible to women in districts located further from a factory (Jensen, 2012). Similarly to Heath and Mobarak (2015), throughout I use a binary variable denoting the access to a factory for. For the majority of districts (80 percent) with factory openings between 2000 and 2014, this was the first and only factory in the district that operated within the period. In addition, with the exception of districts in and around Phnom Penh, which were treated before 2000, the average district with a first factory opening had a population of fewer than 100,000 people [5], while the average factory size of factories that opened since 2000 has been estimated at 1200 workers [6]. This translates into a significant shock to local labour markets and women's employment: an average of at least 600 garment jobs (not considering subcontractors) per 10,000 women aged 15-30. Therefore, this estimation strategy allows me to capture both the direct and indirect effects associated with the arrival of garment manufacturing in a district, discussed in section 2.

Work-related migration is an important concern for the unbiasedness for my results. The confounding effect of migration could go in both directions. Healthier and more educated women are more likely to migrate to a district following the opening of a factory, resulting in a positive coefficient of factory exposure. Similarly, women who are less likely to marry or have children early would be more prone to migrating across district borders. On the contrary, as factories are located mostly in urban areas, women migrating from less urbanized localities may have lower educational attainment or worse nutritional status. Even though the latest waves of the Cambodian DHS do not contain individuals' migration history, I am able to address this concern. I use the fact that migrants commonly arrive and live alone, and limit my sample to individuals who live in households with at least one member who is not a spouse and who was born before the opening of the factory in treated districts. According to the Cambodian Rural-Urban Migration Project, approximately 75 percent of urban migrant women's households consti-

[5] Data from the World Bank DataBank shows that Cambodia has much smaller population density than other major garment producers: as of 2018 it is estimated at 92 people per km^2, more than three times less than Vietnam, five times less than India, and more than ten times less than Bangladesh.

[6] Data on current employment in a subset of factories registered with the Garment Manufacturers Association of Cambodia (GMAC).

tute of a single member, and 67% of migrant women garment workers have never been married (Kheam and Treleaven, 2013). Moreover, multi-generational and extended families represent about 70 percent of households' living arrangements and unmarried non-migrant men and women would usually live with their parents (Heuveline and Hong, 2016). By limiting my sample following this procedure, I exclude 9 percent of households. This number fits with estimates of migrant population in urban centres from the 2008 Census of between 6 percent and 24 percent (Ministry of Planning, 2012). The average number of household members thus becomes 6.03 , compared to 5.9 in the original sample. I estimate all but the labor supply regressions using this sample definition. In the case of the labor supply model, I allow for the effect of migration to capture the overall employment effects in garment districts.

After controlling for gender-specific district trends, ruling out reverse causality and mitigating the threat posed by migration, the remaining major identification concerns are measurement error and omitted variables correlated with both the dependent variables and the introduction of garment manufacturing. Measurement error stems from the fact that my dataset contains only garment manufacturers registered with the Ministry of Commerce, omitting any unregistered factories. In Cambodia, in order to export, factories need to be registered with the Ministry, which means that my analysis uses all formal exporters. To minimize costs, subcontracting entities would normally locate near the larger exporting factories, which diminishes the probability of arrival of omitted factories in control districts. If such cases do exist, they would be biasing my results towards zero, and in this scenario my findings would be interpreted as lower bounds of the true effects of garment manufacturing.

A remaining concern for the interpretation of the results is presented by any omitted variables, such as the construction of schools, new roads or hospitals, correlated with both the arrival of garment factories and education, health and marriage. I argue that several aspects of the results make it unlikely that such omitted factors could be explaining the whole set of results. To begin with, I find much stronger effects on girls' heath and schooling than boys, so the impacts of any infrastructure investment accompanying garment factories would need to be gender-specific. Even if this is the case, the fact that the results on school attendance are positive for younger girls, and negative for older girls, also goes against such a possibility, since children in Cambodia typically have to travel further for secondary school than primary school. Similarly, there is only an effect on girls' marriage and childbearing outcomes, and no effect on boys. This makes it unlikely that the results are driven by investments in garment districts that would have changed the marriage market as a whole, or changed the cost of having children. Therefore, while each result could be related to a specific omitted variable individually, it is difficult to imagine omitted variables that can explain the entire set of different results on males and females.

1.4.2. Labor supply

I begin by confirming that the presence of a factory increases the probability of women working in the garment sector, being in wage employment, and working longer hours, relative to men. Because of missing information for men in the earlier waves of the DHS, I use the seven yearly waves of the Cambodian Socio-Economic Survey (2010-2016) to examine employment and wage work status, as well as the number of hours worked per week. I also compare estimates for the effect on garment employment from both datasets. I estimate:

$$\begin{aligned} \textit{LaborOutcome}_{idt} &= \alpha_1 \boldsymbol{G}_{dt} + \alpha_2 \boldsymbol{G}_{dt} \times Female_{idt} \\ &+ \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_{df} t + \delta_{pft} + \epsilon_{idt}, \end{aligned} \tag{1.1}$$

where G_{dt} is equal to 1 if there was a garment factory in district d in year t. The coefficient α_2 represents the labor supply effect that factory exposure in their districts has on women relative to men within treated districts versus women relative to men in non-garment districts. To account for time-invariant differences across treated and untreated districts, I include district fixed effects, δ_d. I also include wave fixed effects, δ_t/. The district-level panel nature of the data allows me to account for the fact that outcomes trended differently across districts and by gender. To this end I include district–gender-specific time trends, $\delta_{df}t$. I also flexibly control for trends across provinces (which are the next largest administrative unit), by including gender-specific province-year dummies, δ_{pft}. The vector X_{idt} includes dummies for individuals' age and ethnicity, as well as primary-sampling-unit-level information denoting whether they live in an urban area, distance from national borders, travel time to largest city and population density.

If younger women are more likely be employed in garment manufacturing (Heath and Mobarak, 2016; figure 1.5), then the effect of the opening of a garment factory in a district is likely to vary by age. I estimate equation 1.1 using interactions with age dummies $\mathbb{1}(Age_{idt} = a)$, where $a = 15, 16, ..., 49$. The equation in this case is:

$$
\begin{aligned}
&\textit{GarmentWork}_{idt} = \alpha_1 \boldsymbol{G}_{dt} + \alpha_2 \boldsymbol{G}_{dt} \times Female_{idt} \\
&+ \sum_{a=15}^{49} \beta_{a,g} \boldsymbol{G}_{dt} * \mathbb{1}(Age_{idt} = a) + \sum_{a=15}^{49} \beta_{a,g,f} \boldsymbol{G}_{dt} * \mathbb{1}(Age_{idt} = a) * Female_{idt} \\
&+ \sum_{a=15}^{49} \beta_{a,f} \mathbb{1}(Age_{idt} = a) * Female_{idt} + \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_{df} t + \delta_{pft} + \epsilon_{idt},
\end{aligned}
\tag{1.2}
$$

where $GarmentWork_{idt}$ is equal to 1 if individual i in district d is employed in garment manufacturing in year t.

1.4.3. Education and school enrollment

I examine the impact of the garment industry on girls' educational attainment by considering the effects of the number of years of exposure to a garment factory between ages 5 and 18. These are the ages in which most students would complete primary and secondary education, and are thus the most relevant age group for this analysis (figure 1.6). In addition to district-gender time trends and province-gender-year fixed effects, I allow for increasing educational attainment across cohorts by including gender-specific district birth-year trends $\delta_{df} b$. I estimate:

$$
\begin{aligned}
&\textit{EducationYears}_{idt} = \alpha_1 YearsExposure_{idt} \\
&+ \alpha_2 YearsExposure_{idt} \times Female_{idt} + \alpha_3 \boldsymbol{G}_{dt} \times Female_{idt} + \alpha_4 G_{dt} \\
&+ \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_{df} b + \delta_{df} t + \delta_{pft} + \epsilon_{idt}
\end{aligned}
\tag{1.3}
$$

where $\textit{EducationYears}_{idt}$ stand for individuals' years of schooling and $Years$ $Exposure_{idt}$ represent the number of years between ages 5 and 18 in which they had a garment factory in their district.

Since, in addition to expectations for future employment, contemporaneous exposure to garment employment is of significant importance for current-year school attendance, especially for older students, I estimate the effect of factory openings on girls' school attendance by using the binary variable G_{dt}, equal to 1 if there is a factory in district d in year t and interact it with gender using the sample of individuals aged between 5 and 18. As we expect the effect on enrollment to vary with age (Atkin, 2016; Heath and Mobarak, 2015), I include interactions with age dummies $\mathbb{1}(Age_{idt} = a)$, where $a = 5, 6, .., 18$. To allow for gender-specific linear increase in the enrollment rate within districts I once again include gender-specific district birth-year trends $\delta_{df}b$. As before, I include province-gender-year fixed effects and gender specific district time trends.

$$
\begin{aligned}
&\mathit{Enrolled}_{idt} = \alpha_1 \boldsymbol{G}_{dt} + \alpha_2 \boldsymbol{G}_{dt} \times Female_{idt} \\
&+ \sum_{a=5}^{18} \beta_{a,g} \boldsymbol{G}_{dt} * \mathbb{1}(Age_{idt} = a) + \sum_{a=5}^{18} \beta_{a,g,f} \boldsymbol{G}_{dt} * Female_{idt} * \mathbb{1}(Age_{idt} = a) \\
&+ \sum_{a=5}^{18} \beta_{a,f} Female_{idt} * \mathbb{1}(Age_{idt} = a) + \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_{df}t + \delta_{df}b + \delta_{pft} + \epsilon_{idt}
\end{aligned}
\tag{1.4}
$$

Note that the vector $\mathbf{X}_{idt}$ everywhere contains age dummies, as well as other individual- and village-level characteristics defined in section 1.4.2.

1.4.4. Marriage and childbearing

I use the information on the date of men and women's first marriage and first birth to estimate a discrete-time hazard model of the probability of marriage and first birth on years of garment exposure. This allows me to capture both the contemporaneous effect of garment work opportunities and the longer-term effect of additional years of education (where long-run exposure matters) (Heath and Mobarak, 2015). I use retrospective marriage and birth history data from the 2014 DHS wave (which includes marriage and fertility data for both women and men), tracing individuals from birth until their year of marriage or first birth. The dependent variable takes on a value of 0 in periods when an individual is not married (or has not had any children), and 1 in the year when that event occurs, after which they exit the sample. Never married observations or those without children take on a value of 0 in all periods. Here, instead of DHS survey wave, t stands for the year in which women and men were of a certain age. I estimate the following equation:

$$\mathit{Married}_{idt} = \alpha_1 \mathit{YearsExposure}_{idt} + \delta_d + \delta_t + \delta_d t + \delta_{pt} + \epsilon_{idt} \tag{1.5}$$

$\mathit{YearsExposure}_{idt}$ represents the number of years since birth in which an individual in the sample was exposed to a garment factory in their district until year t. $\delta_d t$ are district-level time trends and δ_{pt} are province-year fixed effects. The coefficient α_1 represents the effects of an additional year of exposure to garments job on the probability that a woman gets married or begins childbearing in a given year. In this part of the analysis I estimate the models separately for men and for women, as, in the absence of changes on the marriage market, it is likely for garment exposure to affect men and women similarly in this respect.

As the pottential marriage and fertility delays are likely to be stronger for younger individuals, I also estimate the marginal effects by age:

$$\begin{aligned}&\mathit{Married}_{idt} = \alpha_1 \mathit{YearsExposure}_{dt} \\ &+ \alpha_2 \mathit{YearsExposure}_{dt} \times Age_{idt} + \alpha_3 \mathit{YearsExposure}_{dt} \times Age^2_{idt} \\ &+ \delta_d + \delta_t + \delta_d t + \delta_{pt} + \epsilon_{idt},\end{aligned} \tag{1.6}$$

where Age_{idt} is the age of individual i in district d in year t of the retrospective sample.

1.4.5. Health

Children

I begin by examining the link between garment exposure and height-for-age, weight-for-age and anemic status of female and male children aged 0-5. To isolate the health effects of mothers' garment employment from the consequences of delayed marriage and fertility, I re-estimate (1.7) below limiting the sample to children who were already born at the time a garment factory opened in their district. Instead of dummies for age in years, in the regressions of children's health outcomes the vector $\mathbf{X}_{idt}$ includes dummies for children's age in months along with other previously described characteristics.

$$\begin{aligned}&\mathit{HealthOutcome}_{idt} = \alpha_1 \mathit{YearsExposure}_{idt} + \alpha_2 \mathit{YearsExposure}_{idt} \times Female_{idt} \\ &+ \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_{df} t + \delta_{df} b + \delta_{dfp} + \epsilon_{idt}\end{aligned} \tag{1.7}$$

Here $YearsExposure_{idt}$ represents the number of years in which a child in district d and year t was exposed to a garment factory in their district during their lifetime. The remaining variables are as defined above. The coefficient α_2 represents the effect of an additional year of garment exposure on girls relative to boys. Netting out the potential effects of garment manufactories on mothers' schooling, marriage and fertility, we expect this coefficient to be positive if garment opportunities favor redistribution of resources to girls, give mothers additional access to resources or increases their bargaining power within the household.

Women

Unfortunately, data for children aged 5-15 is not collected by the DHS; the survey does, however, contain health information for women aged 15-49. I analyze the long-term effects of garment exposure on women's adult height by examining the effects of garment exposure at different ages. A sizeable body of literature suggests the importance of investments in early childhood for adult outcomes (Currie and Vogl, 2013; Case and Paxson 2008a; Schoellman, 2016). That is why, we could expect the effect of an additional year of exposure on adult height to decrease with age of exposure. I estimate the following model:

$$\begin{aligned} \mathit{Height}_{idt} &= \alpha_1 \mathit{YearsExposure}_{dt} \\ &+ \alpha_2 \mathit{YearsExposure}_{dt} \times AgeExposed_{idt} + \alpha_3 \mathit{YearsExposure}_{dt} \times AgeExposed^2_{idt} \\ &+ \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_d t + \delta_d b + \delta_{pt} + \epsilon_{idt}, \end{aligned} \tag{1.8}$$

where $AgeExposed_{idt}$ is equal to the age at which the woman was first exposed to a garment factory in her district and the remaining variables are as previously defined.

1.4.6. Mechanisms

To identify the drivers behind the observed increases in investment in girls' health, and to evaluate to what extent the shifts in the marriage market played a role, I test a number of different outcomes related to women's bargaining power within the household: the likelihood that they earn more than their husband, the degree to which they can decide on household purchases, as well as the difference in schooling and age between them and their husband. Unsurprisingly, all of these variables are available for married women only. The first two could be directly positively affected by women's increased earning opportunities or their higher income (Friedberg and Webb, 2006). The latter group could be an indirect (negative or positive) consequence of the introduction of garment manufacturing, either through women's change in earnings and schooling (assortative matching, for instance), or through altering the pool of potential matches (through a delay in marriage or through exposure to different people) (Atkin, 2009).

To alleviate the concern of omitted variables related to selection of garment districts, I compare women most likely to be affected when a factory was introduced to women who were older at the time of the factory arrival. Figure 1.9 shows that the probability of garment employment increases most for women younger than 30. That is why, I compare women who were younger than 30 at the arrival of a factory with women who were older than this threshold. I also use an alternative threshold of 28, which is the age by which 95 percent of the women in my sample marry. Hence, the group of women who were exposed at ages 30 (or 28) and older were a lot more likely to be already married and a lot less likely to start working at a factory. I also examine whether these findings are driven by the subsample of women who were already married at the time of factory arrival, or whether they were present for women who married after. I estimate the following equations:

$$\begin{aligned} \textit{DecisionOutcome}_{idt} &= \alpha_1 G_{dt} \times Exposed_{y0idt} + \alpha_2 G_{dt} + \alpha_3 Exposed_{y0idt} \\ &+ \mathbf{X}'_{idt}\gamma + \delta_d + \delta_t + \delta_d b + \delta_d t + \delta_{pt} + \epsilon_{idt}, \end{aligned} \tag{1.9}$$

where $Exposed_{y0idt}$ denotes exposure to a garment factory in the district before the respective age threshold $y0$. The coefficient α_1 measures the difference between the effect that garment exposure had on women who were most likely to be affected versus women who were less likely to be affected by a factory opening within a district.

1.5. Results

1.5.1. Labor supply

Table 1.2 confirms that women in districts with a garment factory are five percentage points more likely to be employed in the sector, controlling for district fixed effects and time trends . These results are almost identical in the DHS and the CSES samples, regardless of the different time-periods covered. This increase in garment employment is accompanied with a proportionate increase in women's wage employment relative to men (column (3)).

Given that the share of garment employment in the DHS sample was about five percent (column (1) in table 1.2) and the average share of wage employment in the CSES sample was 39 percent, these results represent more than a 100 percent and 10 percent increases in garment and wage employment respectively. This confirms that the location of a factory within a district is a major shock to the local labor market, which significantly expands women's wage employment opportunities. In addition, relative to men, women in garment districts work nearly two hours more per week, compared to women in non-garment districts. This is consistent with reports that garment workers often have to work overtime (Better Factories Cambodia, 2020).

It is also important to note that there is no associated change in women's employment. Unlike India and Bangladesh, which have been the focus of the majority of works on the garment industry (Sivasankaran, 2014; Heath and Mobarak, 2015, Kabeer, 2002), Cambodia has had continued high female labor force participation, which was more than 76 percent in 2017 based on latest national estimates from ILOSTAT. Therefore, the access to garment employment in this setting represents a shift towards wage work, without changing the probability of employment.

Figure 1.9 shows that the arrival of garment manufacturing in a district affects predominantly young women, with the effect being strongest among women aged 17 to 24, and becoming insignificant after age 30. This indicates that garment employment could have significant implications for marriage and childbearing decisions discussed in section 5.3.

1.5.2. Educational attainment and enrollment

I begin examining the effects of expanding women's wage employment opportunities by focusing on years of schooling. In this section, I define years of exposure as the number of years in which an individual was exposed to a garment factory between ages five and 18, which are the years in which people usually would attend primary and secondary school in Cambodia. Columns 2 and 3 of table 1.3 show that an additional year of garment exposure had a significantly positive effect on girls' schooling in both absolute terms and relative to boys. Column 2 displays results for the entire sample, while column 3 reports the results from the estimation on a sample of men and women who are currently older than age 18 and are thus more likely to have completed their education. Similarly to Heath and Mobarak (2015), I find that the absolute positive effect on girls' schooling is stronger for younger individuals (column (2)). I also find a negative effect on boys' schooling in the cohort with complete schooling (column (3)) and a significantly lower magnitude of the positive effect on girls' schooling in absolute terms. The negative coefficient on boys may point to redistribution of resources between boys and girls within households at a time when primary schooling was less accessible. The mean years of exposure in garment districts is approximately four years (six for the older sample), which would translate to a sizeable increase in girls' educational attainment ranging from 0.60 in the older cohort to 2.08 years in the younger one. These results are comparable in magnitude to Heath and Mobarak (2015), who find an increase between 0.92 and 3 years in Bangladesh.

I next turn to school attendance.The first column of table 1.3 shows the overall effects of factory exposure on school enrollment in a sample of five- to 18-year-olds. We do not observe an aggregate increase in enrollment. Nevertheless, the age-specific estimates displayed in figure 1.10 show that this insignificant coefficient masks the heterogeneous effects of garment presence by age. There is a significant and sizeable positive effect of garment exposure on the probability of girls' enrollment in primary schooling, which decreases and becomes negative (at

10% level of significance) at age 12. The particular age at which the coefficient becomes negative holds an important meaning, as it is at age 12 that Cambodian students are expected to start lower secondary schooling. This indicates that, while the presence of a garment factory in a district significantly increases girls' primary enrolments, it also reduces their probability of transitioning into secondary education.

The negative effect on enrolments is strongest at age 15, which is the minimum working age with parental consent in Cambodia. While the minimum working age is higher than 12, minors aged as young as 12 could still legally hired to perform light work (ILO and National Institute of Statistics, 2014). In addition, recent reports show that violations of the minimum working age (by presenting fake documents, for instance) are not uncommon, especially outside of Phnom Penh (United Nations Refugee Agency, 2017). These all point to an increase in the opportunity cost of schooling once primary education is completed, and are the likely explanation for the significant drops in attendance at the secondary school level.

1.5.3. Marriage and childbearing

I next turn to examining the effects of improved access to garment employment on delays in marriage and childbearing. Tables 1.4 and 1.5 report the results of discrete time hazard models that examine the effects of years of exposure to garment factories (up to a certain year) on the hazard of a girl getting married by that year, or the hazard of the girl giving birth to her first child by that year.

Exposure to a garment factory is defined as the number of years a girl has lived in a district with a garment factory since the first factory opened in the district. Starting with marriage, the first column of table 1.4 indicates that women living in garment districts, where factories have been operating for 6.6 years (the average exposure in garment districts) have a (6.6×0.00063) = 0.4 percentage point lower probability of getting married by that year relative to women living in control districts in the same province. As the probability of marriage in an average sample year is 3.5 percent, this 0.4 percentage point decrease represents an 11 percent decrease in the hazard of getting married.

The second column in table 1.4 delves into whether this delay in marriage varies by age by adding interaction terms between the years of garment exposure by a given year, and the woman's age in that year. Figure 1.11 plots the marginal effects from this estimated regression. It shows that exposure to the garment sector has significant negative effects on the probability of marriage up until age 25. The effect size is the largest at ages 17 to 21, which coincide with the ages in which women are most likely to be employed in the garment sector following the opening of a factory (see figure 1.9). This result, combined with the finding in the previous section that girls were actually less likely to be in secondary school in districts with garment openings (figure 1.10), indicates that it is likely the increase in garment employment that drives these results. There are particular aspects of garment work, such as working overtime and commuting away from home (versus work on the household farm, for instance), which would make it incompatible with early marriage and childbearing in the Cambodian setting.

Turning to the hazard of childbearing, column 1 in table 1.5 shows that women living in garment districts with 6.6 years of factory exposure are (6.6×0.0005) = 0.33 percentage points less likely to have given birth by that year compared to women living in control districts in the same province. The probability that a woman has her first child during the average sample year is 3 percent, which means that the 0.33 percentage point decrease represents a 10 percent reduction in the hazard of having their first child. The second column in table 1.5 confirms that there was variation in this effect across different age groups. Figure 1.12 displays these heterogeneous effects. Similarly to the postponements in marriage described above, the delay in childbearing was strongest among younger women aged 17 to 21, which also confirms that the effects of garment exposure on marriage and childbearing are linked.

The last two columns of tables 1.4 and 1.5 show that the garment sector had no significant effect on the marriage or first-birth timing for men. This findings indicate some potential shifts in marriage market dynamics. For instance, men could be more likely to marry outside the district. I explore some of these dynamics in section 5.5, but a more in depth analysis is a worthwhile avenue for future research in order to better understand the long-term welfare implications of garment manufacturing.

1.5.4. Health

I identify the effects of an additional year of exposure to garment manufacturing jobs on children's health by exploiting variation between boys and girls aged 0 to 5 across garment and control districts within provinces over time. Years of exposure here is defined as the number of years in which there was a garmen factory in a child's district since birth until their current age.

Column 1 of table 1.6 shows that an additional year of exposure to a garment factory in their district is associated with increase in girls' height-for-age z-score relative to boys in garment districts compared to control districts within the same province. To establish whether any health effects are associated with delayed fertility or marriage , I re-estimate the model limiting the sample to children who were born before the opening of a factory within treated districts. Column 2 in table 1.6 shows that the results for height-for-age are stronger for children born before the opening of the factory. This indicates that the shifts in marriage and fertility, observed in the previous section are not accountable for these changes. I find that an additional year of garment exposure increases young girls height-for-age z-score by about 22 percent of a standard deviation relative to boys. At 1.3 years of average garment exposure in treated districts for children born before the arrival of factories, this is equivalent to 0.28 standard deviations or 15 percent increase from the mean in the average sample year. Columns 3 to 6 of table 1.6 show that there are no effects on children's weight-for-age z-score or probability of being anemic. These results are policy relevant for a number of reasons. Height-for-age is one of the most important indicators of children's nutrition and early improvements have been shown to have lasting positive effects on human capital development (Case and Paxson, 2008a; Spears, 2012; Steckel, 2009). In addition, the fact that the effect is gender-specific points to intrahousehold distribution channels linked with either the increased perceived return on investment in girls or mothers' higher bargaining power within the household.

I next point my attention to the long-term health implications of garment exposure on a sample of women aged 15 to 49. Years of exposure are defined as the number of years since birth in which a woman has had a garment factory in her district. As early childhood investments are particularly important for individuals' height and long-term health, I also allow the effect of years of exposure to vary with women's age of first exposure to garment manufacturing in the district.

Columns 1 and 2 in Panel A of table 1.7 confirm that garment exposure in early age can have long-lasting positive effects on women's height. Panels A and B in column 2 also show that the significance of garment exposure diminishes with the age a woman was initially exposed. An year of garment exposure between ages 0 to 4 is associated with approximately 0.2 additional centimeters of height. At the mean exposure of children under 5 exposed between these ages of 1.3 years, this would translate to a 0.28 cm gain in adult height. For the average woman in my sample, who is 152.6 centimeters tall, this means a 0.3 percent increase in height. The diminishing strength of garment exposure by age at exposure is also illustrated in figure 1.13. The magnitude of the coefficients and the standard errors indicate that I am likely underpowered in the estimation of the marginal effects, as the largest age-at-exposure group consists of 290 observations. These findings are consistent with the previous result on female children's height-for-age z-scores.

They are also in line with previous research emphasizing the importance of early investments in health on adult outcomes (Case and Paxson, 2008a; Smith, 2009; Schoellman, 2016).

1.5.5. Mechanisms

I find that women's self-reported household decision-making power increases as a result of the introduction of garment employment opportunities. Importantly, the magnitude of the effects on earnings (relative to husband's earnings) and say over household purchases is similar to the coefficient of the increase in garment employment in affected districts. Table 1.8 and table 1.9 also clearly indicate that these gains in bargaining power are concentrated among women who were already with their partner at the time of factory exposure. This indicates that garment manufacturing may introduce changes in the marriage market that limit women's decision-making authority within the household. Indeed, table 1.10 shows that the gap between women and their spouses' education increases by more than 25 percent, and this increase is most likely driven by couples married after the arrival of the factory. I don't find any effects on the difference in age between partners (1.11). The difference in spouses' age and schooling has been shown to be a predictor of women's bargaining power in a marriage (Friedberg and Webb, 2006; Strauss, 1995).

These results, combined with the finding that garment manufacturing does not affect men's timing of marriage, may mean that, at a given age, the fewer women on the marriage market would be able to marry a more educated partner. Garment workers' higher earning potential also indicates that they would match with a higher earning (and likely more educated) partner. The finding that power-dynamics shift only in couples formed prior to the opening of a factory also aligns with the findings that girls' height-for-age increases more for children who were born before the arrival of garments in their district (table 1.6). Therefore, the potential shifts in marriage markets and the long term consequences of garment expansion are an important avenue for future research.

1.6. Conclusion

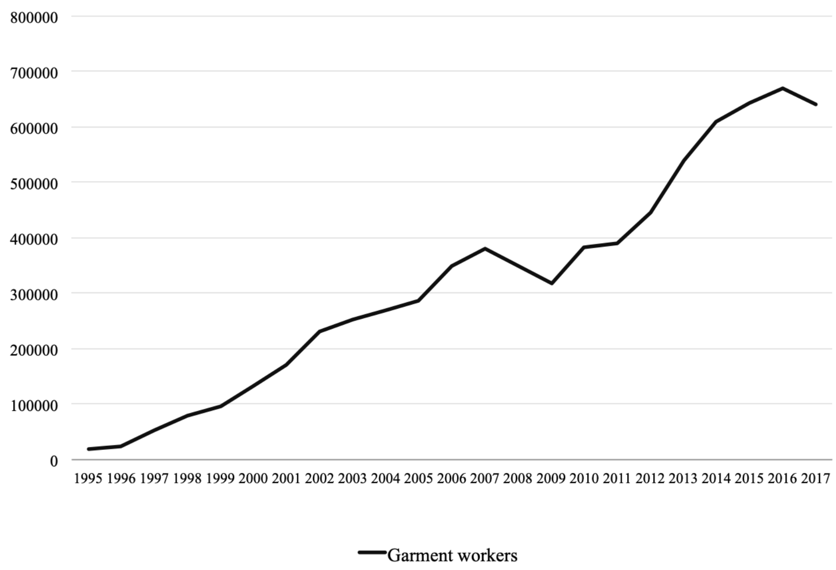
This paper studies the effects the rapid expansion of garment manufacturing over the past two decades has had on the lives of Cambodian women and children. It shows that the industry's explosive growth led to a sizeable increase in women's schooling, delayed marriage and childbearing, and brought about potentially long-lasting height gains for female children. While there is a sizeable positive net effect on girls' schooling, it mainly driven by women who would have obtained little or no schooling in the absence of a factory. At the secondary level of schooling, female enrollments actually decrease as a result of garment exposure. This is crucial, considering the country's great progress in increasing primary completion, and current efforts to meet the United Nations development goal of facilitating access to secondary schooling (World Bank, 2019). I also examine the mechanisms behind these findings. The fact that girls' schooling increases only up to the level required by garment work, means that to a large degree these higher investments in girls' human capital are driven by an increased return to their early education. I also show that the delays in marriage and childbearing are strongest among the age groups most likely to begin work in a garment factory after its arrival in their district. I further present evidence that women's decision making power and access to household resources increase, following a factory opening, which is likely related to the increased investment in girls' health relative to boys. Nevertheless, these empowerment gains do not seem to hold for women married after garment exposure, which raises the question of whether any gains associated with women's increased bargaining power would persist over time.

My findings serve to inform policy on the importance of wage employment opportunities for women's schooling, health, and timing of marriage and childbearing. All of these outcomes are of heightened importance in developing contexts, and the garment sector's potential expansion to poorer Asian and African economies (Younus and Yamagata, 2012; de Haan and Vander Stichele, 2007), makes an improved understanding of the industry's effect on women's lives crucial. Moreover, for countries like Cambodia, which have already experienced decades of garment growth, and that seek to upgrade their place in the global value chains, recognizing the challenges and opportunities presented by the industry is an important step on their way towards joining the middle-income group.

1.7. Appendix

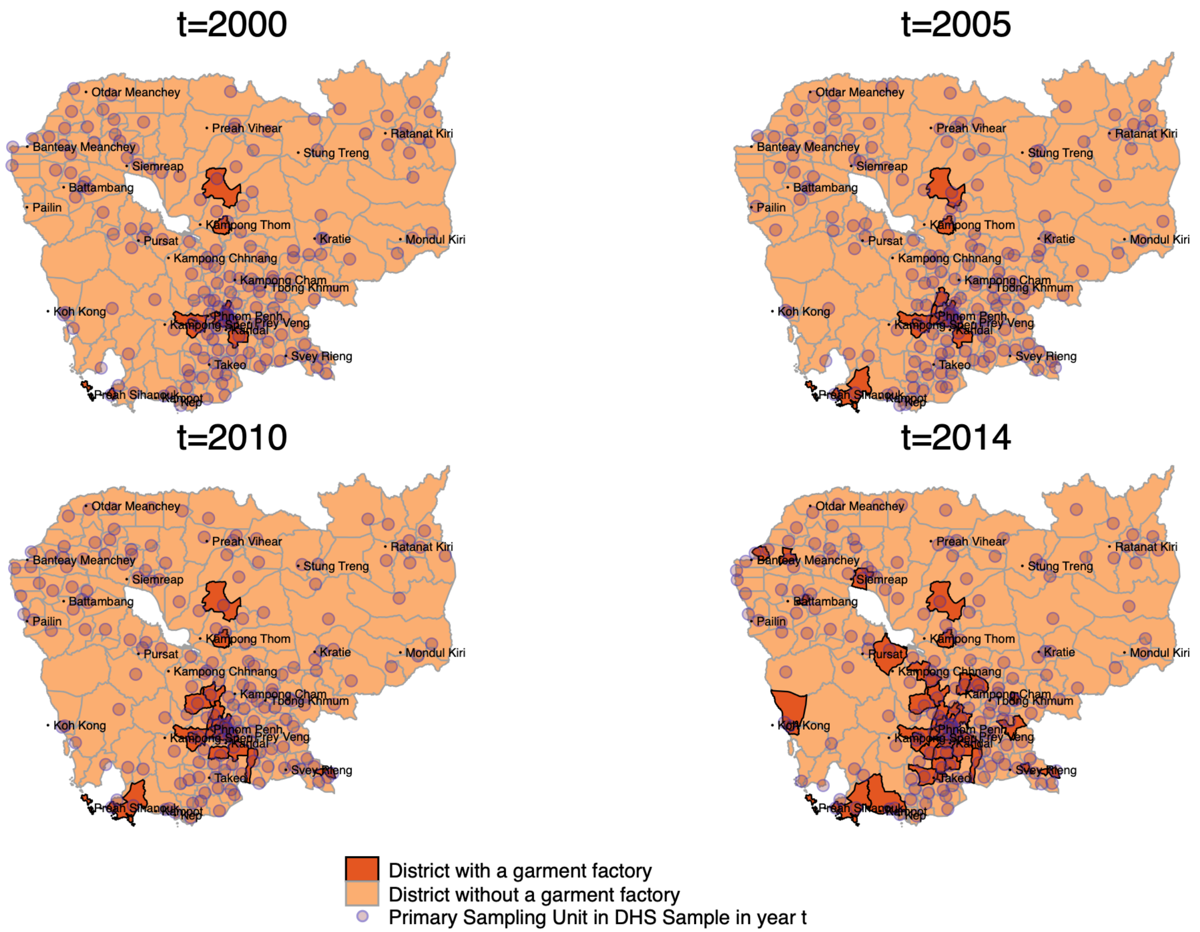
1.7.1. Background and summary statistics

Figure 1.1: Growth of garment employment in Cambodia



Note: Data from ILO's Cambodian Garment Sector Bulletin (2017).

Figure 1.2: Geographic expansion of garment manufacturing in Cambodia

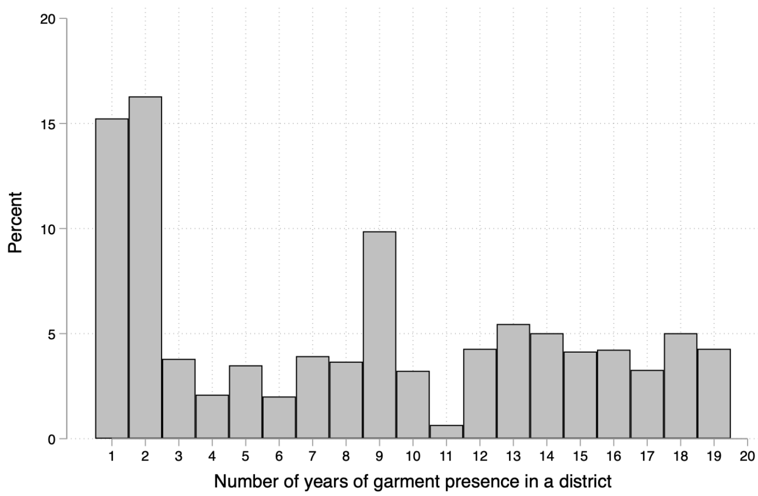


Notes: Data on factory location and year of opening from the Cambodian Chamber of Commerce Registry. Factory and primary sampling units coordinates DHS 2000-2014 merged with administrative district boundaries from Department of Geography of the Ministry of Land Management, Urbanization and Construction (2008). Province names and locations of province capitals are shown as well.

Figure 1.3: Main trade corridors in Cambodia

Notes: Map from Hong Kong Trade and Development Council (2017).

Figure 1.4: Length of factory presence in treated districts

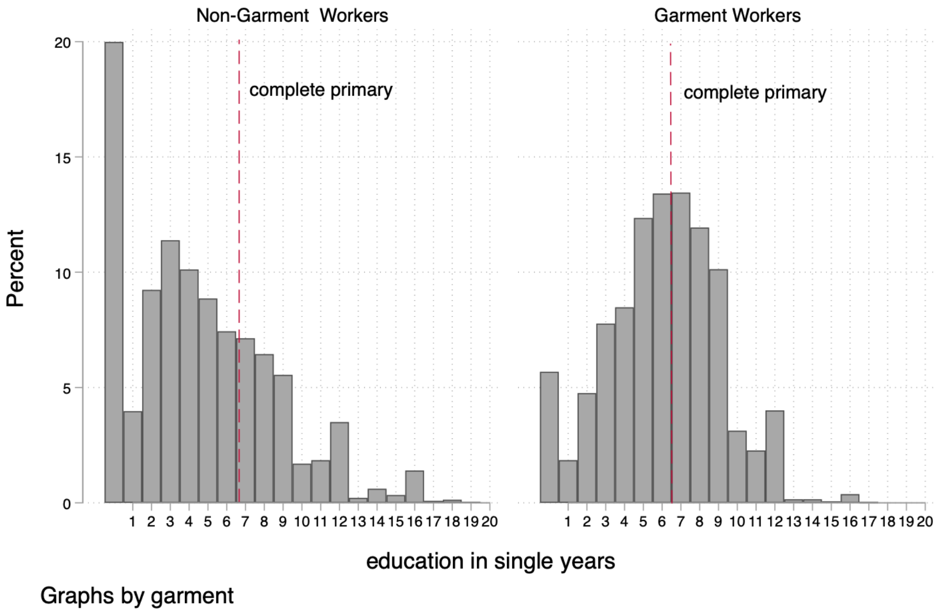


Note: Data from the 2000-2014 Cambodia DHS

Figure 1.5: Garment vs. non-garment workers' age

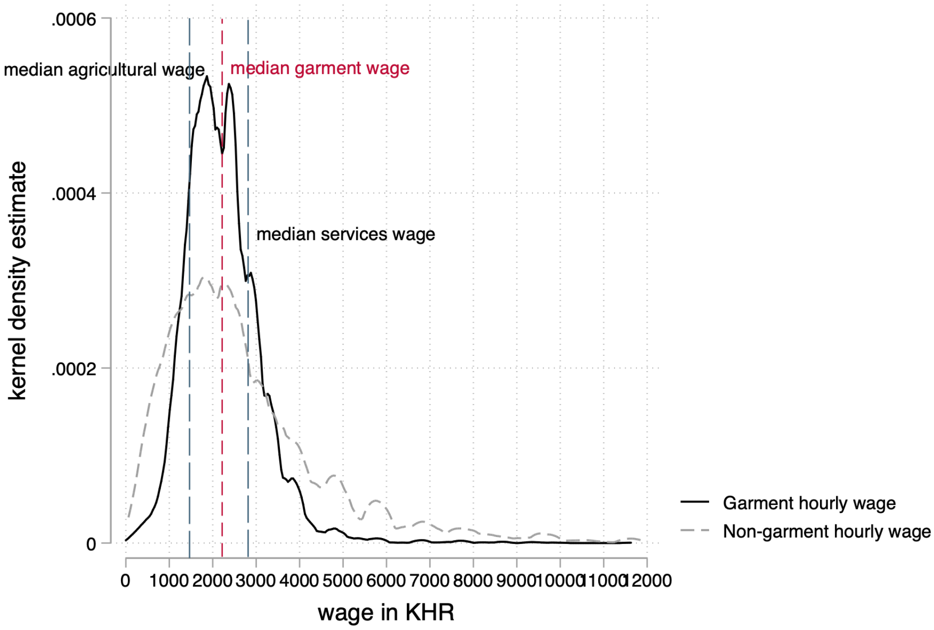
Note: Data from the 2000-2014 Cambodia DHS

Figure 1.6: Garment vs. non-garment workers' educational attainment



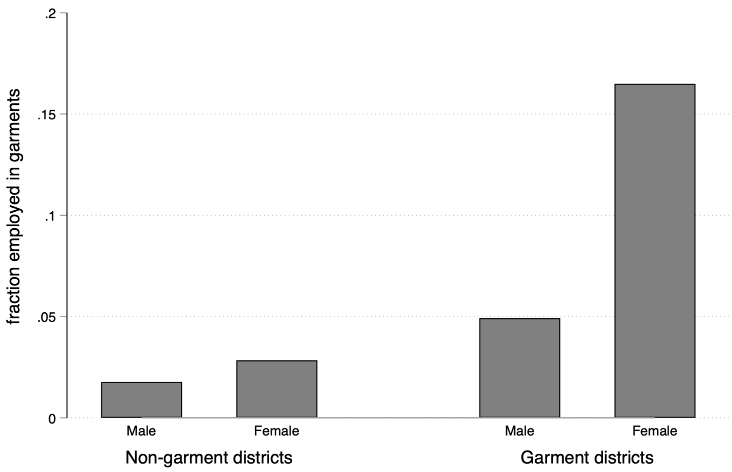
Note: Data from the 2000-2014 Cambodia DHS

Figure 1.7: Garment and non-garment wages



Note: Wage data from the pooled Cambodian Socio-Economic Survey 2010-2016.

Figure 1.8: Garment employment in treated and non-treated districts



Note: Data from the 2000-2014 Cambodia DHS

Table 1.1: Differences between garment and non-garment districts

	Garment (1)	N (2)	Non-Garment (3)	N (4)	(1)-(3) (5)	SE (1) -(3) (6)
A. Individual						
Age at first birth (women 30+)	21.96	3702	21.70	19382	0.26*	0.15
Age at first marriage (women 30+)	20.60	3806	20.49	19863	0.12	0.14
Height in cm (women 30+)	152.55	2006	152.46	11166	0.09	0.23
Years of schooling (women 30+)	3.58	4058	2.95	20944	0.63**	0.24
Years of schooling (men 30+)	4.71	3547	4.20	14635	0.52*	0.28
B. Village						
Distance to national borders (km)	66.27	14515	58.02	70146	7.75	10.65
Travel time (min) to city (100,000+)	112.49	14515	238.17	70146	-125.91***	18.85
Population density per km^2	741.05	14515	397.57	70146	341.10	301.28

Notes: Village here stands for primary sampling unit in the DHS survey. Reported values for garment districts are in waves prior to factory openings. Standard errors clustered on district level reported in column (6). *** p<0.01, ** p<0.05, * p<0.1

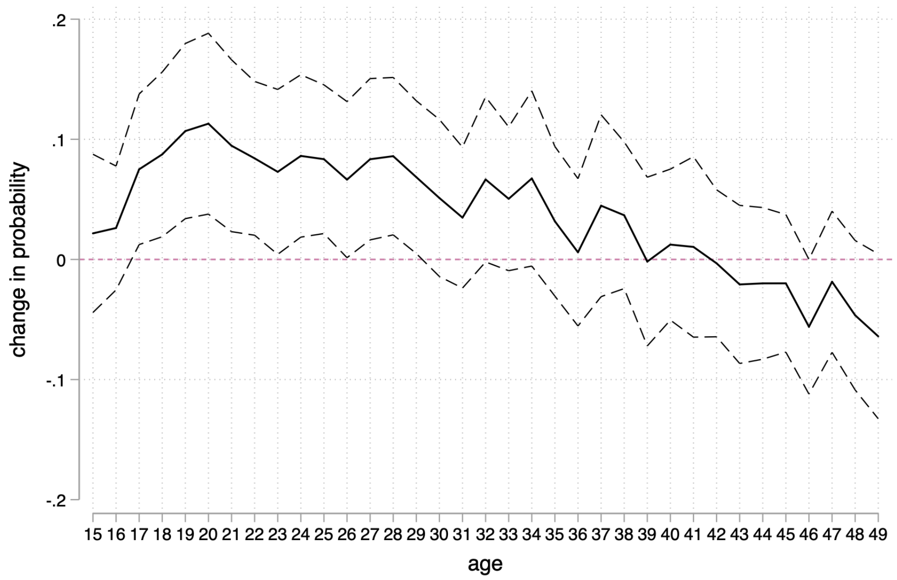
1.7.2. Results

Table 1.2: Garment exposure and labor supply outcomes

Dependent variable	Garment manufacturing	Garment manufacturing	Employed	Wage employment	Hours per week
Dataset	DHS 2000-2014	CSES 2010-2016	CSES 2010-2016	CSES 2010-2016	CSES 2010-2016
	(1)	(2)	(3)	(4)	(5)
Garment district	-0.0305	-0.0239	0.00575	-0.0822*	-2.687
	(0.0187)	(0.0252)	(0.0231)	(0.0434)	(2.155)
Garment district × Female	0.0628**	0.0560***	-0.00158	0.0501**	1.881**
	(0.0279)	(0.0168)	(0.0145)	(0.0212)	(0.931)
Observations	87,342	82,220	82,220	82,220	82,220
R-squared	0.188	0.246	0.148	0.162	0.141
Mean of dep. var	0.0465	0.0968	0.831	0.391	36.98

Notes: Estimates from model 1.1. Data from the main DHS sample in column (1) and the Cambodia Socio-Economic Survey 2010-2016 in columns (2) to (5). Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1.9: Marginal effects of garment exposure on women's garment employment



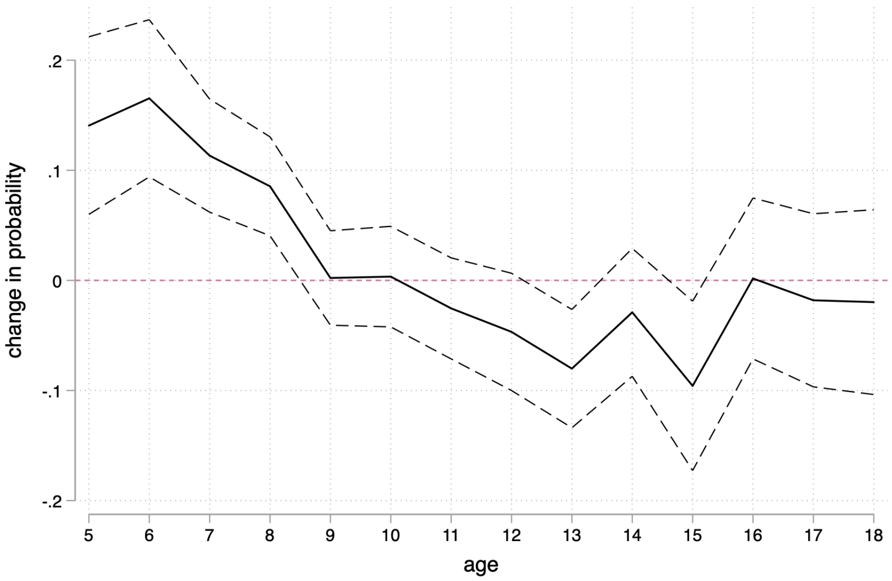
Notes: Estimates on absolute increases in female garment employment from model 1.2. Data from DHS 2000-2014. Dashed lines represent 95% confidence intervals.

Table 1.3: Effect of garment exposure on enrollment and educational attainment

Dependent variable	Enrolled	Education years	Education years
Age	Aged 5 - 18	Entire sample	Older than 18
	(1)	(2)	(3)
Garment district	0.0226		
	(0.0271)		
Garment district × Female	-0.00970		
	(0.0204)		
Years Exposure		0.435***	-0.0909***
		(0.0159)	(0.0239)
Years Exposure × Female		0.0826***	0.186***
		(0.0136)	(0.0268)
Observations	96,638	272,839	145,210
R-squared	0.121	0.247	0.373
Mean of dep. var	0.672	3.601	5.118

Notes: Estimates from 1.3. Panel 3 includes individuals that are currently older than 18 and are more likely to have completed their education. Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1.10: Marginal effects of garment exposure on girls' school attendance



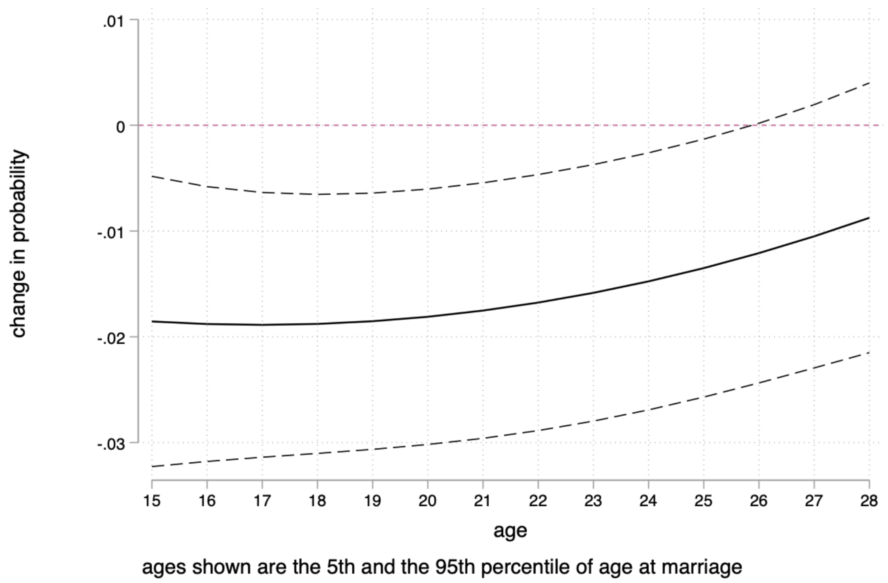
Note: Absolute changes in female enrolments estimated in model 1.4. Dashed lines represent 95% confidence intervals.

Table 1.4: Garment exposure and the timing of marriage

Dependent variable	Ever married	Ever married	Ever married	Ever married
Gender	Female	Female	Male	Male
	(1)	(2)	(3)	(4)
Years exposure	-0.000631***	4.88e-06	-0.000336	-2.29e-05
	(0.000230)	(0.000101)	(0.000211)	(3.69e-05)
Years exposure × Age		-7.86e-06		2.41e-07
		(8.80e-06)		(2.78e-06)
Years exposure × Age^2		2.44e-07		8.80e-09
		(1.93e-07)		(5.38e-08)
Observations	337,272	337,272	101,659	101,659
R-squared	0.283	0.271	0.286	0.323
Mean of dep. var	0.035	0.035	0.029	0.029

Notes: Estimates from model 1.5 in columns (1) and (3) and model 1.6 in columns (2) and (4). Retrospective sample based on individuals' reported age at first marriage and year of factory opening. The dependent variable equals 1 in the year in which an individual was married. Individuals are in the sample from birth until either the time of marriage, or the time of the survey (if unmarried). Years of exposure represent the years up to marriage that an individual in the regression was exposed to the garment industry (i.e. number of years in which there was a garment factory in their district). It equals zero for all individuals not in garment districts. Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1.11: Garment exposure and the timing of marriage



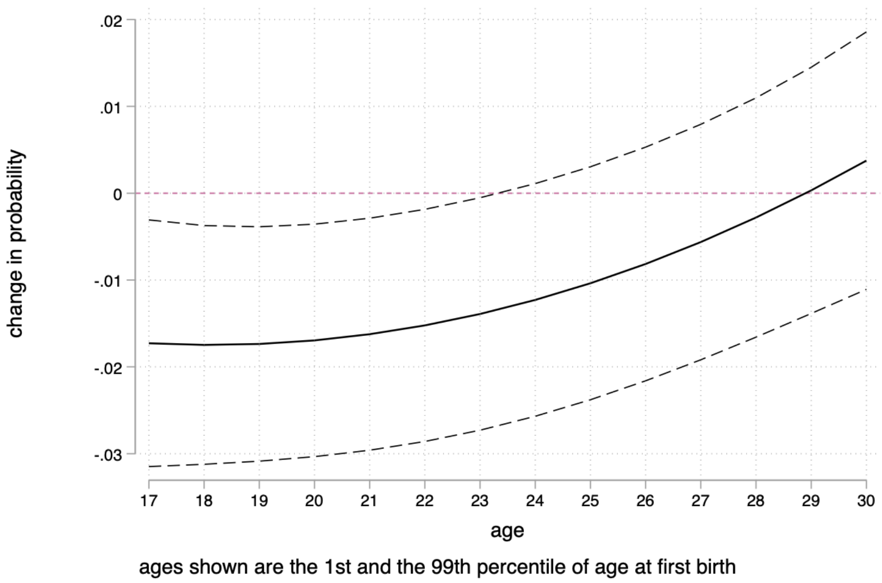
Notes: Estimates from model 1.6. Dashed lines represent 95% confidence intervals.

Table 1.5: Garment exposure and timing of first birth

Dependent variable	First birth	First birth	First birth	First birth
Gender	Female	Female	Male	Male
	(1)	(2)	(3)	(4)
Years exposure	-0.000502**	2.13e-05	-0.000155	-1.13e-07
	(0.000214)	(2.49e-05)	(0.000209)	(3.48e-06)
Years exposure × Age		-3.63e-06*		-1.02e-07
		(2.12e-06)		(2.50e-07)
Years exposure × Age^2		1.00e-07**		3.26e-09
		(4.53e-08)		(4.60e-09)
Observations	353,500	353,500	105,970	105,970
R-squared	0.181	0.283	0.274	0.333
Mean of dep.var	0.031	0.031	0.025	0.025

Notes: Estimates from model 1.5 in columns (1) and (3) and model 1.6 in columns (2) and (4). Retrospective sample based on individuals' reported age at first birth and year of factory opening. The dependent variable equals 1 in the year in which an individual had their first child. Individuals are in the sample from birth until either the time of first birth, or the time of the survey (if no children). Years of exposure represent the years up to first birth that an individual in the regression was exposed to the garment industry (i.e. number of years in which there was a garment factory in their district). It equals zero for all individuals not in garment districts. Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1.12: Marginal effects of garment exposure on the timing of first birth



Notes: Estimates from model 1.6. Dashed lines represent 95% confidence intervals.

Table 1.6: Effect of garment exposure on children's health (age 0-5)

Dependent variable	HAZ	HAZ	WAZ	WAZ	Anemic	Anemic
Sample	All	Born before	All	Born before	All	Born before
	(1)	(2)	(3)	(4)	(5)	(6)
Years exposure	-0.0432	-0.127	-0.0674	-0.0783	-0.0208	-0.0270
	(0.0699)	(0.0911)	(0.0457)	(0.0579)	(0.0200)	(0.0290)
Years exposure × Female	0.135*	0.220**	0.0525	0.0965	0.000768	-0.0144
	(0.0732)	(0.100)	(0.0554)	(0.0792)	(0.0238)	(0.0301)
Observations	11,894	10,510	11,894	10,510	11,894	10,510
R-squared	0.231	0.238	0.223	0.227	0.226	0.236
Mean of dep. var	-1.785	-1.843	-1.494	-1.533	0.584	0.588

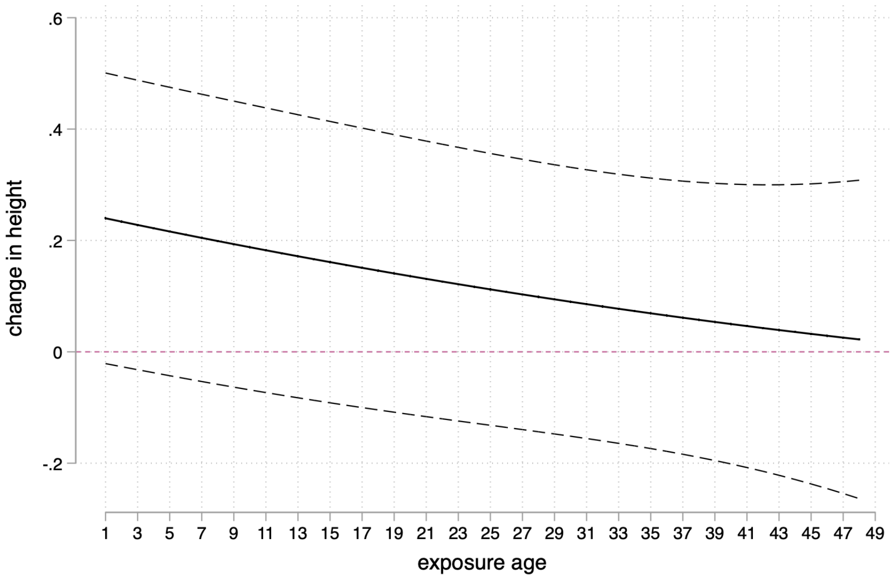
Notes: Estimates from model 1.7. Samples in columns (2), (4), and (6) are limited to children in treated districts born before the factory was open. This limits the probability that their mother migrated to the treated district, as most garment migrants move alone (Ministry of Planning, 2012) Standard errors clustered on the district level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.7: Garment exposure and women's health

Dependent variable	Height (cm)	Height (cm)
	(1)	(2)
A. Coefficients		
Years exposure	0.140	0.225*
	(0.111)	(0.120)
Years exposure× Age exposed		-0.00703***
		(0.00247)
Years exposure× Age exposed ^{2}		6.11e-05
		(7.90e-05)
B. Marginal effects		
Age Exposed = 1		0.2177*
		(0.1189836)
Age Exposed = 2		0.2109*
		(0.1182701)
Age Exposed = 3		0.2041*
		(0.1176072)
Age Exposed = 4		0.1975*
		(0.1169895)
Age Exposed = 5		0.19106
		(0.1164122)
Age Exposed = 10		0.1605
		(0.1139975)
Age Exposed = 15		0.1269
		(0.1110106)
Observations	35,900	35,900
R-squared	0.066	0.067
Mean of dep, var	152.6	152.6

Notes: Estimates from model 1.8. Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1.13: Marginal effects of garment exposure on women's height in centimeters



Notes: Estimates of the marginal effects of garment exposure by age of exposure from model 1.8. Dashed lines represent 95% confidence intervals.

Table 1.8: Garment exposure and women's earnings relative to their husbands

Dependent variable	Earns more than husband	Earns more than husband	Earns more than husband	Earns more than husband
Sample	All	Married before	Married after	All
	(1)	(2)	(3)	(4)
Garment district	-0.0718	-0.0677	0.0686	-0.0684
	(0.0517)	(0.0506)	(0.0919)	(0.0521)
Garment district × Exposed_{30}	0.0518***	0.0555***	-0.107	
	(0.0196)	(0.0206)	(0.0812)	
Garment district × Exposed_{28}				0.0586***
				(0.0193)
Observations	30,561	28,369	26,611	30,561
R-squared	0.102	0.106	0.109	0.102
Mean of dep.var	0.359	0.352	0.349	0.359

Notes: Estimates from model 1.9. Data from DHS 2005, 2010 and 2014 (dependent variable not in DHS 2000). Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.9: Garment exposure and women's decision over large household purchases

Dependent variable	Decides on purchase	Decides on purchase	Decides on purchase	Decides on purchase
Sample	All	Married before	Married after	All
	(1)	(2)	(3)	(4)
Garment district	-0.0661	-0.0785	0.00263	-0.0689
	(0.0604)	(0.0598)	(0.145)	(0.0599)
Garment district × Exposed $_{30}$	0.0543**	0.0761***	0.00210	
	(0.0228)	(0.0257)	(0.127)	
Garment district × Exposed $_{28}$				0.0895***
				(0.0214)
Observations	26,747	24,252	23,218	26,747
R-squared	0.134	0.135	0.152	0.134
Mean of dep.var	1.044	1.048	1.035	1.044

Notes: Estimates from model 1.9. 0= has no say, 1 = decides with someone else, 2 = decides alone. Data from DHS 2005, 2010 and 2014 (dependent variable not in DHS 2000). Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.10: Garment exposure and difference in partners' education

Dependent variable	E_m- E_f	E_m- E_f	E_m- E_f	E_m- E_f
Sample	All	Married before	Married after	All
	(1)	(2)	(3)	(4)
Garment district	0.0953	0.113	0.389	0.0830
	(0.166)	(0.162)	(0.501)	(0.169)
Garment district × Exposed $_{28}$	0.413***	0.0904	0.524	
	(0.156)	(0.156)	(0.375)	
Garment district × Exposed $_{30}$				0.342***
				(0.130)
Observations	45,853	43,007	40,370	45,853
R-squared	0.199	0.201	0.199	0.198
Mean of dep.var	1.565	1.539	1.509	1.565

Notes: Estimates from model 1.9. Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.11: Garment exposure and difference in partners' age

Dependent variable	A_m- A_f	A_m- A_f	A_m-A_f	A_m- A_f
Sample	All	Married before	Married after	All
	(1)	(2)	(3)	(4)
Garment district	0.142	0.0151	-0.931	0.122
	(0.276)	(0.288)	(0.798)	(0.277)
Garment district × Exposed $_{28}$	-0.151	0.161	0.191	
	(0.209)	(0.218)	(0.734)	
Garment district × Exposed $_{30}$				-0.0668
				(0.213)
Observations	41,578	38,931	36,789	41,578
R-squared	0.058	0.057	0.061	0.058
Mean of dep.var	3.212	3.147	3.192	3.212

Notes: Estimates from model 1.9. Sample includes individuals that live with at least one family member (other than a spouse) who was born before the first factory opening in the district. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.12: Pre-garment trends in educational attainment, enrollment, marriage, first birth, and height

Dependent variable	Years of schooling (1)	Enrolled ages 5-18 (2)	Ever married (3)	Any Children (4)	Height for age children (5)	Height in cm adult women (6)
Years to factory	-0.0227	0.000435	0.00236	0.00146	0.00742	-0.0534
	(0.0165)	(0.00224)	(0.00179)	(0.00155)	(0.0229)	(0.0413)
Years to factory × Female	-0.00625	0.000764			-0.00967	
	(0.00530)	(0.00123)			(0.0208)	
Observations	182,713	67,553	42,275	42,275	6,341	21,451
R-squared	0.167	0.104	0.378	0.406	0.222	0.053
Mean of dep. var	4.117	0.642	0.702	0.647	-1.705	152.2

Notes: Years to factory is calculated by subtracting the year of the observation from the year of arrival of garment manufacturing in a district. Regressions include individual's age and village characteristics, as well as district fixed effects, year fixed effects, and province-year or, where applicable, province-year-female dummies. Regressions in column (5) also include dummies for age in months. Standard errors clustered on the district level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Chapter 2

WHO NEEDS ELEMENTARY MATH? CURRICULUM REFORM AND THE GENDER GAP IN HIGH-SCHOOL MATH AND SCIENCE IN AUSTRALIA

2.1. Introduction

Subject choices in high school are an established predictor of gender gaps in career and wage outcomes (Goodman, 2019; Card and Payne, 2017; Joensen and Skyt Nielsen, 2009 and 2016; Altonji, Bloom and Meghir, 2012; Goldin, Katz and Kuziemko, 2006). Girls tend to take fewer or easier mathematics and science subjects than boys, which leads to fewer women in fields like science, technology, engineering and mathematics (STEM) (Delaney and Devereux, 2019; Card and Payne, 2017). These consequences have motivated a growing body of research analysing the role in subject choices of acquired and inherent gender-specific characteristics (Buser, Niederle and Oosterbeek, 2014; RodrÃguez-Planas and Nollenberger, 2017; Delaney and Devereux, 2019). However, there is little analysis of how making high-school STEM subjects compulsory affects these gender differences. If a requirement that students take mathematics or science is more binding for girls than boys, due to girls' lower propensity to take these subjects in the first place, then it is likely to narrow the gender gap in STEM subject choices and the gaps in educational and career outcomes that result from these choices. In this paper I show that compulsory math does indeed reduce the STEM subject gap. Moreover, I demonstrate a positive externality from compulsory math, driven by the intuitive notion that students face complementarities between subject choices: it is difficult to study science without any mathematics. Using a natural experiment in which high-school math became non-compulsory, I show that the reduction in math uptake among girls also leads to a reduction in uptake of the STEM subjects to which girls are typically more inclined, such as biology.

By demonstrating complementarities in subject choices and their effect on gender gaps, this paper contributes to the areas of literature examining the presence and causes of gendered patterns in subject choices, the consequences of STEM subject choices, and the role of curriculum policies in affecting gender gaps. Joensen and Skyt Nielsen (2009, 2016) show that girls are less likely to take mathematics when it is made more costly by requiring that it be taken with physics. My results indicate that the marginal benefit of pairing a subject with math is not only determined by the preferences towards the second subject, but also by any complementarities it has with math. I also build on works on the behavioral origins of gender gaps in STEM subjects, (Buser et al., 2014; Niederle and Vesterlund, 2010; Morgan et al., 2013; Kamas and Preston, 2014), by showing that significant gender gaps in mathematics and science uptake persist after controlling for ability (including comparative advantage), confidence, and early STEM career plans. This evidence is consistent with present bias in subject choices, with women foregoing the long-term benefits of choosing math and science for the short-term utility gains from opting out (Lavecchia, Liu and Oreopoulos, 2016). Several papers show that taking advanced mathematics subjects lead to higher future earnings (Altonji et al., 2012; Joensen and Skyt Nielsen, 2009), while Goodman (2019) shows a positive earnings effect of elementary mathematics. My results indicate that elementary math subject choices could have different longer-term benefits, by widening the range of complimentary subjects/careers available to the student. Finally, my study also demonstrates that understanding the spillover effects of the uptake of mathematics on other subjects is crucial for policy, as it could imply that providing flexible options and making at least some mathematics compulsory in high school can encourage more females to study science as well.

The identification strategy exploits a 2001 curriculum reform in the Australian state of New South Wales (NSW). The reform removed the requirement that students take either mathematics or science, and increased the difficulty of the elementary mathematics class. No similar curriculum reforms took place in other states, with math remaining either compulsory or noncompulsory depending on the state. Data on the effects of the reform come from a nationally representative survey that contains school identifiers, demographic characteristics, test results and students' perceived ability. This dataset permits analysing pre-reform subject choices and estimating differences in outcomes for cohorts of male and female students in NSW and control states just before and after the reform. While the trends in subject uptake differed across states, they were not significantly different between boys and girls within states.

I begin the analysis by showing that, for students who would otherwise not study any mathematics or science, the pre-reform requirement was fulfilled by taking elementary mathematics. Similar to students in the United States (U.S. Department of Education, 2018), very few students took any science class without studying at least some mathematics, with more quantitative science classes being increasingly paired with more difficult mathematics classes. This is evidence that taking mathematics is valuable for students who want to study science, further supported by the fact that high-school science subjects and university science majors have math requirements. Among the complimentary sciences, I find the strongest effects for biology, which has been shown to be a science that is particularly favoured by girls (Card and Payne, 2019; Cheryan et al., 2017; Delaney and Devereux, 2019).

I then present a simple framework in which there are costs to taking science without mathematics, and show that making mathematics non-compulsory can lead to dropouts not only by students who would otherwise take a single class (elementary mathematics) to meet the requirement, but also by those studying mathematics and biology. The framework demonstrates that this spillover effect is exacerbated by the curriculum reform component that increased the difficulty of the elementary mathematics class, which further discouraged uptake. In this setup, more female students would drop out of the science than male students if females' preference for the science relative to mathematics is higher than males'. This is consistent with findings that, among STEM subject choices, females overwhelmingly prefer biology (Delaney and Devereux, 2019). This simple setup emphasizes the role that the cost of studying science without mathematics plays in students' decision making. While intuitive, this mechanism has been overlooked in previous research on students' subject choices.

Next, I show that the estimated impacts of the 2001 reform in NSW are consistent with the proposed mechanism. I report a large decline in mathematics and science uptake for females relative to males, and a corresponding increase in the share of females taking neither subject, following the curriculum reform. This was driven by 8 percentage points more females dropping out of biology along with elementary mathematics than males, while there was no significant decrease in females relative to males studying only elementary mathematics. I also demonstrate that this was associated with nearly a 3 percentage point decrease in females enrolling in biology-related sciences majors (life sciences) at university. My findings indicate that less than 10 percent of these results were explained by mathematics ability (including relative to English ability), confidence, and plans to work in STEM as stated at the beginning of high school. I do not find changes in probability of university graduation or earnings within five years of finishing high school, although the sample suffers non-random attrition, so better longer-term data is needed to establish effects on earnings. These findings verify the prediction that a disincentive for the study of any mathematics is also a disincentive for the study of science, both at high-school and university, and especially for females.

Finally, to better understand the welfare implications of these findings, I show that the negative effects that the reform had on science and math uptake, and on life science university enrollments, were concentrated among students with ex-ante mathematics performance close to the median. These findings are consistent with the presented framework because, regardless of gender, if the optimal level of mathematics is positively correlated with ability: (i) students at the bottom of the skill distribution are less likely to combine mathematics and biology with or without a compulsory mathematics requirement, and (ii) those at the top of the skill distribution would be less likely to drop out (and also less likely to study elementary mathematics), regardless of gender. The result that the gender gap opened around the median of the skill distribution is important, as it means that the females who left biology along with basic mathematics had the potential to perform well and benefit from studying these subjects. This also suggests that, beyond educational and career considerations, the increased gender gap in maths and science uptake may represent a reduction in the pool of scientific talent.

In summary, these findings present evidence that curriculum reforms that affect the study of mathematics and science are not gender-neutral. They reveal that the role of mathematics in science is an important consideration in students' subject choice, and incentives for the study of math not only reduce gender gaps in math uptake, but also have spillover effects on females' uptake of science. This is a previously undocumented mechanism, which also adds relevant insights to the debate whether at least some mathematics should be compulsory.[1] The rest of the paper is organized as follows. Section 2.2 provides background on the 2001 reform. Section 2.3 describes pre-reform subject uptake patterns and develops a simple framework to demonstrate the expected effects of the reform. Sections 2.4 and 2.5 describe the data and empirical strategy. Section 2.6.4 presents the results and section 2.7 concludes.

[1]For example, this debate has been ongoing in the state of NSW ever since the 2001 reform and the reinstatement of a mathematics requirement is currently being planned (Sydney Morning Herald, 2019)

2.2. The 2001 HSC reform

The credential awarded for completing the final two years (11 and 12) of schooling in New South Wales (NSW) is known as the Higher School Certificate (HSC). Students choose their HSC subjects in year 11, when they take preparatory level classes. In year 12, they take the HSC-level classes of those subjects, and at the end of the year sit external examinations for them. Following a comprehensive review of the HSC (McGaw, 1997), a new curriculum was introduced to the cohort that started year 11 in February 2000. This cohort took the HSC-level classes in 2001, when they were in Year 12, and were the first to complete the HSC under the new curriculum.

The 2001 HSC reform eliminated a number of overlapping subjects, made changes to the way examinations marks were reported, and left English as the only mandatory subject in the HSC curriculum. Previously, the Breadth of Study requirement had mandated one unit from the Math, Science or Technology areas, two units of English, and one unit from the Humanities, Languages or Creative Arts areas.

The reform's most noteworthy changes were to elementary mathematics. Prior to the reform, while there were various low-level subjects that students took to fulfil the mandatory arts and humanities requirement, in the mathematics and science subject area, the math or science requirement was most commonly fulfilled by taking one of the two elementary mathematics subjects, Mathematics in Society and Mathematics in Practice (see next section). After the reform, these subjects were replaced by a single new course General Mathematics, which was more difficult and comprehensive than its predecessors (Coupland, 2006). In addition to basic algebra, data analysis and geometry, the new elementary mathematics class now included additional topics in financial mathematics, statistics and modelling (Ayres and McCormick, 2006).

Aside from increasing the difficulty of the elementary mathematics class, and removing some overlapping subjects, the 2001 HSC reform did not limit the subject areas that students could choose from or the content of other key courses in mathematics, science and technology. It also did not affect the subjects that students took before entering years 11 and 12, at which stage students couldn't choose their mathematics and science classes and at least one subject in science, math, humanities and English was taken by everyone. Therefore, without affecting previous studies, other subject areas or university admission procedures in a major way, the 2001 HSC reform disincentivized the study of elementary mathematics by making it non-compulsory and increasing the difficulty of the elementary mathematics course.

2.3. Pre-reform subject uptake and expected effects

2.3.1. Pre-reform subject choices

Figure 2.1 confirms that prior to the reform, nearly all male and female students in NSW took some mathematics, while in the seven other (control) states, some of which did not have compulsory mathematics, 27 percent of females and about 20 percent of males did not take any. This indicates that the vast majority of students in NSW who would prefer to not study any subject in the math and science area would meet the requirement by taking mathematics. Figure 2.1 also shows that nearly all students who were taking science subjects (biology, chemistry or physics), were studying some mathematics as well. There was a small number of students, mostly females, who took biology without any mathematics, in NSW and in control states. This is consistent with the framework I present in the next subsection.

Figure 2.2 pools subject choices prior to the 2001 reform across all states, and illustrates how students tended to pair science with mathematics. The pattern that emerges is that biology was most commonly paired with elementary mathematics, chemistry was most commonly paired with intermediate or advanced mathematics, and physics was most commonly paired with advanced mathematics. This type of pattern is not a uniquely Australian phenomenon (see, for instance, U.S. Department of Education, 2018). The fact that more female students choose biology over chemistry (Figure 2.1), and more so over physics, is also a well documented phenomenon (Card and Payne, 2019; Cheryan et al., 2017).

Mathematics was not a formal prerequisite for the uptake of any science subject, but the expected knowledge outlined in syllabuses, and the contents of HSC exams in biology, chemistry and physics before and after the reform, show that knowledge and application of mathematical concepts is expected in all three subjects, with increasing complexity in that order (Table 2.11). This indicates that taking these subjects without an appropriate level of mathematics would reduce students' chances of performing well, and would increase the effort associated with passing. In the long run, not studying any mathematics would also reduce students chances to do well in science-related university majors. Advice provided to prospective HSC students by the Australian Universities Admissions Centre confirms that, for university majors related to biology (such as biological science, general science, biotechnology, marine science, environmental science etc.), prior uptake of at least elementary mathematics is assumed (Universities Admissions Centre, 2017). The report also outlines that majors related to biology are less likely to specifically assume or require prior uptake of an advanced mathematics subject, compared to those where chemistry and physics are recommended.

2.3.2. Who needs elementary math? A simple framework

To understand the effects of the reform on the study of elementary mathematics and of science, I present a framework in which there are costs to taking science without mathematics. The costs could represent the difficulty of performing well in the science, and, over the longer run, of potentially failing university majors that assume knowledge of at least elementary mathematics. The evidence for these costs is discussed in section 2.3.1. The framework demonstrates how the 2001 reform, which led fewer female students to take mathematics, also reduced the share of females relative to males studying biology.

Let's assume no subject requirements and a discrete set of science options $s_n \in \{0, s_1, s_2\}$ and mathematics options $m_n \in \{0, m_1, m_2\}$, ordered by their quantitative difficulty. Since the reform affected elementary math, for simplicity I disregard more difficult mathematics and science classes. Throughout, "more/less difficult" will mean "requiring more/less advanced mathematics concepts". Each student has an optimal allocation m_i^* and s_i^* of science and mathematics, which can be any nonnegative number, not limited to integers. Given the discrete options available, they choose their preferred subject combination (m_n, s_n) by maximizing the utility function:

$$U_i = f(|s_n \hat{a} s_i^*| + |m_n \hat{a} m_i^*| + a \times 1_{(s_n > m_n)}(s_n - m_n)), \tag{2.1}$$

where $f(\cdot)$ is a monotonically decreasing function. Similar to a standard multinomial discrete choice model, students choose the subject bundle which minimizes the distance to their optimal one. What is new in this setup is the incorporation of the cost $a \times 1_{(j>k)}(s_j - m_k))$, which they incur if they choose a science level that is more mathematically demanding than their chosen mathematics level. In this case, the cost is equal to the difference between the chosen science and mathematics class. The constant a represents the severity of the cost, which is assumed to be the same across all students.

Within this framework, students would not incur an additional cost if they took the easiest mathematics with the easiest science, but they would if they took the easiest science without any mathematics, or the easiest mathematics with a harder science. Given that biology is the least mathematically demanding of the science subjects, this is why the choice whether to take elementary mathematics is expected to be linked with the choice to study biology, but not more difficult science subjects.

In order to understand the implications of the reform by gender, it is important to see how its effects would vary by ability. In this setting, students' optimal choices of science s_i^* and mathematics m_i^*, could be represented as a monotonically increasing function $g(\cdot)$ of randomly assigned and continuously distributed ability and specific subject preference:

$$\begin{aligned} s_i^* &= g(\sigma_i, \pi_{si}) \\ m_i^* &= g(\sigma_i, \pi_{mi}), \end{aligned} \tag{2.2}$$

where σ_i stands for student i's quantitative skill, π_{si} is their assigned preference for science and π_{mi} is their assigned preference for mathematics. Given that in this setup subjects are ranked based on their quantitative difficulty, it is reasonable to assume that the probability of preferring a more quantitatively demanding science class would depend on mathematics ability. At the same time, extrinsic (such as stereotypes, for instance, Kessels (2005)) and intrinsic (such as competitiveness, see Buser, Niederle, Osterbeek (2014)) factors would determine individual students' preferences for mathematics and science separately. This is particularly pertinent for analyses of subject choices from a gender perspective, as women have been documented to prefer biology and chemistry to physics, even when they choose advanced mathematics classes (U.S. Department of Education, 2018; Joensen and Skyt Nielsen, 2009; Card and Payne, 2017).

Prior to the reform, the option to take no science and no mathematics was not available: $(m_k, s_j) = (0, 0)$ was not a feasible bundle. Figure 2.3 maps students' continuous optimal bundles (m_i*, s_i*) into corresponding subject choices for three scenarios: prior to the reform (panel A); after making mathematics and science non-compulsory (that is, (0,0) becomes an option) (panel B); and combining the non-compulsory math and science with increased difficulty of elementary math ($m_{1new} > m_{1old}$) (panel C). I do not look at students with $s^* > s_1$ and/or $m^* => m_1$, because these students' chosen bundle would never be (0,0).

Panels A, B and C in Figure 2.3 illustrate the spillover effect that a reform making mathematics and science non-compulsory has on students who would fulfil the prior requirement by taking either mathematics or mathematics and biology. Panel A shows that, in the presence of the described complementarity between elementary mathematics and biology, the pre-reform constraint would make it optimal for some students that like biology but don't like mathematics to take both subjects. These are the students shaded green that have $m^* < 0.5$. Panel B demonstrates that when (0,0) became an available option, some of these students' optimal bundle would shift to (0,0). This is crucial, as it shows that students with a relatively strong interest in biology drop out of it along with elementary mathematics. Note that, because of the presented complementarity, these effects would be present in a setting where only math was compulsory in Panel A. Unsurprisingly, the students whose optimal preference is for very little math and even less or no science, would be taking only elementary mathematics in Panel A, but would choose (0,0) in Panel B. Due to the cost of studying biology without any mathematics, only a small fraction of students take only biology, despite the fact that biology can satisfy the compulsory requirement. Once the requirement is removed (Panel B), this share decreases further. Panel C shows that the increase in difficulty of the elementary class reinforced the effect of removing the compulsory subject requirement, with even more students dropping out of both biology and elementary mathematics. These are the students that are shaded green in Panel B but grey in panel C.

Holding ability constant, females would be more likely to drop out of biology together with elementary mathematics in both panels B and C if the density of their science preferences is concentrated closer to 1, while their math preferences are clustered closer to 0, relative to boys. That is, if females' optimal bundle is more likely, on average, to be above the dashed diagonal line (area (a)) in Figure 2.3, disproportionately more females would end up with no mathematics or science after the reform. As previously discussed, females' sorting into biology and preference for no or less-intensive mathematics classes is well documented, and is also observed in the Australian setting.

On the other hand, regardless of potential disparities in preferences, gender differences in the response to the reform would be less likely to emerge for students at the top and at the bottom of the ability distribution. This is easy to see, as students at the top of the distribution would be less likely to be taking elementary mathematics, regardless of their science or math subject preferences, and those at the bottom would be less likely to study a math and science combination prior to the reform, and would all drop out of mathematics, regardless of gender.

Finally, Figure 2.4 confirms that calibrating this simulation with various severities of the cost (a=0.5, a=0.9) and various changes in difficulty ($\frac{m_{new}}{m_{old}} = 1.1$, $\frac{m_{new}}{m_{old}} = 1.3$), yields results that are qualitatively the same. It also presents potential explanations for why more students in control states took only biology (see the previous subsection). This could be explained by, for instance, the cost of taking biology without math being lower (due to the math-intensity of the class), or the elementary mathematics class being harder. Increasing a leads to a larger share of students, who have a strong preference for biology but do not like mathematics (Figure 2.4.1 and 2.4.2), taking biology only. Increasing the difficulty of the new elementary mathematics class also induces more students to drop out of biology and mathematics, as well as mathematics only (Figure (2.4.3 and 2.4.4)). Figure 2.4 also shows that it is possible to observe an increase in the share of female students (with a strong relative preference for biology) who took only biology after the reform. However, because of the cost of not studying mathematics in this setup, the net effect on the uptake of biology, would be negative.

To summarize, the presented setting yields the following testable predictions on the main gendered effects of the reform on math and science:

- If women's preference for math is lower, as math uptake becomes less attractive, more women, relative to men, would drop out of elementary mathematics and mathematics altogether (controlling for ability);
- If women have a higher preference for biology relative to math compared to men, the presented complementarity between math and science would lead more women to drop out of biology along with elementary mathematics (controlling for ability);
- These gendered effects will be weaker at the top and bottom tails of the math ability distribution.

2.4. Data

For the main part of the analysis I use the last pre-reform and first post-reform waves (1995 and 1998) of the nationally representative Longitudinal Surveys of Australian Youth (LSAY). The LSAY follows cohorts of Year 9 students (15 years old) for a period of ten years after the initial interview. The 1995 and 1998 cohorts were in year 12 (the last year of high school) in 1998 and 2001 respectively. The 1998 cohort therefore captured the first students to be exposed to the new curriculum requirements. The LSAY includes information on high-school subject choices, field of tertiary education, university completion, and earnings, until students reach 25 years of age. It also includes results from standardized mathematics and English tests taken in the first survey year (i.e. year 9). I use these results as a proxy for mathematics and English ability. In addition, the LSAY contains baseline information on self-concept of ability in mathematics, English and overall performance, which I use to proxy for student's confidence. Students in both cohorts were also asked what their preferred future occupation was, with approximately 87 percent of students providing an answer. The dataset also contains information on socio-economic status and parents' education and background.

Table 2.1 reports summary statistics of the above described characteristics by sex and state in the pre-reform cohort. While there are some differences between boys and girls in both New South Wales and control states, the only significant difference between boys and girls across treatment status is in their plan to work in STEM: at age 15, girls in NSW were more likely to plan a career in STEM than girls in control states, relative to boys.

The main sample used for this analysis consists of 6,613 students before and 5,886 students after the reform, interviewed from age 15 to age 20 (two years after high-school graduation). The LSAY data shows that more than 92% of students who went to university had enrolled within two years from graduating high school. To examine potential changes in earnings and university completion, I also use the last available survey wave, 10 years after students were first interviewed. As survey attrition was particularly high after students left high school, the sample size for the earnings and university completion regressions is significantly smaller (3,160 and 2,868 pre- and post-reform participants respectively). Previous analysis of the survey (Rothman, 2009) has presented evidence that attrition was non-random, which combined with the fact that survey participants were just aged 25 in the final wave, makes this reduced sample less reliable for studying longer-term impacts of the reform.

Finally, to test the parallel-trends assumption, I supplement the pre-reform LSAY cohort with six waves (1990-1995) of the Australian Youth Survey (AYS), which is less rich than LSAY, but is the only nationally representative survey to contain individual-level data on students' Year 12 subject uptake prior to 1998.

2.5. Empirical strategy

The empirical analysis uses a difference-in-difference strategy and exploits the variation between male and female students within states. The regressions compare NSW students that completed year 12 in 1998 (pre reform) to those that completed year 12 in 2001 (post reform). Students from other states in these two cohorts serve as a control group. The regressions focus on how differences between boys and girls vary between the pre and post reform groups. Specifically, I estimate the following equation:

$$\begin{aligned} Y_{ishc} = \beta_0 + \beta_1 NSW_s \times Post_c + \boldsymbol{\beta_2} Female_i \times NSW_s \times Post_c \\ + \delta_{sh} + X'_{ishc}\gamma + \epsilon_{ishc} \end{aligned} \tag{2.3}$$

where: i represents individual, s represents state, h represents high school and c

represents cohort; δ_h is a high-school fixed effect and X_{ishc} is a vector of controls that includes $Post_c$, $Female_i$, and $Post_c \times Female_i$, mathematics and reading scores, self-concept in mathematics and English, future plans to work in STEM, and socio-economic status and parental characteristics; and the dependent variable is either a binary variable representing subject choice, university major, or another post-secondary outcome. The main coefficient of interest is β_2 in (2.3), which represents the change in subject uptake, university major or other outcome of interest for girls, relative to boys, associated with the 2001 reform.

The identification strategy relies on the assumption that while trends in subject uptake may have been different across states, within states these trends were the same across genders. Tables 2.2, 2.3 and 2.4 support this assumption. In annual data between 1990 and 1998, trends in subject uptake did differ between NSW and control states, but there was no significant difference once boys were used as an additional control group. A potential threat to this identification strategy is the possibility that, in anticipation of the reform, males and females in the later cohort made different schooling choices before years 11 and 12. However, the post-reform cohort began high-school (in year 7) in January 1996, well before the first report announcing the government's plan for curriculum changes was published in late 1997 (Aquilina, 1997). This makes it highly unlikely that the reform affected the choice of school for students in my sample. Moreover, prior to the last two years of high school, all students in NSW study mathematics, science, English, humanities and social sciences, which makes it very unlikely that the reform affected subject choices before the senior secondary level.

To address self-selection into schools and variation in subjects offered across schools, in my main specification I include school fixed effects. For robustness and to better understand the mechanisms behind my findings, I also report the results with state fixed effects only, as well as including covariates for math and English ability, confidence and future career plans. To address potential correlation in the standard errors, they are clustered at the 72 levels of the interaction between state, region (metropolitan, regional or remote) and school sector (government, private, independent). As a robustness check I also cluster standard errors on the school level, and the significance of coefficients remains consistent throughout.

2.6. Results

2.6.1. The 2001 reform and the uptake of mathematics and science

This section begins by exploring the impact that the 2001 reform had on girls' mathematics and science subject uptake relative to boys. I start by showing that there were significant gender differences in the reform's impact within subject areas. Table 2.5 demonstrates that it led to more girls dropping mathematics and science than boys. The results increase in size and significance as more variance is captured by fixed effects and controls for ability, confidence and career plans. Table 2.5 also demonstrates that, in contrast to mathematics and science, there is no gendered effect of the reform on English or technology. While it is outside of the scope of this paper, the removal of the humanities requirement (see section 2) also led to more boys dropping humanities than girls, which is consistent with previous findings that boys have lower preferences than girls for humanities subjects (Card and Payne, 2017).

The next results separate the mathematics and science choices by the particular type of mathematics or science subject taken (Table 2.6). As expected, the relative decline in girls taking mathematics and science is driven by a decline in elementary mathematics and in biology, respectively. When controls are included (columns 2, 3, 5 and 6), there is no significant effect in intermediate or advanced mathematics, or in chemistry or physics. The confinement of the mathematics effect to elementary mathematics is unsurprising, given that a requirement to take at least minimal mathematics or science is unlikely to be binding for any students, male or female, that have chosen to take more than the minimum mathematics. The confinement of the science effect to biology is consistent with the fact that more girls took biology before the reform (see section 3.1) and with the framework in section 3.2.

Table 2.7 presents this paper's key result – that the decline in girls taking biology is driven by girls dropping elementary mathematics and biology simultaneously. The decline in girls taking biology, discussed above, cannot be explained by girls dropping only biology, because the reform led to an increase in the number of girls taking only biology (in line with the predictions in section 3.2). Nor can it be explained by girls that were taking math and biology dropping only biology, because there was no significant increase in girls taking only mathematics. There is, however, a significant increase in girls that take neither mathematics or biology. This confirms the importance of studying at least some mathematics for students', and especially females', choice to study biology. In line with predictions, there is also a small increase in the number of students choosing to study only biology.

The magnitudes of these effects are non-negligible and have persisted over time. The decrease in the share of girls taking math and biology relative to boys is equivalent to more than a third of the pre-reform difference (see figure 2.1). Analysis of subject choices over time has shown that the reform's sizeable effects on the gender gap in mathematics and science uptake identified in this paper were also long-lasting and could have potentially intensified over time (Wilson and Mack, 2016). For instance, in 2014, female year-twelve students in New South Wales were 10 percentage points more likely to study no mathematics or science, compared to male students. For a comparison, in the pre-reform cohort, the gap was approximately 1.5 percentage points[2](figure 2.1), and in the first pre-reform cohort it grew by approximately six percentage points (table 2.7.

2.6.2. High school graduation and university majors

I next explore whether these effects on subject uptake translated into longer-term differences in outcomes by gender. Table 2.8 shows that the reform was not associated with changes in girls' relative probability of high school graduation (columns (1) to (3)), their probability of entering university (columns (4) to (6)), or their university admission scores[3] (columns (7) to (9)). The fact that girls' relative scores did not change is of particular interest. As admission scores are based on students' scores in their chosen subjects, this result indicates that girls did not necessarily switch to subjects in which they performed better. Moreover, as the scores were also representative of their ranked overall performance, this also indicates that the effect of the increased difficulty of the new elementary mathematics class, which would potentially affect girls' score relative to boys in a positive direction, was likely offset by the fact that girls switched to classes in which their performance was worse.

Moving to university majors, table 2.9 demonstrates that there was a significant decrease of around 2.8 percentage points in the share of females entering into biology-related life science university majors, including biology, microbiology, and marine science among others (see table 2.12). Given that it was girls' relative uptake of elementary mathematics and biology which was affected by the reform, these results are very intuitive and complement previous findings that gender gaps in advanced mathematics and science combinations in high-school are linked with

[2]This was likely driven by students who finished high school without obtaining the Higher School Certificate.

[3]The university admission score represents students' rank based on HSC exam marks.

gender gaps in enrollments in university STEM majors (Delaney and Devereux, 2019; Card and Payne, 2017).

Of particular relevance is that for many degrees, with the exception of engineering, while there were recommended subjects, no formal subject requirements were in place and admission decisions were made solely based on students' admission rank (University Admissions Centre, 2017). As discussed above, the results also show that the observed decrease in life science enrollments was not due to changes in students' admission scores or probability of high school graduation (table 2.8). This suggests that the observed link between the decrease in mathematics and biology uptake and the lower enrollments in biology STEM majors is likely a reflection of students' lower preparedness for the field. Considering the seven percentage point decrease in girls' elementary mathematics and biology enrolments, this could indicate that a percentage point decrease in elementary mathematics and biology uptake translated into 0.4 percentage points decrease of enrolments in life-science university majors.

2.6.3. Results by mathematics performance

To better understand the mechanisms in play and the welfare implications of the above results, I now turn to analyzing the effects of the 2001 reform on subject uptake and university majors by mathematics performance. Figure 2.5 displays the β_2 coefficients (in equation 2.2 using the specification in Table 2.7 columns (3) and (6), estimated for various subsamples by levels of mathematics ability. Mathematics ability is proxied by students' standardized test scores in Year 9.

2.5 shows that the females who opted out of life science majors and those who dropped out of the elementary mathematics and biology subject combination were concentrated around the median of the math score distribution (scores around 13). These findings are also consistent with the predictions from the framework in section 2.3.2 because, if the optimal level of mathematics is positively correlated with mathematics ability, then: (i) both males and females at the bottom of the skill distribution would drop out of mathematics (with or without biology), and (ii) those at the top of the skill distribution would be less likely to drop out (and also less likely to study elementary mathematics), regardless of gender. The result that the gap opened around the median of the skill distribution is also policy relevant, as it means that the females who left basic mathematics and biology altogether would have likely been able to perform well in these subjects.

2.6.4. Short-term effect on earnings and university completion

This last section analyses the short-term impacts that the HSC reform may have had on the gender gap in earnings. The LSAY follows students and provides earnings data for 8 years after high-school graduation. This means that the last available salary information is from ages 25-26. Therefore, the results on earnings would only provide a short-term effect on earnings and would not be indicative of effects that would materialize in the long-run.

With the above in mind, other studies (Rose and Betts, 2010; Goodman, 2019) have found mixed long-term effects of elementary mathematics on earnings. My results so far show that the reform induced more females to shift away from life science, which does not produce any straightforward prediction on the effect on earnings among university graduates. Among those who did not go to university, an elementary mathematics class (or the lack thereof) could still have an impact on earnings through signalling ability or enhanced productivity (Rose and Betts, 2010). I find that the reform had no significant impact on earnings or university completion status at the time students reached their mid-twenties (Table 2.10). However, better and longer-term data would be necessary to evaluate the potential career and earnings implications of the reform over students' lifetime.

2.7. Conclusion

This paper presents evidence that the study of mathematics in high school has important implications for the uptake of science, especially for females. For female students, reducing the uptake of elementary mathematics was associated with a significant decrease in those studying biology. This further translated into fewer females studying biology-related science majors in university.

I present an explanation for this unintended effect, which, while very intuitive, has not received much attention previously: students view some mathematics as an informal prerequisite for performing well in science. Elementary mathematics provided sufficient preparation for biology but not physics or chemistry, as biology is the least quantitatively demanding of the three science subjects. Given females' stronger sorting into biology over chemistry and physics, and their preference for less-demanding mathematics classes, by disincentivizing the uptake of elementary mathematics, the reform disproportionately affected female students' uptake of biology.

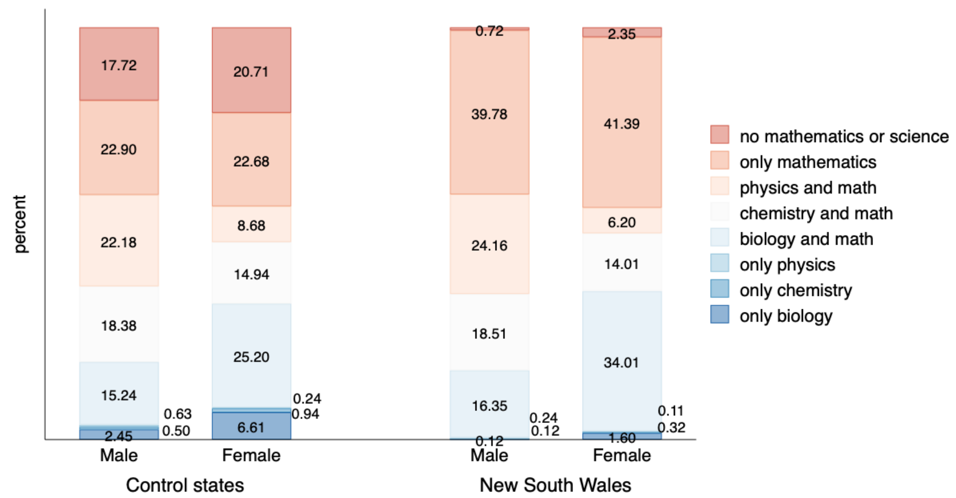
The framework presented shows that both removing the non-compulsory mathematics and science requirement and the increasing of the difficulty of the mathematics class could lead to more females leaving biology, along with mathematics. While it is not possible to disentangle the impacts of the two components, the key message is clear: encouraging more females to take mathematics can have positive spillover effects on the uptake of science. Another important message that emerges from these findings is that lumping students' preference for mathematics and science together, especially from a gender perspective, may not be accurate. Acknowledging that female students' optimal choice of science may exceed that of mathematics and considering how the study of mathematics complements that of science (and not vice versa), could be an improvement to modelling subject choices.

From a policy perspective, the fact that the impact of the reform was strongest among female students around the median of the math performance distribution indicates that the increased gender gap in math and science uptake may represent a lost pool of talent. While data limitations do not allow me to conclusively examine the longer-term implications of the reform on earnings, it is important that future studies aim to provide better understanding of the links between studying basic mathematics and science and gender gaps in earnings. This paper also makes it clear that there are stark gender gaps in preferences for mathematics and science, which emerge before the final years of high school. While I show how these pre-existing gaps interact with policy changes, improved knowledge of when and why these differences emerge is a meaningful avenue for future studies to inform policy on how to effectively address the gender divide in STEM education.

2.8. Appendix

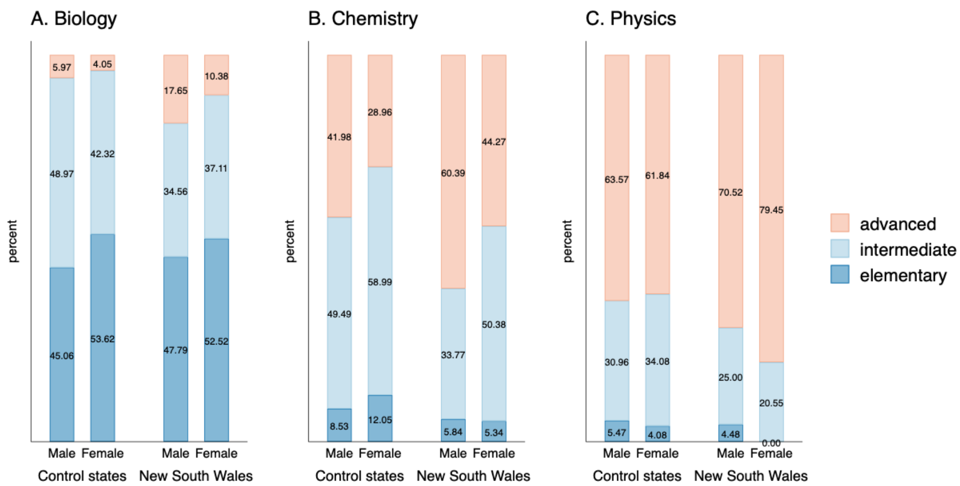
2.8.1. Background and expected results

Figure 2.1: Pre-reform math and science subject uptake



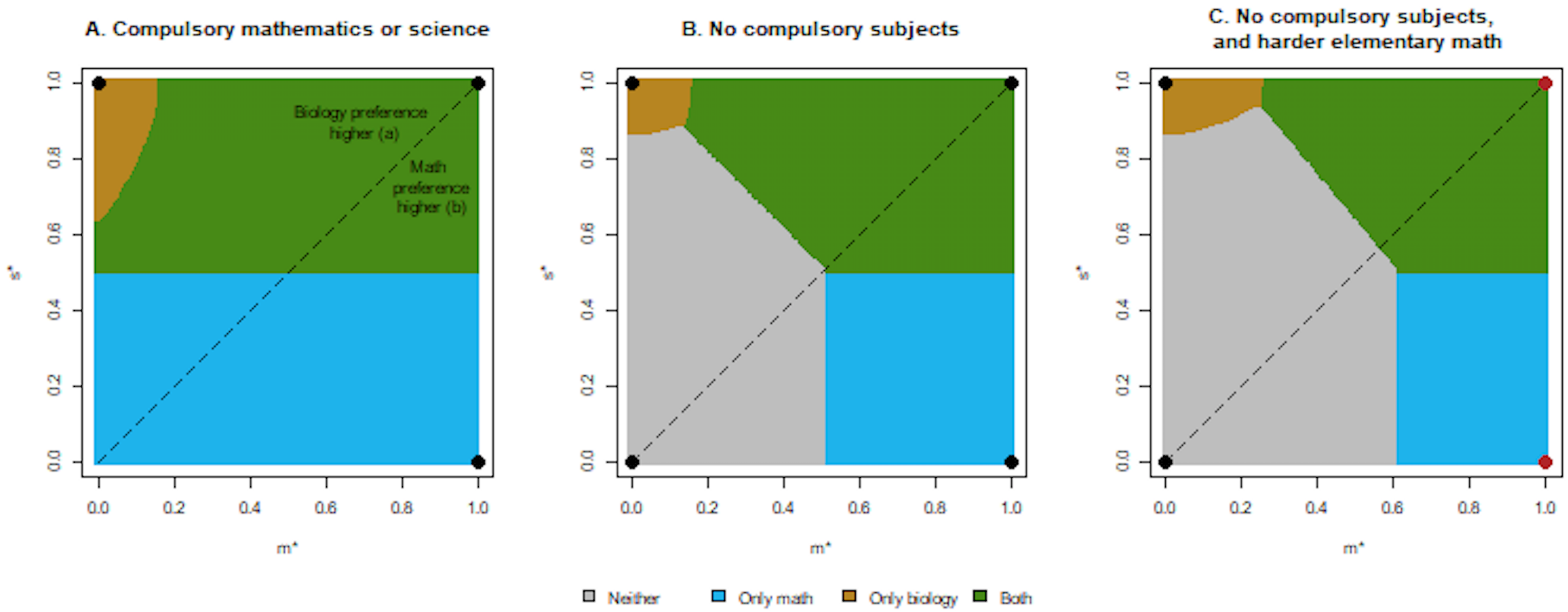
Notes: Data on Year 12 students' subject uptake in 1998 from the Longitudinal Survey of Australian Youth.

Figure 2.2: Math subjects taken by students in different science classes



Notes: Data on Year 12 students' subject uptake in 1998 from the Longitudinal Survey of Australian Youth.

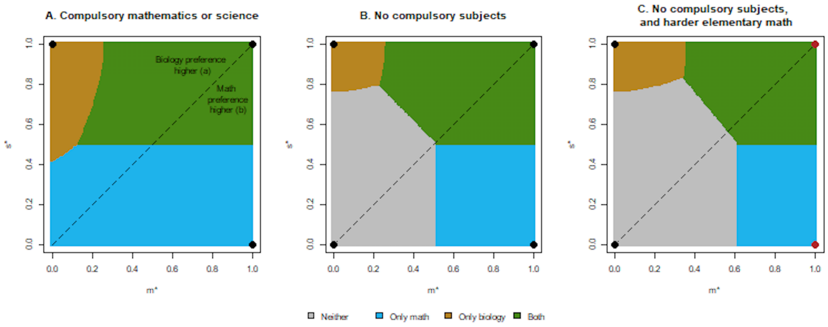
Figure 2.3: Expected effects of the 2001 reform



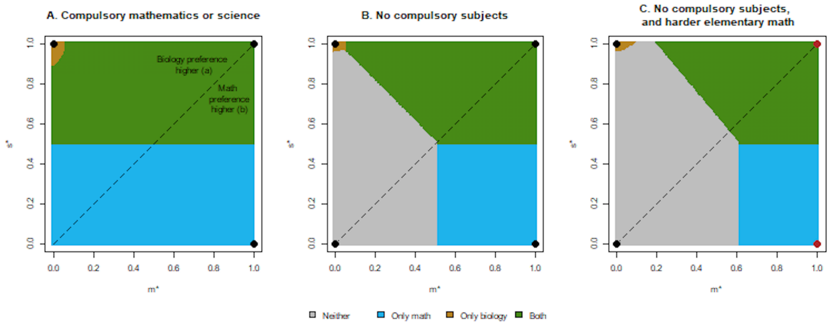
Notes: $a = 0.7$ $m_{11} = 1.2m_{01}$. The new allocations in red in Panel C reflect the change in difficulty of the elementary mathematics class.

Figure 2.4: Expected effects of the reform: different calibrations

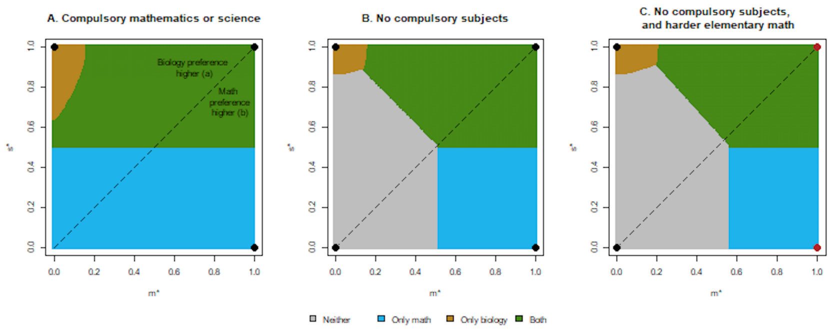
4.1. $a = 0.5\ m_{11} = 1.2m_{01}$



4.2. $a = 0.9\ m_{11} = 1.2m_{01}$



4.3. $a = 0.7\ m_{11} = 1.1m_{01}$



4.4. $a = 0.7\ m_{11} = 1.3m_{01}$

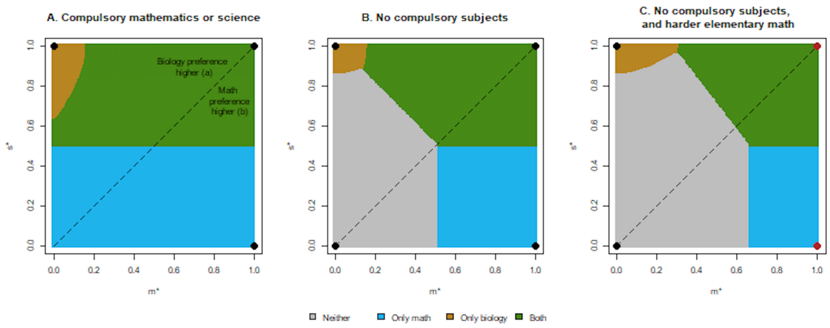


Table 2.1: Pre-reform summary statistics and means tests

	New South Wales				Control states				Difference	
	N	Female	F-M	p-value	N	Female	F-M	p-value	(3) -(7)	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Score: math	1446	12.87	-1.22	0.00	5167	12.86	-0.89	0.00	-0.33	0.10
Score: English [4]	1446	14.66	0.83	0.02	5167	14.30	0.44	0.00	0.39	0.19
Confidence: math	1446	3.52	-0.25	0.00	5167	3.50	-0.24	0.00	-0.00	0.98
Confidence: English	1446	3.64	0.06	0.26	5167	3.71	0.14	0.00	-0.09	0.14
Confidence: overall [5]	1446	3.67	-0.11	0.07	5167	3.72	-0.07	0.01	-0.04	0.46
Plan to work in STEM [6]	1446	0.04	-0.05	0.00	5167	0.04	-0.09	0.00	0.04	0.01
Socio-economic index [7]	1446	37.56	-0.59	0.65	5167	36.46	-0.60	0.40	0.01	0.99
Number of siblings	1446	2.03	0.15	0.01	5167	2.07	0.07	0.05	0.08	0.12
Mother: university	1446	0.50	-0.09	0.02	5167	0.51	-0.03	0.01	-0.06	0.10
Father: university	1446	0.64	-0.07	0.02	5167	0.64	-0.04	0.01	-0.03	0.28
Mother: not from Aus	1446	0.33	-0.04	0.28	5167	0.30	-0.01	0.44	-0.02	0.46
Father: not from Aus	1446	0.36	-0.05	0.10	5167	0.33	-0.01	0.57	-0.04	0.19
Mother works in STEM	1446	0.01	0.00	0.13	5167	0.01	0.00	0.46	0.00	0.31
Father works in STEM [8]	1446	0.07	0.01	0.40	5167	0.06	-0.01	0.32	0.02	0.19

Notes: [2] The test scores range from 1 to 20 and are from standardized mathematics and English tests conducted when students were in Year 9 of high school. [3] Reported in Year 9: 1 = Significantly below than average, 2 = Below average, 3 = Average, 4 = Above average, 5 = A lot better than average. [4] This includes those students who didn't have any specific plans yet when they were asked in Year 9. The analysis reflects that and includes a dummy denoting whether students already had a preferred field when they were asked. [5] International Socio-Economic Index of parents occupational status provided by the survey. Calculated as the average of parents' indices where both working or not missing; in cases where a parent's index is missing, the other parent's index is used instead. [6] Similar to future plans, this includes those who didn't respond, which helps explain why numbers are low; this is reflected with a dummy in the analysis as well.

2.8.2. Results

Table 2.2: Trends in subject areas

Dependent variable:	Math	Science	Computing and technology	English	Humanities
	(1)	(2)	(3)	(4)	(5)
NSW × t	-0.0140***	-0.0330***	0.0242***	0.0297***	-0.0326***
	(0.00435)	(0.00559)	(0.00487)	(0.00382)	(0.00432)
Female × NSW × t	0.00240	-8.87e-06	0.00295	-0.000838	-0.000537
	(0.00593)	(0.00761)	(0.00663)	(0.00519)	(0.00588)
Observations	9864	9864	9864	9864	9864
R-squared	0.0727	0.0130	0.0699	0.0940	0.353
F-statistic	110.4	18.59	105.8	146.1	769.1

Notes: Standard errors reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1. Data of Year 12 subject uptake from the 1990-1995 annual waves of the Australian Youth Surveys and the 1998 Year 12 cohort of the Longitudinal Survey of Australian Youth.

Table 2.3: Trends in subject area combinations

Dependent variable:	Math and science	Math only	Science only	No math or science
	(1)	(2)	(3)	(4)
NSW × t	-0.0384***	0.0245***	0.00545*	0.00850**
	(0.00564)	(0.00515)	(0.00301)	(0.00359)
Female × NSW × t	0.00316	-0.000759	-0.00317	0.000767
	(0.00768)	(0.00702)	(0.00410)	(0.00489)
Observations	9864	9864	9864	9864
R-squared	0.0301	0.0144	0.0282	0.0411
F-statistic	43.63	20.62	40.92	60.42

Notes: Standard errors reported in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data of Year 12 subject uptake from the 1990-1995 annual waves of the Australian Youth Surveys and the 1998 Year 12 cohort of the Longitudinal Survey of Australian Youth.

Table 2.4: Trends in specific subjects and subject combinations

Dependent variable:	Elementary	Biology	Elementary only	Biology only	Elementary and biology
	(1)	(2)	(3)	(4)	(5)
NSW × t	-0.0154***	-0.0115**	-0.00187	0.00211	-0.00622*
	(0.00524)	(0.00519)	(0.00420)	(0.00238)	(0.00358)
Female × NSW × t	0.000589	-0.000282	0.00313	0.000211	-0.00181
	(0.00714)	(0.00707)	(0.00572)	(0.00324)	(0.00487)
Observations	9864	9864	9864	9864	9864
R-squared	0.0694	0.0310	0.0386	0.0210	0.0222
F-statistic	105.1	45.09	56.50	30.26	31.98

Notes: Standard errors reported in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data of Year 12 subject uptake from the 1990-1995 annual waves of the Australian Youth Surveys and the 1998 Year 12 cohort of the Longitudinal Survey of Australian Youth.

Table 2.5: The 2001 and subject areas

	STEM subjects			English and humanities		
	(1)	(2)	(3)	(4)	(5)	(6)
Subject area		Mathematics			English	
NSW × Post	-0.0590**	-0.0788**	-0.0855**	-0.0452*	-0.0524	-0.0536*
	(0.0240)	(0.0364)	(0.0364)	(0.0232)	(0.0320)	(0.0320)
Female × NSW × Post	-0.0674**	-0.0795***	-0.0836***	0.0110	0.00816	0.00339
	(0.0287)	(0.0271)	(0.0277)	(0.0207)	(0.0250)	(0.0226)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.288	0.339	0.389	0.457	0.502	0.524
Subject area		Science			Humanities	
NSW × Post	-0.00753	-0.0185	-0.0354	-0.170***	-0.193***	-0.190***
	(0.0387)	(0.0481)	(0.0421)	(0.0421)	(0.0497)	(0.0485)
Female × NSW × Post	-0.0673	-0.0959*	-0.0934**	0.0695**	0.0825**	0.0757**
	(0.0425)	(0.0504)	(0.0465)	(0.0319)	(0.0320)	(0.0333)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.104	0.173	0.243	0.213	0.268	0.280
Subject area		Technology				
NSW × t	-0.0482	-0.0621	-0.0660			
	(0.0537)	(0.0553)	(0.0547)			
Female × NSW × t	0.0194	0.0385	0.0428			
	(0.0446)	(0.0500)	(0.0532)			
Observations	12491	12491	12491			
R-squared	0.144	0.209	0.219			
School FE	No	Yes	Yes	No	Yes	Yes
Ability, Confidence and STEM Plan	No	No	Yes	No	No	Yes

Notes: Standard errors clustered on the state-area-school sector level reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1. All regressions include socio-economic status and dummies for country of birth, parents' country of birth and education, and number of siblings. Ability includes dummies for mathematics and English test scores, confidence is self-reported rank in mathematics, English and overall school performance, and STEM plan stands for whether a student indicated they wanted to work in a STEM field in the future. STEM plan also includes a dummy for whether a student provided an answer to the question. The proxies for ability, confidence and plans to work in STEM were reported when students were in Year 9.

Table 2.6: The 2001 reform and subject levels

	Mathematics subjects			Science subjects		
	(1)	(2)	(3)	(4)	(5)	(6)
Subject		Elementary			Biology	
NSW × t	-0.00739	-0.0504	-0.0402	0.0377	0.0341	0.0303
	(0.0295)	(0.0359)	(0.0325)	(0.0402)	(0.0389)	(0.0372)
Female × NSW × Post	-0.122***	-0.0827**	-0.0715**	-0.112***	-0.0872*	-0.0768*
	(0.0354)	(0.0360)	(0.0305)	(0.0406)	(0.0498)	(0.0439)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.102	0.168	0.259	0.0845	0.139	0.160
Subject		Intermediate			Chemistry	
NSW × t	-0.0491	-0.0538	-0.0557	-0.00349	-0.0129	-0.0301
	(0.0400)	(0.0483)	(0.0434)	(0.0198)	(0.0221)	(0.0213)
Female × NSW × Post	0.00439	0.0450	0.0392	0.00221	0.00844	0.0106
	(0.0357)	(0.0436)	(0.0406)	(0.0207)	(0.0264)	(0.0310)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.0460	0.101	0.163	0.0746	0.128	0.259
Subject		Advanced			Physics	
NSW × Post	-0.0409	0.00138	-0.0110	-0.0152	0.0100	-0.00554
	(0.0404)	(0.0301)	(0.0283)	(0.0240)	(0.0301)	(0.0220)
Female × NSW × Post	0.0734*	-0.0211	-0.0287	0.0302	-0.000723	-0.00493
	(0.0414)	(0.0355)	(0.0313)	(0.0256)	(0.0311)	(0.0329)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.0870	0.151	0.287	0.115	0.165	0.297
School FE	No	Yes	Yes	No	Yes	Yes
Ability, Confidence and STEM Plan	No	No	Yes	No	No	Yes

Notes: Standard errors clustered on the state-area-school sector level reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1. All regressions include socio-economic status and dummies for country of birth, parents' country of birth and education, number of siblings. Ability includes dummies for mathematics and English test scores, confidence is self-reported rank in mathematics, English and overall school performance, and STEM plan stands for whether a student indicated they wanted to work in a STEM field in the future. STEM Plan also includes a dummy for whether a student provided an answer to the question. The proxies for ability, confidence and plans to work in STEM were reported when students were in Year 9.

Table 2.7: Main result: the 2001 reform and its effect on elementary mathematics and biology

	No elementary math combinations			Elementary math combinations		
	(1)	(2)	(3)	(4)	(5)	(6)
Subject combination	Biology only			Elementary annd Biology		
NSW × Post	0.0197**	0.0235**	0.0229**	0.0134	0.0106	0.0105
	(0.00758)	(0.00923)	(0.00896)	(0.0249)	(0.0184)	(0.0194)
Female × NSW × Post	0.0154	0.0333**	0.0373***	-0.0882***	-0.0800***	-0.0734***
	(0.0139)	(0.0142)	(0.0138)	(0.0285)	(0.0250)	(0.0245)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.0593	0.114	0.120	0.0660	0.114	0.136
Subject combination	No math or science			Elementary only		
NSW × Post	0.0464**	0.0722*	0.0782**	-0.00708	-0.0144	-0.00520
	(0.0223)	(0.0368)	(0.0356)	(0.0309)	(0.0427)	(0.0393)
Female × NSW × Post	0.0607**	0.0579**	0.0605***	-0.0415	0.0104	0.0141
	(0.0246)	(0.0243)	(0.0192)	(0.0391)	(0.0482)	(0.0443)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.265	0.316	0.364	0.0504	0.115	0.164
School FE	No	Yes	Yes	No	Yes	Yes
Ability, Confidence and STEM Plan	No	No	Yes	No	No	Yes

Notes: Standard errors clustered on the state-area-school sector level reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1. All regressions include socio-economic status and dummies for country of birth, parents' country of birth and education, and number of siblings. Ability includes dummies for mathematics and English test scores, confidence is self-reported rank in mathematics, English and overall school performance, and STEM plan stands for whether a student indicated they wanted to work in a STEM field in the future. STEM plan also includes a dummy for whether a student provided an answer to the question. The proxies for ability, confidence and plans to work in STEM were reported when students were in Year 9.

Table 2.8: The 2001 reform, high school completion and admission marks

Dependent variable	Obtained High School Certificate			Enrolled in university			University admission score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NSW × Post	-0.0188**	-0.0227*	-0.0220*	0.0226	0.0354	0.0240	3.226*	4.119**	3.208**
	(0.00908)	(0.0121)	(0.0118)	(0.0303)	(0.0315)	(0.0251)	(1.702)	(2.038)	(1.473)
Female × NSW × Post	0.0121	0.00988	0.00812	-0.0115	-0.0240	-0.0276	0.421	-0.219	-1.309
	(0.0101)	(0.0129)	(0.0130)	(0.0380)	(0.0254)	(0.0233)	(2.052)	(1.805)	(1.507)
Observations	12491	12491	12491	12491	12491	12491	8887	8887	8887
R-squared	0.0240	0.0769	0.0843	0.142	0.208	0.333	0.275	0.350	0.526
School FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ability, Confidence and STEM Plan	No	No	Yes	No	No	Yes	No	No	Yes

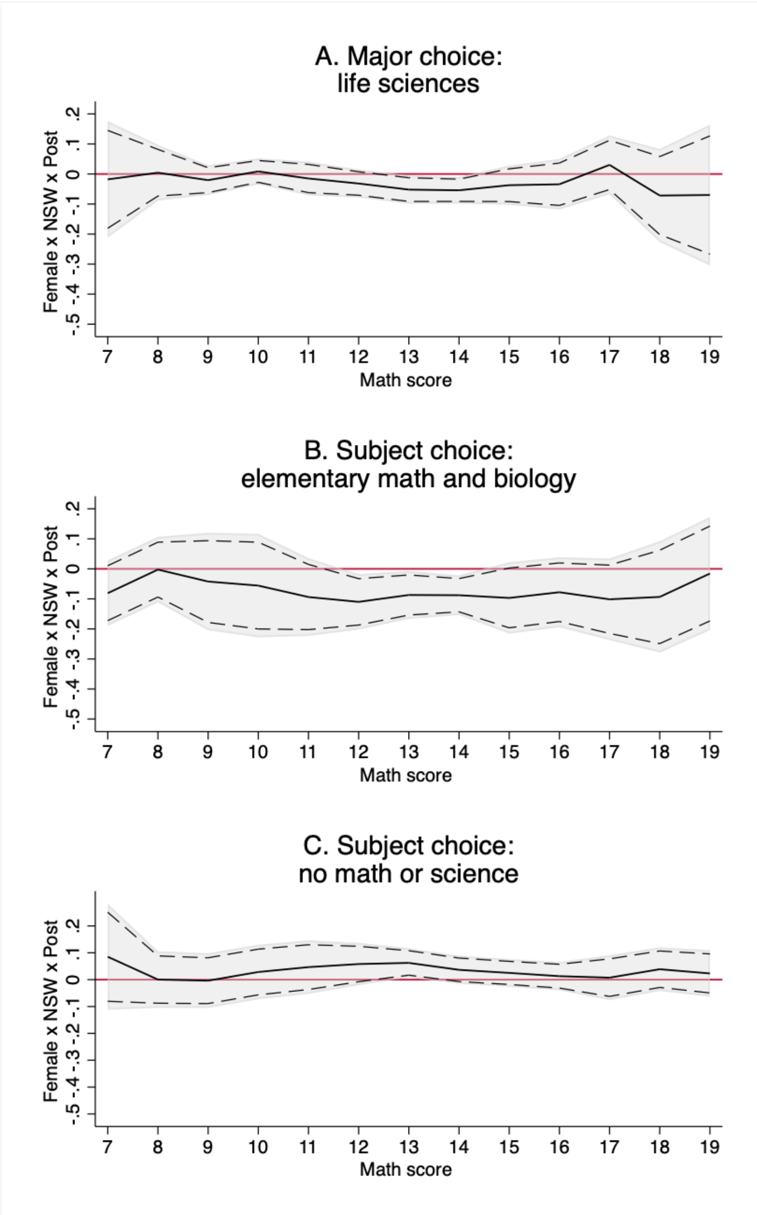
Notes: Standard errors clustered on the state-area-school sector level reported in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. All regressions include socio-economic status and dummies for country of birth, parents' country of birth and education, and number of siblings. Ability includes dummies for mathematics and English test scores, confidence is self-reported rank in mathematics, English and overall school performance, and STEM plan stands for whether a student indicated they wanted to work in a STEM field in the future. STEM plan also includes a dummy for whether a student provided an answer to the question. The proxies for ability, confidence and plans to work in STEM were reported when students were in Year 9.

Table 2.9: The 2001 reform and university majors

	STEM Majors			Non-STEM Majors		
	(1)	(2)	(3)	(4)	(5)	(6)
Majors	Life sciences			Medical sciences		
NSW × Post	0.0289***	0.0169*	0.0130	-0.0166	-0.00484	-0.00491
	(0.00976)	(0.00857)	(0.00782)	(0.0107)	(0.0119)	(0.0121)
Female × NSW × Post	-0.0218*	-0.0291**	-0.0286**	0.00539	-0.00655	-0.00728
	(0.0120)	(0.0126)	(0.0137)	(0.0184)	(0.0190)	(0.0185)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.0137	0.0516	0.0687	0.0328	0.0703	0.0837
Majors	Technical sciences			Arts and Humanities		
NSW × Post	0.0158	0.0267	0.0215	0.0191	0.0239	0.0261
	(0.0189)	(0.0200)	(0.0188)	(0.0160)	(0.0305)	(0.0290)
Female × NSW × Post	-0.00731	-0.0204	-0.0222	-0.0519**	-0.0376	-0.0414
	(0.0165)	(0.0170)	(0.0170)	(0.0254)	(0.0397)	(0.0350)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.0370	0.0659	0.0974	0.0391	0.0799	0.104
Majors	Engineering			Business and Social Sciences		
NSW × Post	0.00709	0.0208	0.0128	-0.0262	-0.0302	-0.0225
	(0.0149)	(0.0163)	(0.0161)	(0.0194)	(0.0223)	(0.0220)
Female × NSW × Post	-0.00515	-0.00607	-0.00528	0.0524**	0.0393	0.0302
	(0.0165)	(0.0183)	(0.0165)	(0.0261)	(0.0263)	(0.0280)
Observations	12491	12491	12491	12491	12491	12491
R-squared	0.0409	0.0719	0.126	0.0389	0.0786	0.0962
School FE	No	Yes	Yes	No	Yes	Yes
Ability, Confidence and STEM Plan	No	No	Yes	No	No	Yes

Notes: Standard errors clustered on the state-area-school sector level reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1. All regressions include socio-economic status and dummies for country of birth, parents' country of birth and education, and number of siblings. Ability includes dummies for mathematics and English test scores, confidence is self-reported rank in mathematics, English and overall school performance, and STEM plan stands for whether a student indicated they wanted to work in a STEM field in the future. STEM plan also includes a dummy for whether a student provided an answer to the question. The proxies for ability, confidence and plans to work in STEM were reported when students were in Year 9.

Figure 2.5: Heterogeneous effects of the reform by mathematics ability



Notes: Shaded area represents 95% confidence intervals; dashed line represents 90% confidence intervals. Estimates from moving samples of sizes (x-1;x+1), where x represents math scores reported on the x-axis. 5th to 95th percentile of mathematics scores is displayed.

Table 2.10: Degree completion and wages

Dependent variable	Completed undergraduate			Log (hourly wages) (completed undergraduate)			Log(hourly wages) (without undergraduate)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NSW × Post	0.0253	-0.0107	0.0194	-0.103	-0.541*	-0.558*	-0.0182	-0.0846	0.0186
	(0.0327)	(0.0374)	(0.0416)	(0.139)	(0.307)	(0.283)	(0.200)	(0.273)	(0.288)
Female × NSW × Post	0.0472	0.133**	0.0889	0.0163	0.161	0.203	-0.192	-0.210	-0.240
	(0.0391)	(0.0511)	(0.0552)	(0.202)	(0.353)	(0.343)	(0.303)	(0.543)	(0.542)
Observations	6026	6026	6026	3032	3032	3032	2510	2510	2510
R-squared	0.103	0.185	0.264	0.117	0.241	0.254	0.153	0.285	0.302
School FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ability, Confidence and STEM Plan	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Standard errors clustered on the state-area-school sector level reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1. All regressions include socio-economic status and dummies for country of birth, parents' country of birth and education, number of siblings. Ability includes dummies for mathematics and English test scores, confidence is self-reported rank in mathematics, English and overall school performance, and STEM plan stands for whether a student indicated they wanted to work in a STEM field in the future. STEM Plan also includes a dummy for whether a student provided an answer to the question. The proxies for ability, confidence and plans to work in STEM were reported when students were in Year 9.

2.8.3. Subject contents and university majors

Table 2.11: Summary of mathematics concepts in Year 12 science classes

Subject	Summary of math requirements
Biology	data analysis, basic functions and graphs
Chemistry	data analysis, functions and graphs, solving mathematical problems
Physics	data analysis, functions and graphs, mathematical modelling and analysis

Notes: Source: Biology, Chemistry, and Physics syllabuses available on NSW Board of Studies website and pre-reform science syllabuses/exams found on high-school websites.

Table 2.12: Classification of university majors

Category	Majors included
Engineering	Engineering (transport, mechanical, computer, aerospace), Surveying
Technical Science	Computer science, Mathematics (pure & applied), Physics, Geology, Astronomy, Hydrology, Chemistry
Life Science	Biology, Zoology, Marine science, Food science, Biochemistry, Physiology
Business &	Business administration, Management, Marketing, Accounting, Real estate, Actuarial studies
Social Sciences	Economics, Econometrics, Behavioral Science, Psychology
Arts & Humanities	History, Philosophy, Literature, Theater, Music, Fine arts
Education	Primary & Secondary teaching, Counseling
Medicine	Medicine, Medical science, Veterinary science
Legal studies	Law (international, taxation, etc.), Law enforcement

Notes: Biology-related majors are divided into Life sciences and medicine. Based on the UAC (2017) report, medicine majors usually require advanced or at least intermediate mathematics. Despite this, medicine is not usually considered as part of STEM. Majors were reported using FOSCTEC (Field of Study Classification of Tertiary Education Courses) in the pre-reform wave and ASCED (Australian Standard for Classification of Education) in the post-reform wave. For consistency across the waves, I generated the broad subject categories above using the FOSCTEC-ASCED conversion table provided in the LSAY technical documentation.

Chapter 3

PRIVATE SCHOOL GROWTH AND INEQUALITY IN EDUCATIONAL OUTCOMES: EVIDENCE FROM INDIA

3.1. Introduction

Improving access to education in developing countries has been one of the most important goals for international education policy over the past decades (Muralidharan, 2013). However, despite increased spending on expanding public provision and nearly universal access to free public primary schools, a striking trend has emerged: the parallel growth in fee-charging private primary schools. Recent estimates show that private schools now account for over 20 percent of primary school enrolments in low-income countries (Baum et al., 2014). This raises important questions regarding the implications of private schools for equity in educational opportunities and the optimal policy response to their growth.

India is a prominent example of the simultaneous expansion of the public resource base and growth in private provision. In 2018, when public primary schooling was universally accessible, more than 30 percent of rural primary-school aged children were enrolled in private institutions – a staggering increase from less than 19 percent in 2006 (Figure 3.1). This shift is seen as a testimony to persistent inadequate standards of instruction at government schools (Muralidharan and Kremer, 2006) and parents' increased demand for quality of schooling (Muralidharan, 2013). Statistics from rural India document a widespread failure of the education system, especially the public sector, to instill basic literacy and numeracy skills. For instance, in 2018, only about 42 percent of students at the end of primary school at rural government schools could do two-digit subtraction, compared to 54 percent in rural private schools (Figure 3.2).

These trends in enrolments and learning have been accompanied by large and growing income inequalities in learning outcomes over the past decade. Analysis of The Young Lives Survey in Andhra Pradesh revealed that disparities between primary-school students from the wealthiest and poorest quartiles have grown (Rolleston et al., 2014). Data from the two waves of the Indian Human Development Survey in Figure 3.3 show that the worsening quality of learning across rural India has disproportionately affected poorer students and gaps in basic functional numeracy and literacy skills between the top and bottom 25 percent have expanded over time. As a result, in 2012 only about 40 percent of students in the bottom income quartile could do two-digit subtraction at the end of primary school, compared to 70 percent of students at the highest quartile. Even more strikingly, only 30 percent of the poorest children could read a short story at the end of primary school, compared to 50 percent of the wealthiest ones. Such gaps in basic learning could translate into longer-term disadvantage in education and labor market opportunities, and perpetuate a cycle of poverty for a large share of the population (Rose et al. 2017; Ravallion, 2011).

Despite these striking patterns, there is conflicting evidence on the accessibility of private schools to disadvantaged groups, and little research examining the links between the expansion of private schooling and income- and gender-driven gaps in learning outcomes. Some studies have described the emergence of income stratification in private education in separate locations in India (Streuli et al., 2011; Härmä, 2011). In other analyses, private primary schools are still described as "low-fee" and believed to be accessible to students from all backgrounds (Tooley, 2013).

In this paper, I use a nationally representative dataset and show that for at least a quarter of rural students, the fees charged by private primary schools are not "low", and are, in fact, prohibitive. I demonstrate that private primary school openings increase private enrolment gaps by socioeconomic status and gender within villages. Using standardized test performance data for students aged eight to eleven, I also present evidence that increased access to private schooling could also contribute to gaps in learning outcomes between students from low- and high-income households. These findings emphasize the policy-relevance of studying the role of private primary schools, and more generally the demand for school quality, for improving early learning outcomes.

There are a number of mechanisms through which private school openings and increasing income-driven inequality in private enrolments can widen gaps in learning outcomes. If private schools improve student performance (Muralidharan and Kremer, 2006; Bold et al., 2011; Tabarrok, 2013; Andrabi et al. 2011; Singh, 2015), then, because of their better access to private schooling, higher-income students' learning would improve disproportionately. In addition, evidence from Andhra Pradesh indicates that students with high-starting ability, which are likely to be in wealthier households, are more likely to benefit from private enrolments (Rolleston and Moore, 2018). The arrival of a private school could also affect the performance of the existing public school in the village. There is evidence that the threat of competition and lower enrolments can worsen public school performance (Macleod and Urquiola, 2013; McMillan, 2004; Bukowski and Kobus, 2018). When a private school opens in a village, it may also "cream skim" higher-income and potentially higher-performing students away from public schools (Macleod and Urquiola, 2015). This loss of better students and high-quality peers can have negative spillovers on the performance of teachers and students in public schools (Hsieh and Urquiola, 2006).

To establish whether private schools contributed to such unequal outcomes in enrolments and learning, I exploit the geographic expansion of private primary schools in India and use the rural village- and household-level panel from the 2005 and 2012 waves of the Indian Human Development Survey. The main threat to identifying the link between private openings and learning outcomes stems from the non-random assignment of private primary schools to villages: if private schools select into locations where public provision and poor students' performance is worsening over time, any findings of increased income gaps in learning outcomes could not be plausibly attributed to the arrival of a private school in a village. I alleviate these concerns by showing that enrollments and grade repetition for poorer students were improving faster in treated relative to control villages, and that there were no baseline differences in scores. The panel-structure of the data also allows for the use of district-year fixed effects. I also report the estimates with and without village-level changes in employment, transportation and presence of higher-level private and public schools, and show that such trends had a positive influence on low-income compared to high-income students' scores in treated relative to control villages.

The analysis begins by revealing that there is a stark income-driven inequality in households' response to the opening of a private primary school in a village: students at the top 25 percent of the income distribution are 13 percentage points more likely to enrol in the newly opened private school than those at the bottom 25 percent, and few students below the median students enrol in private schools altogether. These results align with other state-level analyses that show a pattern of income stratification in private enrolments (Streuli et al., 2011; Härmä, 2011). To my knowledge, this is the first nationally-representative study to document this pattern and link improved access to private primary schooling with the increasing income divide between students attending public and private schools in India.

I also show that, at most income levels, girls are less likely to enter the newly opened private school than boys. Using household fixed effects to address the endogeneity of child gender in India (Jayachandran and Pande, 2017; Bhalotra and Cochrane, 2010), I confirm that this finding is driven by parents' systematic preference for investing in boys' education, especially in poorer households. This is consistent with findings in the literature that income constraints exacerbate discriminatory investments between boys and girls in India (Rose, 1999). To my knowledge, there is only one other paper using nationally representative data to estimate such a gap within households (Maitra, Pal, and Sharma, 2016) and no other work formally linking the increasing gender gap in private school enrolments with the expansion of private schooling on a national scale.

Next, I present evidence that private school expansion is linked with increasing income-driven gaps in mathematics and reading scores among primary school children. I show that in villages with private school openings gaps between students above and below the median income increased, and this result was concentrated among those at the bottom and top 25 percent of the distribution, for whom the gap in private enrolments was highest. In particular, the gap in mathematics expanded by 4.5 percent out of a score of 3, while the one in reading – by 7.5 percent out of a score of 4. Translating these results into specific achievement levels, these gaps were driven by fewer poor students reaching the higher learning benchmarks of doing subtraction and division in math, and reading a story. Relative to students at the top 25 percent of the income distribution, those at the bottom were approximately six percentage points less likely to be able to subtract two digit numbers, seven percentage points less likely to be able to do division and nine percentage points less likely to be able to read a one-page story.

The findings in this paper shed light on the nation-wide implications of the expansion of private schooling for equity in educational opportunities and learning in India. They provide insights to the long-standing debate on the trade-off between efficiency and equity linked with the growth of private provision in developing countries (Muralidharan and Sundararaman, 2016; Pedro, Leroux and Watanabe, 2015). The fact that a large share of low-income and female students are precluded from entering private institutions means that, even if private schools are more efficient, they cannot be relied upon to address deficiencies in public provision and bridge existing gaps in educational outcomes. Studies on the causes on consequences of human capital investments, including by gender, are also closely related (Attanasio et al. 2020, Jensen, 2010; Munshi and Rosenzweig, 2006). The paper also contributes to the literature on the relationship between private school provision and student performance (Bold et al., 2011; Tabarrok, 2013; Andrabi et al. 2011; Singh, 2015), as well as works on the aggregate effects of private expansion on the school system (Hsieh and Urquiola, 2006; Muralidharan and Sundararaman, 2015). This is also the first study to document a link between the growth of private schooling and the persistent inequalities in learning outcomes between high- and low-income students in India. The findings clearly show that, in order for policy interventions to successfully improve the quality of learning in the country, the interplay between private and public provision should not be ignored.

The remainder of this paper is organized as follows. In section 3.2, I outline the role of private schools in India's educational landscape. In section 3.3 I describe the data and in section 3.4 I discuss the empirical strategy. Section 3.5 presents the results and section 3.6 concludes.

3.2. Private schools in rural India

There are three main types of schools operating across India: government schools, government-aided private schools, and private unaided schools. Schools also vary by the levels of education they provide: primary (or lower primary, referred to as primary from here on) at ages six to eleven, middle (upper primary) at ages 11-14, lower secondary at ages 15 and 16, and upper secondary at ages 17 and 18. In 2002, the 86th Constitutional Amendment made free, compulsory elementary education a Fundamental Right. Government and government-aided primary schools are prohibited from charging tuition fees. By contrast, private unaided schools, which I refer throughout as just private schools, charge tuition fees. They conform to the stereotypical idea of what private institutions are: autonomous fee charging schools run by private managements, recruiting their own teachers and determining their salaries (Kingdon, 2020).

The number of private schools at all levels has been steadily growing in rural communities (Kingdon, 2020; Muralidharan and Kremer, 2006). Data from the two waves of the India Human Development Survey used in this paper show that only between 2005 and 2012, the percentage of villages with a private primary school grew from 31 percent to 41 percent (Figure 3.4). Government primary schools were present in almost all villages in both waves, with coverage growing from

94.5 percent to 98.5 percent. The National Sample Survey further reveals that in 2014, 22.5 percent of rural primary school aged children (6-10) attended a private school, up from 14 percent in 2007 (National Sample Survey Office 2014). These numbers are very similar to the 24.9 percent and 13.3 percent observed in the 2012 and 2005 waves of the India Human Development Survey.

Private primary schools have been found to operate better than public ones in India. Muralidharan and Kremer (2006) conduct a detailed nationally-representative survey of private primary schools in rural India. They document that in private primary schools, teachers are less likely to be absent, are younger, are more likely to have a college diploma and are more likely to come from the local community. Also, while private schools are generally smaller, they hire more teachers at lower salaries, which results in lower student-to-teacher ratios and less frequent multi-grade teaching. Muralidharan and Kremer (2006) also show that private primary schools are a lot more likely to start teaching English in early primary grades, which parents find desirable. In addition to operational efficiency, there is also evidence confirming that private school attendance is linked with better student performance as well, although some of the findings may suffer from selection bias (Muralidharan and Kremer, 2006; Bold et al., 2011; Tabarrok, 2013; Andrabi et al. 2011; Singh, 2015; French and Kingdon, 2010). On the other hand, in a randomized voucher experiment, Muralidharan and Sundararaman (2015), found private school students to have better scores only in Hindi, which is not taught in government schools in the state.

The tuition fees and supplementary spending on books, uniforms, and private lessons are a lot higher in private schools relative to government ones. Muralidharan and Kremer (2006) report that the monthly median fee charged by a private primary school in their sample was 63 rupees. This is also the median tuition fee in private primary schools in the pooled (2005 and 2012) India Human Development Survey sample (in 2005 prices) (Figure 3.6). However, attending a private school entails additional expenditure, which is not necessarily incurred in government schools. The two waves of the India Human Development Survey contain information on all school-related expenditures by households. Figure 3.6 displays the distribution of tuition fees and total fees (which include tuition fees, transportation, books, uniforms, and private tutoring) across different educational levels in government and private schools. The median total monthly expenditure associated with sending a child to a government primary school was about 19 rupees, while the monthly spending needed to send a child to a private primary school was about 141 rupees, more than seven times higher. These account for about 4 percent and 31 percent respectively of the 2005 median monthly household per capita consumption. Such costs would therefore be prohibitive for the poorer segment of the rural population.

There are both demand and supply factors driving the decision to send a child to a private primary school. In their qualitative study in Andhra Pradesh, Morrow and Wilson (2013) reveal that, despite the higher fees that parents need to pay in order to send their child there, parents overwhelmingly prefer to enrol their children in private schools, if affordable. In particular, the education private schools provide is perceived to be of better quality, they start teaching English earlier, and, as parents pay high tuition, private schools are believed to be more accountable (Morrow and Wilson, 2013). In the interviews conducted, parents also reported that they decided which child to send to the private school, based on how much they believe the child would benefit from better schooling. On the supply side, Morrow and Wilson (2013) reveal that distance is the most important determinant, especially in the case of primary-aged children. Primary-aged children are too young for boarding and parents often don't have the time or means to transport them everyday (Morrow and Wilson, 2013).

In this setting, the opening of a private primary school in a village, would induce a significant number of parents to enrol their children in the private school. The fact that government primary schooling was almost free and available in most villages makes it unlikely that the arrival of a private primary school would affect parents' decision whether to enrol children in school altogether. Instead, it would change the composition of enrolments, with more children studying in a private primary school. Given the high costs required, the opening of a private primary school in a village would induce large increases in private enrollments among higher-income students, while many low-income ones would not be able to enter. In addition, if boys are perceived as more likely to benefit from private schooling, parents would be more likely to enrol boys, which would also create gender gaps in private enrolments[1].

3.3. Data

This paper uses the 2005 and 2012 waves of the Indian Human Development Survey (IHDS). In both waves of the survey, in rural areas, along with individual- and household-level data collection, a village questionnaire was administered. It contains information on village government, demographics and infrastructure, including private and government schools providing different levels of education

[1] Note that private primary schools are usually co-educational.

(primary, middle, secondary, and higher secondary). More than 90 percent of villages were covered in both waves, and approximately 83 percent of households in the first wave were re-interviewed in 2012 as well.

Crucially, in both waves standardized learning tests were administered to all children aged eight to eleven within surveyed households. In both survey waves, students were tested on functional literacy, numeracy and writing skills and scored according to the same scale. Table 3.1 contains details on the learning benchmarks against which results within subjects were measured in the two waves.

Panel A of table 3.2 displays summary statistics from 2005, the first wave, for villages with and without private school openings between 2005 and 2012. As even in 2005 public primary schools were almost universally accessible, there are no baseline differences in the presence of a public primary option between "treated" and "control" villages. At the same time, villages with future private primary openings were more likely to have both government and private higher-level (middle, upper and lower secondary) schools. The treated villages were also more likely to be larger, have lower employment-to-population ratios and share of agricultural employment, and were more likely to have a bus stop.

Panel B of table 3.2 reveals that, while private schools opened in larger villages, which were less dependent on agriculture, better connected, and with better public and private school presence, there were no baseline differences in learning outcomes and school attendance between students below and above the median levels of income in treated and control villages. Panel B also confirms patterns of stark income inequality in learning and school attendance. For instance, in all villages, wealthier students were about 15 percentage points more likely to be able to write with two or less mistakes than the poorer group. The average student in the wealthier group could also read a paragraph and do subtraction, while lower-income students could on average only read words and recognize some numbers. Finally, lower-income students in all villages were between two and three percentage points less likely to attend school.

3.4. Empirical strategy

The empirical analysis throughout this paper uses households' 2005 per capita monthly consumption as a proxy for ex-ante socioeconomic status prior to the arrival of the private school. Ex-ante consumption expenditure is a more accurate representation of socioeconomic status than income, because the sample is rural and many households engage in subsistence agriculture without a regular source of formal income (Meyer and Sullivan, 2003). From here on, I use consumption and income interchangeably to refer to households' 2005 level of monthly per capita consumption.

3.4.1. Private school openings and enrolments by socioeconomic status and gender

The first specification explores how private-school enrollment varies by household income and gender. The sample is restricted to children aged six to eleven, the official primary-school ages, in the time period t of measurement. Equation 3.1 estimates the marginal effects on private-school enrollments of private school openings, by household income quantiles, and gender, and interactions of these variables. To explore the relationship between enrollments and household income in detail, income quantiles are measured in 20 buckets containing 5 per cent in each, with each quantile denoted Q_q, where $Q_q = [5(q-1), 5q)$ and $q = 1, ..., 20$.

$$\begin{aligned}
&InPrivateSchool_{ihvt} = \alpha_1 P_{vt} + \alpha_2 P_{vt} \times Girl_{ihvt} \\
&+ \sum_{q=1}^{20} \beta_{q,p} P_{vt} * \mathbb{1}(C_h \in Q_q) + \sum_{q=1}^{20} \beta_{q,g,p} Girl_{ihvt} * P_{vt} * \mathbb{1}(C_h \in Q_q) \\
&+ \sum_{q=1}^{20} \beta_{q,g} Girl_{ihvt} * \mathbb{1}(C_h \in Q_q) + \sum_{q=1}^{20} \beta_q \mathbb{1}(C_h \in Q_q) \\
&+ \mathbf{X}_{ihvt}\prime\gamma + \delta_v + \delta_{dt} + \epsilon_{ihvt},
\end{aligned} \tag{3.1}$$

The dependent variable $InPrivateSchool_{ihvt}$ is equal to one if child i in household h and village v was enrolled in a private school in year t. The right-hand side variables include: P_{vt} is equal to 1 if a private primary school was in village v in year t and is 0 otherwise; C_h is the household's 2005 monthly per capita consumption; and the vector $\mathbf{X}_{ihvt}$ contains child- and household-level covariates, including dummies for age and birth order, parents' age and level of education, as well as dummies for households' religion and caste. δ_v are village fixed effects, which limit the estimated effects to villages in which private schools opened between the two waves, and δ_{dt} are district-year fixed effects, which remove any effects of different trends in private enrolments across districts between the two waves.

In equation 3.1, $\alpha_1 + \beta_{q,p}$ represents the change in private enrolments for boys and $\alpha_2 + \beta_{q,g,p}$ is the gender gap in private enrolments at interval $[5(q-1), 5q)$. If private schools were prohibitively expensive for a share of the population or if parents valued investing in boys' education more (Jayachandran and Pande, 2017), we would observe the emergence of gaps in private enrolments between poorer and wealthier families, as well as between female and male students.

3.4.2. Gender gaps in private enrolments by income

We next explore whether gender gaps in private-school enrollment are related to parents' systematic preference for investing in boys' education. To identify the implications of private-school expansion for potential gender gaps in learning, it is crucial to understand what the driver of gender gaps in private enrolments are and if they vary by income.

The IHDS data allows for the use of household fixed effects to address the endogeneity of child gender across households driven by son preference in India. Son preference is a major challenge in attributing the observed gaps to systematic gender bias rather than unobservable parental characteristics, in a setting where, due to pre-natal sex selection and son-biased fertility stopping, boys and girls are not born in the same households (Bhalotra and Cochrane, 2010). In such an environment, the same unobserved parental characteristics that may affect the sex-composition of children could also systematically affect the educational opportunities of boys and girls differently. For instance, girls may be more likely to live in poorer households, with less access to ultrasound technology, who would also be less likely to enrol children in a private school. Therefore, comparing boys and girls within the same household enables us to uncover whether there is a systematic preference for enrolling boys, and not girls, in private schools, netting out the effect of fixed unobservable household characteristics related to children's gender.

I estimate the following equation:

$$\begin{aligned} InPrivateSchool_{ihvt} &= \beta_1 P_{vt} + \beta_2 Girl_{ihvt} \times P_{vt} \\ &+ \beta_3 Girl_{ihvt} + \mathbf{X}_{ihvt}'\gamma + \delta_h + \delta_{dgt} + \epsilon_{ihvt}, \end{aligned} \tag{3.2}$$

where δ_h is a household fixed effect, δ_{dgt} is a district-year-gender fixed effect which controls flexibly for differential trends in private enrolments by gender across districts, and the other variables are defined as above. To uncover the direction in which household characteristics typically associated with child gender affect the coefficients, I estimate (3.2) separately with and without household fixed effects.

I also estimate equation 3.2 for households below and above the median 2005 level of per capita consumption. If the returns to investing in boys' schooling are perceived to be higher (Alderman and King, 1998), it is more likely that gender gaps in private enrollments would be systematically observed in lower-income households, where parents would be less likely to afford enrolling more than one child to a private school. On the other hand, at higher incomes, where parents can afford to send both children to a private school, characteristics linked to unobserved differences in preferences across households are more likely to drive any gender gaps in private enrolments. These have important implications for observed educational outcomes. For instance, if girls in low-income households live with boys who are enrolled in private schools, there may be spillovers influencing both children. Using household fixed effects on the sample of re-interview households (for which there is data on 2005 consumption) limits the sample size significantly, as it would require the same households to have boys and girls aged six to eleven in both waves. Nevertheless, comparing these results still provides important insights into the dynamics of the gap over the income distribution.

3.4.3. Private school openings and gaps in learning outcomes

The empirical strategy outlined in this subsection estimates whether the opening of a private school in a village contributed to worsening income and gender inequalities in literacy, numeracy and writing ability of primary school children.

The main challenge for the empirical strategy lies in the non-random assignment of private schools to villages, which could raise concerns of different trends in learning outcomes and of reverse causality. If private schools opened in places where public schooling was deteriorating and educational outcomes for poor students and girls trended negatively, it would not be possible to attribute emerging gaps in performance to the arrival of private schooling in the village. For instance, although they do not examine trends, Muralidharan and Kremer (2006) document that private primary schools opened in villages where public-school teacher absenteeism was higher.

I address these concerns in three ways. Table 3.3 reports pre-trends in school attendance and in the probability of being below the correct grade for age (a proxy for grade repetition). The estimates show that private schools tended to open in villages with *declining* income and gender gaps in educational outcomes. If trends in school attendance and grade repetition also reflect trends in students' performance (Hunt, 2008; Lewin, 2008; Ampiah and Adu-Yeboah, 2009)[2], the model would be underestimating the potential widening of gaps in learning outcomes associated with private school openings. In addition, as previously observed, table

[2]The literature consistently links low achievement to the risk of early dropouts, as well as grade repetition and overage enrolments.

3.2 shows that there were no significant pre-existing differences in test scores between students in control and treated villages. Finally, the village- and household-panel structure of the data permits the inclusion district-year fixed effects, interacted with income quantile and gender. While it is plausible that prevalence of private schools would be higher in areas where public schools under perform (Muralidharan and Kremer, 2006), and potentially worsen over time, this is less likely to be the case within districts, where individual schools would likely choose to physically locate in less remote and larger villages.

To evaluate the implications of private openings for learning outcomes, I define an independent variable measuring the number of years between ages six and eleven in which a child was exposed to private primary school presence in the village. This is appropriate because it is the cumulative years of exposure at school ages that would affect the accumulation of knowledge in mathematics, reading and writing. For example, ten-year-old children in villages where the private school opened two years ago would not have the same exposure as ten-year-olds in villages where the primary school opened four years ago.[3] Tested children were between ages eight and eleven, so this variable in practice measures the number of years in which the child was exposed since age six.

Thus, I estimate the following equation:

$$\begin{aligned}&TestScore_{ihvt} = \beta_1 YearsExposure_{ihvt}\\&+\beta_2 \mathbb{1}(C_h \leq Q_q)\times YearsExposure_{ihvt}\\&+\beta_3 \mathbb{1}(C_h \leq Q_q)\times YearsExposure_{ihvt}\times Girl_{ihvt}\\&+\beta_4 YearsExposure_{ihvt}\times Girl_{ihvt}+\beta_5 Girl_{ihvt}\times \mathbb{1}(C_h \leq Q_q)\\&+\beta_6 \mathbb{1}(C_h \leq Q_q)+\mathbf{X}_{ihvt}\prime\gamma+\delta_v+\delta_{dqt}+\delta_{dgt}+\epsilon_{ihvt},\end{aligned} \tag{3.3}$$

where $YearsExposure_{ihvt}$ is equal to the number of years in which child i in household h, village v and year t was exposed to a private primary school in their village. The indicator $\mathbb{1}(C_h \leq Q_q)$ is equal to 1 if 2005 household monthly per capita consumption C_h is in the quantile range Q_q, where Q_q will be specified with the results. The vector $\mathbf{X}_{ihvt}$ contains child- and household-level covariates, including dummies for age and birth-order, parents' age and level of education, as well as dummies for households' religion and caste. δ_v is a vector of village fixed effects, and δ_{dqt} and δ_{dgt} are district-year-quantile and district-year-girl fixed

[3]Changing schools at the primary level because of the arrival of an alternative schooling option is not uncommon. Using data from the Young Lives Survey in Andhra Pradesh, Woodhead et al. (2014) reveal that 16 percent of rural eight-year-olds had already changed at least one school, and more than half of these changes were from a government to a private school.

effects. The dependent variable $TestScore_{ihvt}$ measures students' scores in mathematics, reading and writing obtained in the IHDS tests.

While comparing students within the same village would, to some extent, address the exposure to different aggregate shocks across treated and control villages, the opening of a private school could also be associated with changes in trends in village labor markets and infrastructure, which could affect the returns to schooling differently by socioeconomic status and gender. Such improvements in infrastructure, for instance, are likely to benefit disadvantaged groups disproportionately by making transportation and services more accessible. Nevertheless, I account for such village-level changes over time by including covariates for total employment, share of agricultural employment, the opening of a government and private middle and high schools, and the availability of a bus stop in 2005 and 2012, interacted with child gender and socioeconomic status. To confirm the direction in which such changes over time affect the estimates, I report results with and without these time-varying village-level characteristics.

3.5. Results

3.5.1. Private school openings and enrolments by socioeconomic status

Panel A of figure 3.7 displays the marginal effects of a private school opening by household socioeconomic status, estimated in model 3.1 in the previous section. Panel B presents the change in the resulting difference between male and female private enrolments along the income distribution.

Panel A confirms that increased access to private schooling has different implications for poorer and richer households. Those above the median increase enrollments by twice as much as those below the median following an opening, with starker contrasts between households at the bottom and the top of the income distribution. While private enrolments for the top 25 percent increase by nearly 20 percentage points, those at the bottom 25 percent remain largely unaffected. Indeed, for those at the bottom 25 percent of the distribution, the average monthly private school expenses of 141 Rs. would represent at least 40 per cent of 2005 monthly per capita consumption[4].

Panel B reveals a striking pattern: when households are first able to afford private

[4]the 25th percentile is at about 370 Rs.

enrolments (around the 25th percentile), a large gender gap emerges. The size of this gap indicates that between around the 25th and 40th income percentile, boys represent most of private enrolments. This is consistent with findings that the presence of credit constraints exacerbates gaps in investments between boys and girls (Rose, 1999). Panel B also reveals that the gender gap in access to private primary schooling decreases around the median but is still present at higher incomes. This is in line with evidence of stronger, culturally embedded, son preference among higher caste and Hindu families (Jayachandran and Pande, 2017).

Table 3.4 confirms the findings in figure 3.7. It demonstrates the emergence of a nearly 9 percentage point gap in private enrolments between households above and below the median levels of income (column (1)), and a 13 percentage point gap between the top and bottom 25 percent (column (2), with the sample limited to the top and bottom income quartiles). It also shows that the emerging gender gap in private enrolments did not differ above and below the median (column (1)) and the gender gaps were not significantly different at the top 25 and bottom 25 percent. This potentially reflects the fact that those at the bottom of the distribution were not able to enrol any children, while those at the top could afford to send all. Columns (3) and (4) of table 3.4 also confirm that private enrolments for older students were not affected.

3.5.2. Private school openings and the gender gap in private enrolments

Given the patterns in the gaps in enrolments, I next present some evidence of a causal link between child gender, private school expansion and private school enrolments. Column (2) of table 3.5 reports estimates from model 3.2 and shows that, controlling for fixed observable and unobservable household characteristics that make gender endogenous, parents are three percentage points less likely to enrol girls in a private school when it opens in their village.

Columns (4) and (6) reveal that this is driven by households below the median level of consumption and disappears in those above. This indicates that poorer households, who have to make the choice which child to enrol, consistently choose boys. Among higher income households, it is unobservable preferences related to children's gender composition that drive the gap, and the effect disappears once household fixed effects are added to the model. This indicates that among above-median-income households, girls are actually less likely to be born among those prone to sending children to private schools. This could be, once again, associated with the fact that son-preference is stronger among higher-caste families.

3.5.3. Private school openings and gaps in learning outcomes

In this section, I present evidence that gaps in elementary mathematics, reading and writing skills between high- and low-income students expanded more in villages with private primary school openings.

Table 3.6 contains the results from equation 3.3 comparing students above and below the median of the 2005 income distribution. Table 3.7 displays the estimates from comparing students from the bottom and top income quartile. The results reveal that, in villages where a private primary school opened, income-related gaps in mathematics and reading test scores increased. They also show that these gaps were driven by students at the top and bottom 25 percent of the income distribution. Figure 3.7 and Table 3.7 in the first results subsection showed that the increase in the gap in private enrolments between the top and bottom 25 percent was 13 percentage points, with barely any students from the bottom quartile enrolling.

These results are consistent with the previously outlined mechanisms. On the one hand, if private schools lead to better student performance, the gap would be largest among those least and most likely to enroll in a private school. This would be further amplified if highest-income students, due to better initial preparation and socialization, are most likely to benefit from improved instruction (Rolleston and Moore, 2018). On the other hand, if the arrival of a private school reduces instruction quality or student performance in public schools, by decreased enrolments, motivation and loss of high-performing students, it would be those who have no option other than the public school system who would be most affected.

The results also indicate that there are no significant gender gaps emerging. This could be due to the fact that the gender gaps at the top and bottom 25 percent were not significantly different from zero: the least well off were less likely to send any children to a private school, and the most well off were less likely to have to face a trade-off. It could also be due to the specific dynamics of the gap in enrolments in households below and above the median outlined in the previous subsection.

The results in tables 3.6 and 3.7 also confirm that trends in school openings, employment and village infrastructure in treated villages were more likely to favour lower-income students, with the coefficients increasing in magnitude once these time-varying village-level characteristics are included. The magnitude of the coefficients on the gap in test scores, with and without these covariates, is significant. At the average years of exposure of 3.05 years in the treated villages, they translate to 15 percent of a score unit (out of three) in Mathematics and about 30 percent of a score unit (out of four) in Reading.

To better understand the implications of these findings, it is important to identify at what levels of skills the test-score results were concentrated. I define a new dependent variables S_{ihvtj} which is equal to 1 if a child's literacy or numeracy score in a subject was was higher or equal to j. I estimate equation 3.3 with these new dependent variables for the sample of students at the bottom and top income quartiles. Figure 3.8 displays the results. They reveal that the gaps grew at the highest levels of mathematics and reading ability measured. Relative to students at the top 25 percent of the income distribution, those at the bottom were approximately six percentage points less likely to be able to subtract two digit numbers, seven percentage points less likely to be able to do division and nine percentage points less likely to be able to read a one-page story. These findings are consistent with what would be expected following a private school opening. If higher achieving students select into private schools (Hsieh and Urquiola, 2006), improvements of learning outcomes would be at higher levels of literacy and numeracy. If private schools have negative effects on public provision, it is likely that these effects would be strongest at higher levels of mathematics and reading competence. These results are also particularly relevant to the Indian context, where, beyond the very basics, student performance has been poor and declining over time (Figure 3.2).

These results measure the aggregate effects of private primary school arrival in a village and are consistent with either private schools' better performance or negative spillovers on the public school system. Even if the results for high-income students are not significant, the setup does not allow for discerning between these two mechanisms, which is an important gap that future research should address.

3.6. Conclusion

The expansion of private schooling in India reflects a trend of private schools' growing importance in the developing world, where they coexist with free public schools. While in some contexts (such as Latin America) they commonly cater to the richest population, in South Asia, and particularly in India, they have often been seen as accessible to the wider population. In such a scenario, by compensating for shortcomings of public provision, private schools would hold the promise of improved instruction and better outcomes for all.

In this paper, I show that this is not the case and that the rapid expansion of private schooling in rural India has the potential to deepen inequalities in learning outcomes. The finding that, for at least a quarter of the population, these schools are inaccessible, indicates that the private option cannot be considered pro-poor or equitable, and that private schools cannot make up for the current shortcomings of the public education system. The paper also demonstrates that the expansion of private schooling leads to increasing gaps in educational investments between boys and girls, especially for parents below the median levels of income, who can afford to send only some of their children to a private school.

The sizeable increases in these gaps, resulting from the expansion of private primary schooling across India, could have negative implications for equality in learning outcomes. I provide evidence that in villages with private school openings, gaps in mathematics and reading achievement widen noticeably between the bottom and top income quartiles. These findings are consistent with existing evidence of private schools' better performance and previous findings of negative effects of private school openings on the quality of public school provision.

There are important caveats in the empirical strategy of this paper. The non-random assignment of private schools to villages means that these results could be driven by negative trends in the performance of public schools in villages where private schools choose to locate. I alleviate these concerns by showing that enrollments and grade repetition for poorer students were improving faster in treated relative to control villages, and that there were no baseline differences in scores. The data also allows for the use of district-year fixed effects. However, this does assuage doubts completely, as I am unable to to examine trends in learning outcomes prior to the first survey wave or control for village-level trends in learning outcomes.

The second caveat stems from the fact that the empirical strategy is unable to identify the mechanisms behind the observed gaps in scores. Instead, the reported estimates reflect the potential aggregate effect of private school openings within a community. Understanding whether such gaps are attributable to private schools' ability to obtain better outcomes, or to negative spillover effects on public school students and public school performance is an important avenue for future research.

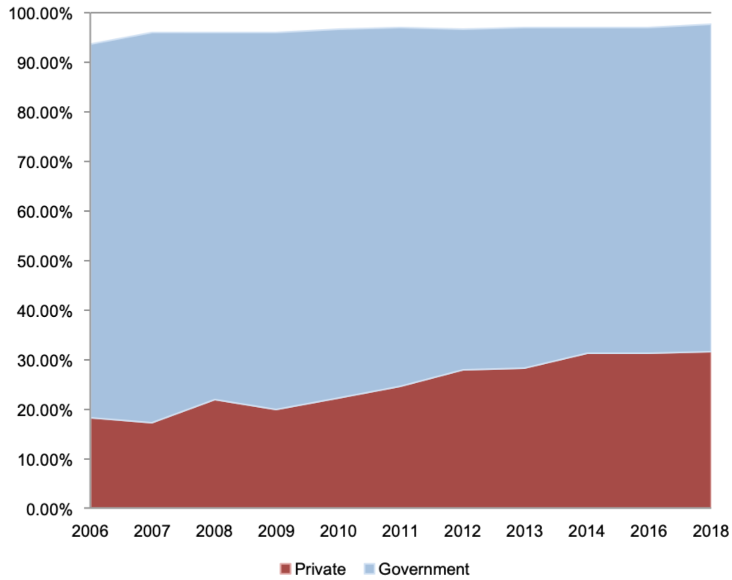
Notwithstanding these caveats, the discussion of the implications that the rise of private provision has for learning outcomes is relevant and very timely in the Indian context. While the past two decades were marked by impressive expansion of the government schooling infrastructure and the attainment of almost universal primary enrolments, the priority over the next one will be dedicated to improving learning outcomes in government schools. This is evidenced by the country's 2020

National Educational Policy, which has committed unprecedented resources to an overhaul of the pre-primary and primary education stages in hopes of advancing the quality of learning. The findings presented in this study underscore some of the important challenges that lie ahead and affirm that acknowledging the role of private provision needs to be an integral part of the process.

3.7. Appendix

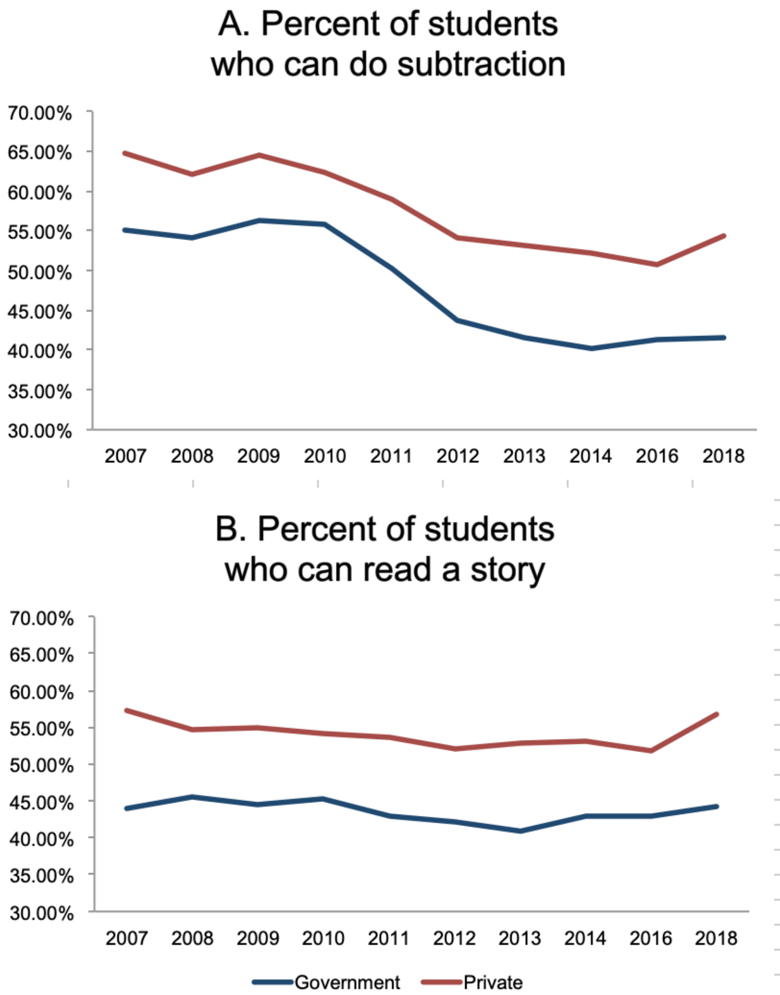
3.7.1. Background and descriptive statistics

Figure 3.1: Enrolments in private and government primary schools in Rural India



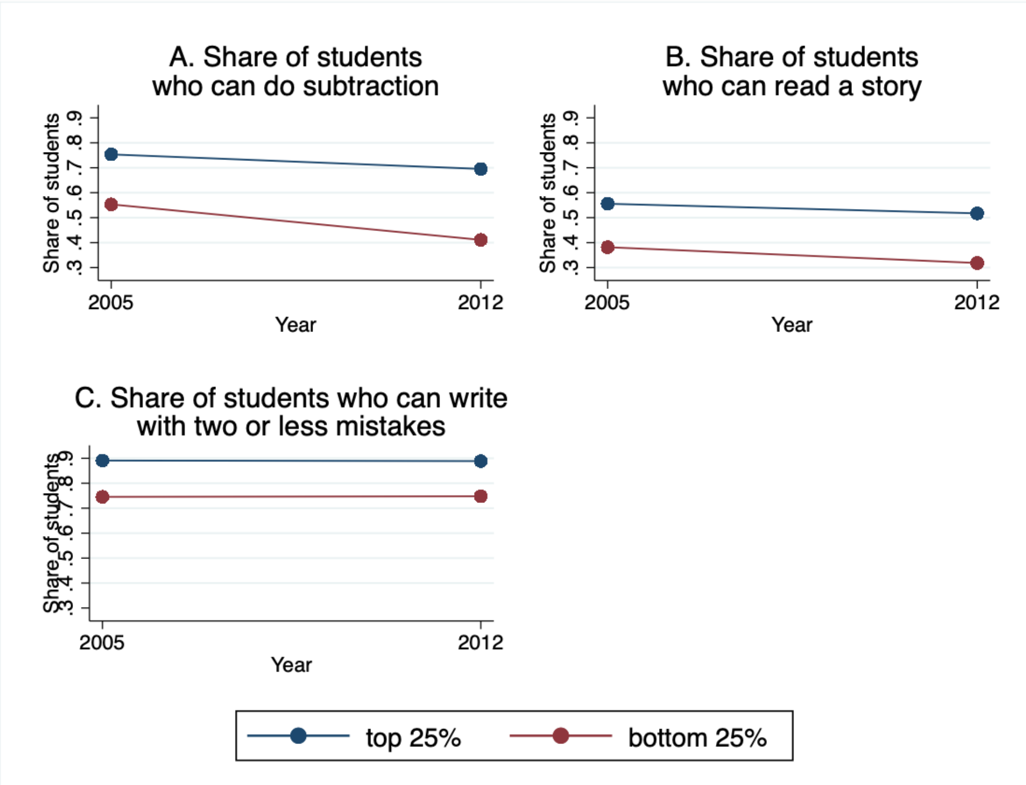
Notes: Annual Status of Education Report 2018. Children aged eight to ten years old (between grades two and five of primary school).

Figure 3.2: Learning outcomes of rural students
at the end of primary school



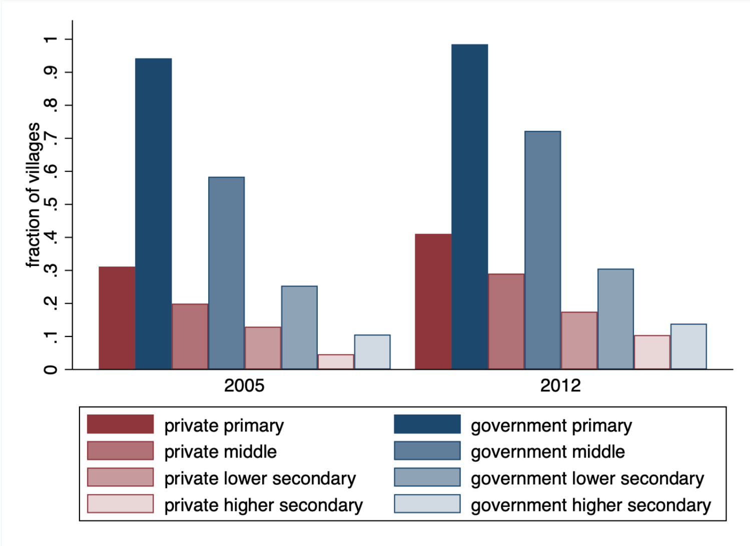
Notes: Annual Status of Education Report 2018. Sample of students in grade five (the final year of primary school). Subtraction refers to subtraction of two-digit numbers. Stories are a page long, at the level of grade two).

Figure 3.3: Income inequality in learning outcomes over time



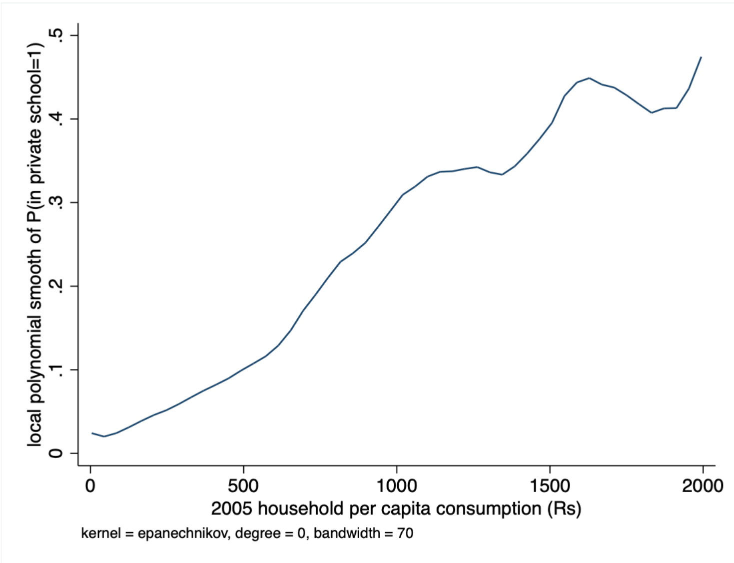
Notes: Test scores from India Human Development Survey 2005 and 2012. Sample of rural students in the last year of primary school (grade five). Top and bottom 25 percent of household per capita consumption in the respective wave. In Panels A and B, Similarly to the learning outcomes measured in the Annual Status of Education Report, stories are a page long, at grade-two level, and subtraction refers to subtraction of two-digit numbers. Panel C displays ability to write a short sentence with two or less mistakes.

Figure 3.4: Geographic expansion of private and government schools between 2005 and 2012



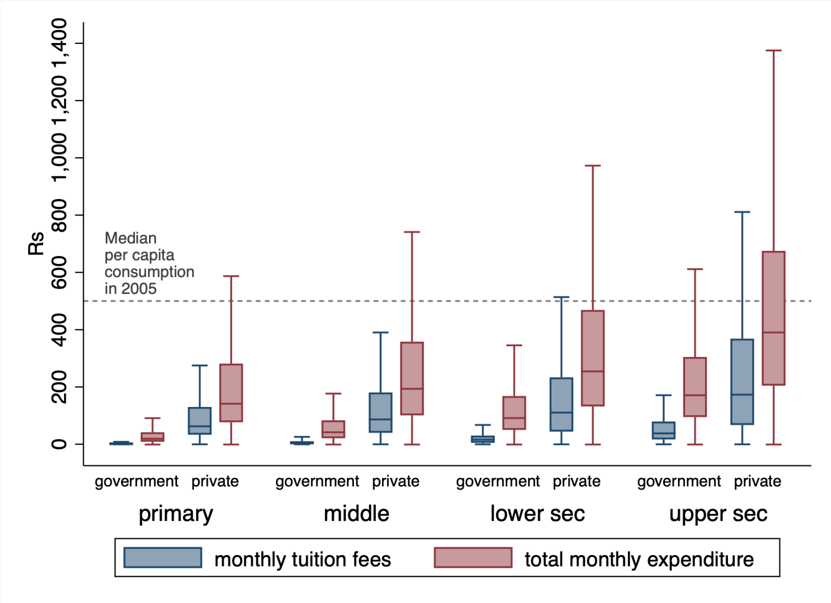
Source: India Human Development Survey 2005 and 2012.

Figure 3.5: Private primary attendance and household socioeconomic status



Notes: India Human Development Survey 2005 and 2012. Sample of children aged 6-11. Levels of household consumption displayed up to the 97th percentile.

Figure 3.6: Tuition fees and total expenditure in rural private and government schools



Notes: Pooled data from IHDS 2005 and 2012. 2012 fees adjusted for inflation to 2005 levels. Total expenditure includes tuition fees, as well as additional expenses on uniforms, books, transportation, and private lessons.

Table 3.1: Coding of test scores in IHDS 2005 and 2012

Score	0	1	2	3	4
Reading score (0 to 4)	Cannot read at all	Recognizes letters but not words	Can read words but not sentences	Can read a short paragraph but not a full story	Can read a page-long story
Mathematics score (0 to 3)	Does not recognize numbers	Recognizes some numbers but cannot do arithmetic operations	Can subtract two-digit numbers	Can divide a three-digit number by a one-digit number	
Writing score (0 and 1)	Cannot write at all	Writes a sentence with two or less mistakes			

Notes: India Human Development Survey. Tests were administered to all children aged 8-11 in interviewed households.

Table 3.2: Summary statistics of villages with and without private primary schools

	All^{+}	N	No private school	N	Private school between 2005-2012	N	Diff (5)-(3)	SE (5)-(3)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Village Level Characteristics (2005)								
Government primary	0.94	1240	0.93	678	0.93	172	-0.00	0.02
Government higher	0.61	1240	0.49	678	0.75	172	0.26***	0.04
Private higher	0.22	1240	0.06	678	0.30	172	0.25***	0.04
Employment to population (15-64): female	0.25	1240	0.29	678	0.23	172	-0.05***	0.02
Employment to population (15-64): male	0.48	1240	0.49	678	0.46	172	-0.04**	0.02
Share of non-agricultural employment: female	0.32	1059	0.26	601	0.32	151	0.06*	0.03
Share of non-agricultural employment: male	0.52	1230	0.46	672	0.53	170	0.07**	0.03
Monthly household per capita consumption	639.99	1240	584.35	678	631.94	172	47.60	35.61
Village has a bus stop	0.53	1182	0.47	656	0.56	166	0.09**	0.04
Percentage of HHs with electricity	67.29	1214	64.89	663	64.40	169	2.99	0.87
Vilage has piped water	0.50	1240	0.48	678	0.49	172	0.01	0.04
Number of HHs in village	581.17	1240	357.63	678	646.59	172	288.96***	51.16
B. Children's education (2005)								
Math score (8-11): Below median	1.20	3056	1.17	1706	1.26	431	0.08	0.07
Math score (8-11): Above median	1.68	2825	1.63	1329	1.57	424	-0.06	0.07
Reading score (8-11): Below median	2.18	3056	2.19	1706	2.23	431	0.03	0.10
Reading score (8-11): Above median	2.75	2825	2.75	1329	2.61	424	-0.14	0.09
Writing score (8-11): Below median	0.56	3056	0.57	1706	0.56	431	-0.004	0.03
Writing score (8-11): Above median	0.73	2825	0.73	1329	0.68	424	-0.05	0.04
School attendance (8-11): Below median	0.97	3056	0.97	1706	0.96	431	-0.01	0.01
School attendance (8-11): Above median	0.99	2825	0.99	1329	0.99	424	-0.004	0.006
Repeated grade (8-11): Below median	0.06	3056	0.07	1706	0.05	431	-0.02	0.015
Repeated grade (8-11): Above median	0.05	2825	0.06	1329	0.06	424	-0.001	0.02

Notes: $^{+}$ Including treated before 2005. "Below median" stands for households below the median level of 2005 household consumption. Standard errors in Panel B clustered on the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

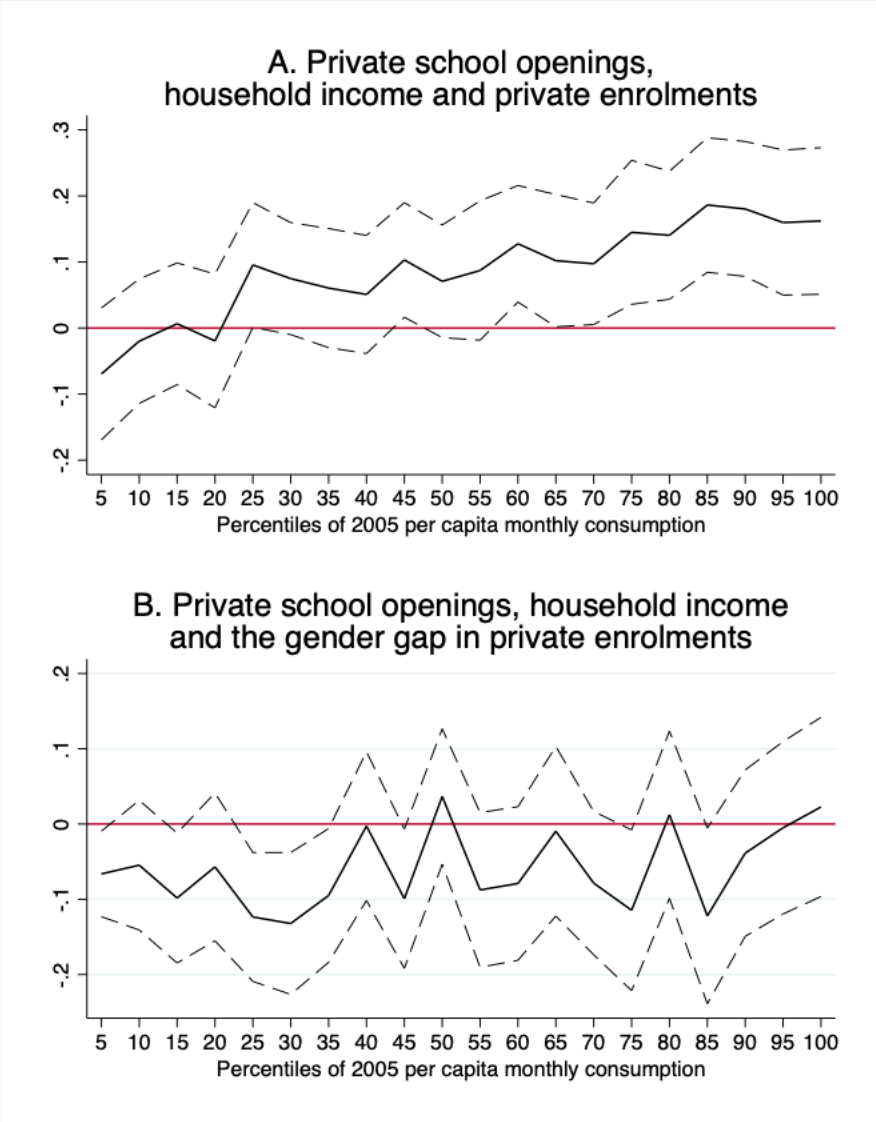
Table 3.3: Pre-trends in school attendance and correct grade for age

Dependent variable	Attended school			Below correct grade		
Period	2000-2005	1990-2005	1980-2005	2000-2005	1990-2005	1990-2005
	(1)	(2)	(3)	(4)	(5)	(6)
Private primary × Year	-0.00994	-0.00808***	-0.00612***	0.0102	0.00544***	0.00524***
	(0.00784)	(0.00189)	(0.00127)	(0.00656)	(0.00161)	(0.00111)
Private primary × Below 50% × Year	0.0230**	0.0107***	0.00597***	-0.00858	-0.00619***	-0.00704***
	(0.0112)	(0.00271)	(0.00184)	(0.00938)	(0.00231)	(0.00160)
Private primary × Below 50% × Girl × Year	-0.0188	-0.00484	-0.00233	0.0210	-0.00237	0.00261
	(0.0160)	(0.00387)	(0.00264)	(0.0134)	(0.00330)	(0.00230)
Observations	27,263	67,240	82,840	27,263	67,240	82,840
R-squared	0.031	0.021	0.030	0.060	0.035	0.059

Notes: Data of children aged 8-11 from IHDS 2005. Retrospective village-, year-, age-, income-, and gender-level samples of students aged 8 to 11 in villages which did not have a private primary school in 2005. Private primary is equal to 1 if a private primary school opened in the village between 2005 and 2012. ”Below median” is equal to 1 if the household was below the median level of 2005 consumption. School attendance in year t is calculated using reported years of schooling in 2005 and assuming a starting age of 6. ”Below correct grade” indicates that a student had obtained fewer that *Age-6* years of education in year t. Standard errors clustered on the village level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

3.7.2. Results

Figure 3.7: Private school openings and enrolments by socioeconomic status



Notes: Sample of students in re-interview households. Coefficients from equation 3.1. Panel A: $\alpha_1 + \beta_{q,p}$. Panel B: $\alpha_2 + \beta_{q,g,p}$. Dashed lines represent 95% confidence intervals. Standard errors clustered on the village level.

Table 3.4: Private school openings and inequality in private school enrolments

Dependent variable	Enrolled in private school			
Age group	6 to 11		12 to 18	
	(1)	(2)	(3)	(4)
Private primary	0.127***	0.187***	0.00139	0.00442
	(0.0250)	(0.0380)	(0.0199)	(0.0285)
Private primary × Girl	-0.0383**	-0.0203	0.00114	-0.00499
	(0.0191)	(0.0304)	(0.0153)	(0.0211)
Private primary × Below 50%	-0.0867***		-0.0169	
	(0.0246)		(0.0169)	
Private primary × Below 50% ×Girl	-0.0141		0.00220	
	(0.0251)		(0.0203)	
Private primary × Below 25%		-0.130***		-0.0435
		(0.0436)		(0.0270)
Private primary × Below 25% × Girl		-0.0613		0.00452
		(0.0406)		(0.0277)
Observations	19006	11989	20740	13179
R-squared	0.437	0.518	0.350	0.425
Mean of Dependent Variable	0.139	0.139	0.102	0.102

Notes: Sample of students in re-interview households. Below 50% denotes below the median level of household monthly per capita consumption in 2005. Below 25% denotes below the 25th percentile of household montly per capita consumption in 2005. Models in columns (2) and (4) compare the 25th to the 75th percentile. Standard errors clustered on the village level in parentheses.

Table 3.5: Child gender, private school openings and gaps in enrolments

Dependent variable	Enrolled in private school					
Income group	All		Below 50%		Above 50%	
	(1)	(2)	(3)	(4)	(5)	(6)
Private primary	0.0634***	0.0673***	0.0607**	0.0581*	0.0927**	0.109**
	(0.0209)	(0.0241)	(0.0293)	(0.0339)	(0.0384)	(0.0468)
Private primary × Girl	-0.0411***	-0.0327***	-0.0558***	-0.0447***	-0.0583***	-0.0166
	(0.0109)	(0.0109)	(0.0139)	(0.0141)	(0.0192)	(0.0203)
Household FE	No	Yes	No	Yes	No	Yes
Observations	23624	23624	10263	10263	8743	8743
R-squared	0.391	0.815	0.372	0.750	0.479	0.854
Mean of Dependent Variable	0.139	0.139	0.139	0.139	0.139	0.139

Notes: Estimates from model 3.2. Sample of all students ages 6 to 11 in columns (1) and (2). Sample of students in re-interview households in columns (3) to (6). Below and Above 50% mean respectively below and above the median level of 2005 household monthly per capita consumption. Standard errors clustered on the village level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.6: Private school openings and gaps in learning outcomes between top and bottom 50% of income

Dependent variable	Math score		Reading score		Writing score	
	(1)	(2)	(3)	(4)	(5)	(6)
Years exposure	0.00121	0.00620	-0.0217	-0.00781	-0.0131	-0.00924
	(0.0137)	(0.0142)	(0.0344)	(0.0350)	(0.0115)	(0.0117)
Years exposure × Girl	-0.0122	-0.0141	-0.0345	-0.0465	-0.00632	-0.00869
	(0.0129)	(0.0134)	(0.0298)	(0.0306)	(0.0106)	(0.0112)
Years exposure × Below 50%	-0.0292**	-0.0375**	-0.0707**	-0.0947***	-0.0110	-0.0166
	(0.0149)	(0.0153)	(0.0344)	(0.0361)	(0.0114)	(0.0121)
Years exposure × Girl × Below 50%	0.0196	0.0199	0.0551	0.0566	0.0109	0.0115
	(0.0183)	(0.0183)	(0.0403)	(0.0402)	(0.0143)	(0.0143)
Girl x Year x Village Covariates	No	Yes	No	Yes	No	Yes
Below 50% x Year x Village Covariates	No	Yes	No	Yes	No	Yes
Observations	9716	9716	9716	9716	9716	9716
R-squared	0.439	0.440	0.684	0.684	0.650	0.650
Mean of Dependent Variable	1.434	1.434	2.460	2.460	0.649	0.649

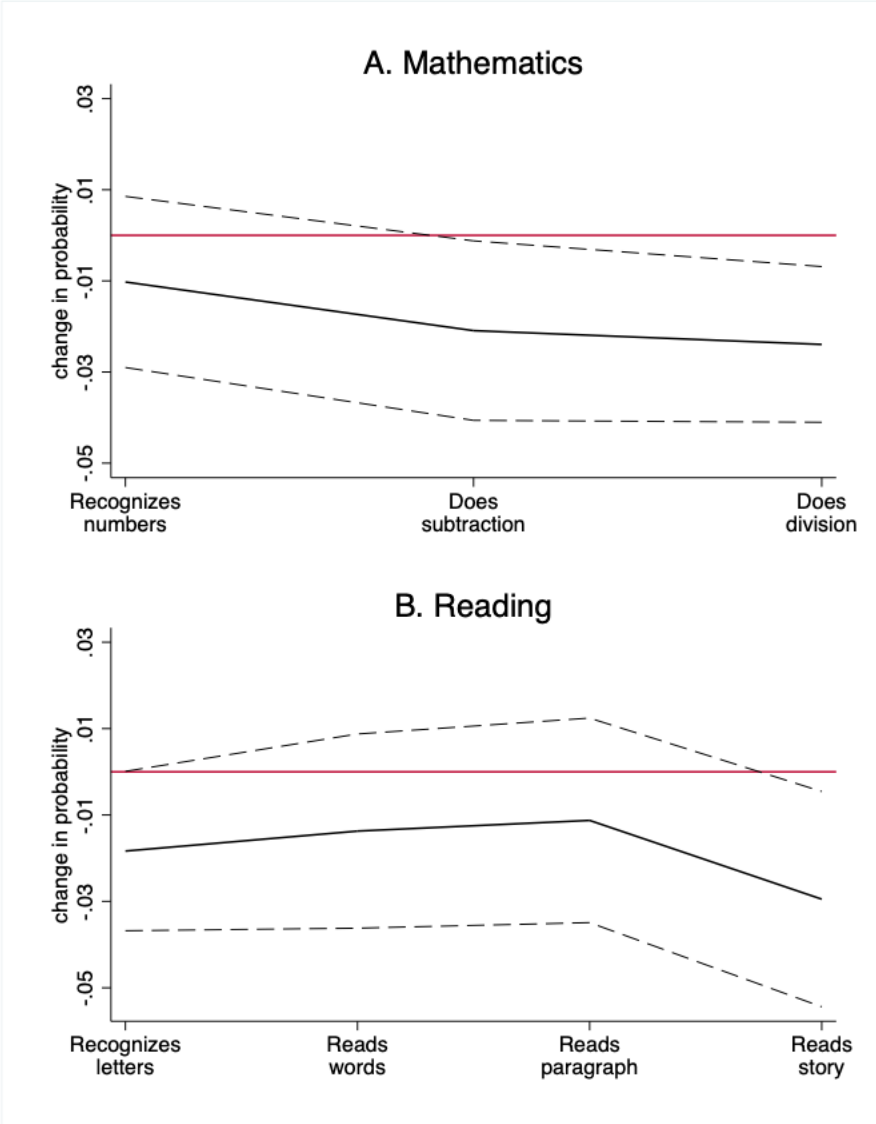
Notes: Sample of students in re-interview households. Below 50% is a dummy equal to 1 if a household was below the median level of 2005 monthly per capita consumption. Estimates from equation 3.3. Years exposure is equal to the number of years since age six in which a child was exposed to a private school within their village. Village covariates include share of agricultural employment, the availability of a bus stop, and the presence of middle, as well as lower- and upper- secondary private and government schools. Standard errors clustered on the village level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.7: Private school openings and gaps in learning outcomes between top and bottom 25% of income

Dependent variable	Math score		Reading score		Writing score	
	(1)	(2)	(3)	(4)	(5)	(6)
Years exposure	0.0157	0.0236	-0.0435	-0.0315	0.00587	0.00754
	(0.0190)	(0.0203)	(0.0576)	(0.0600)	(0.0201)	(0.0202)
Years exposure × Girl	-0.00836	-0.00786	0.0252	0.0132	0.00184	0.00222
	(0.0176)	(0.0190)	(0.0541)	(0.0584)	(0.0183)	(0.0193)
Years exposure × Below 25%	-0.0551***	-0.0693***	-0.118*	-0.121*	-0.0332	-0.0397*
	(0.0204)	(0.0220)	(0.0615)	(0.0626)	(0.0220)	(0.0215)
Years exposure × Girl × Below 25%	0.0258	0.0255	0.0219	0.0262	-0.00219	-0.00242
	(0.0240)	(0.0242)	(0.0691)	(0.0701)	(0.0238)	(0.0240)
Girl x Year x Village Covariates	No	Yes	No	Yes	No	Yes
Below 25% x Year x Village Covariates	No	Yes	No	Yes	No	Yes
Observations	6076	6076	6076	6076	6076	6076
R-squared	0.510	0.510	0.789	0.789	0.774	0.775
Mean of Dependent Variable	1.420	1.420	2.442	2.442	0.638	0.638

Notes: Sample of the top and bottom 25% income percentiles in re-interview households. Estimates from equation 3.3. Years exposure is equal to the number of years since age six in which a child was exposed to a private school within their village. Village covariates include share of agricultural employment, the availability of a bus stop, and the presence of middle, as well as lower- and upper- secondary private and government schools. Standard errors clustered on the village level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 3.8: Private school openings and learning outcomes:
levels of numeracy and literacy skills



Notes: Sample of the top and bottom 25% income percentiles in re-interview households. Estimates from equation 3.3. Estimates of β_3 from equation 3.3 with independent variable S_{ihvtj} equal to 1 if a student's numeracy or literacy level is equal to or higher than j. Dashed lines represent 95% confidence intervals. Standard errors clustered on the village level.

Bibliography

Alderman, Harold and Elizabeth M King, "Gender differences in parental investment in education," *Structural Change and Economic Dynamics*, 1998, *9* (4), 453–468.

Altonji, Joseph G., "The Effects of High School Curriculum on Education and Labor Market Outcomes," *Journal of Human Resources*, 1995, *30* (3), 409–438.

Ampiah, Joseph Ghartey and Christine Adu-Yeboah, "Mapping the incidence of school dropouts: a case study of communities in Northern Ghana," *Comparative Education*, 2009, *45* (2), 219–232.

Annual Status of Education Report, "Annual Status of Education Report - Rural," Technical Report 2018.

Aquilina, John, *Securing their future: the New South Wales Government's reforms for the Higher School Certificate*, NSW Minister for Education and Training Sydney, 1997.

Atkin, David, "Working for the future: Female factory work and child health in Mexico," *Unpublished Manuscript, Yale University*, 2009.

_ , "Endogenous Skill Acquisition and Export Manufacturing in Mexico," *American Economic Review*, August 2016, *106* (8), 2046–2085.

Attanasio, Orazio, Costas Meghir, and Emily Nix, "Human capital development and parental investment in India," *The Review of Economic Studies*, 2020, *87* (6), 2511–2541.

Ayres, Paul L and John M McCormick, "Grade 12 mathematics teachers' views on curriculum reform in New South Wales," 2006.

Baum, Donald, Laura Lewis, Oni Lusk-Stover, and Harry Patrinos, "What matters most for engaging the private sector in education," 2014.

Becker, Gary S, "Investment in human capital: A theoretical analysis," *Journal of political economy*, 1962, *70* (5, Part 2), 9–49.

_ , "A theory of marriage: Part I," *Journal of Political economy*, 1973, *81* (4), 813–846.

Behrman, Jere R, "Intrahousehold distribution and the family," *Handbook of population and family economics*, 1997, *1*, 125–187.

Better Factories Cambodia, "Towards gender equality: lessons from factory compliance assessments 2017-2018," Technical Report, International Labour Organization 2018.

_ , "Gender and Care Responsibilities: Examining the differences for garment workers in Better Factories Cambodia," *ILO-IFC Research Brief*, 2020.

Bhalotra, Sonia R and Tom Cochrane, "Where have all the young girls gone? Identification of sex selection in India," *IZA Discussion Paper*, 2010.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, and Justin Sandefur, "The high return to private schooling in a low-income country," *Center for Global Development Working Paper*, 2011, (279).

Brambilla, Irene, Daniel Lederman, and Guido Porto, "Exports, export destinations, and skills," *American Economic Review*, 2012, *102* (7), 3406–38.

Bukowski, Paweł and Martyna Kobus, "The threat of competition and public school performance: Evidence from Poland," *Economics of Education Review*, 2018, *67*, 14–24.

Burgert, Clara R, Josh Colston, Thea Roy, and Blake Zachary, "Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys," 2013.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness, and Career Choices *," *The Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Card, David and A Abigail Payne, "High school choices and the gender gap in STEM," *Economic Inquiry*, 2017.

Case, Anne and Christina Paxson, "Stature and status: Height, ability, and labor market outcomes," *Journal of political Economy*, 2008, *116* (3), 499–532.

Cheryan, Sapna, Sianna A Ziegler, Amanda K Montoya, and Lily Jiang, "Why are some STEM fields more gender balanced than others?," *Psychological bulletin*, 2017, *143* (1), 1.

Coupland, Mary, "A critical analysis of selected Australian and international mathematics syllabuses for the post-compulsory years of secondary schooling," Technical Report, NSW Board of Studies 2006.

Currie, Janet and Tom Vogl, "Early-life health and adult circumstance in developing countries," *Annu. Rev. Econ.*, 2013, *5* (1), 1–36.

Dasgupta, Indraneel, "Women's employment, intra-household bargaining, and distribution: a two-sector analysis," *Oxford Economic Papers*, 2000, *52* (4), 723–744.

de Haan, Esther and Myriam Vander Stichele, "Footloose investors: Investing in the garment industry in Africa," *Available at SSRN 1660426*, 2007.

Delaney, Judith and Paul J Devereux, "It's not just for boys! Understanding gender differences in STEM," 2019.

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

_ , Pascaline Dupas, and Michael Kremer, "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review*, 2011, *101* (5), 1739–74.

Ear, Sophal, "Cambodia's garment industry: A case study in governance," *Journal of Southeast Asian Economies*, 2013, pp. 91–105.

Filmer, Deon and Norbert Schady, "Getting girls into school: Evidence from a scholarship program in Cambodia," *Economic development and cultural change*, 2008, *56* (3), 581–617.

Fontana, Marzia, "Modelling the effects of trade on women, at work and at home: comparative perspectives," *Economie internationale*, 2004, (3), 49–80.

French, Rob and Geeta Kingdon, "The relative effectiveness of private and government schools in Rural India: Evidence from ASER data," *London: Institute of Education*, 2010.

Friedberg, Leora and Anthony Webb, "Determinants and consequences of bargaining power in households," Technical Report, National Bureau of Economic Research 2006.

Fukunishi, Takahiro, "FDI and Export Participation of Local Firms in Africa: The Case of the Kenyan Garment Industry," *Institute of Developing Economies (IDE), JETRO, Chiba*, 2010.

Glick, Peter, "Women's employment and its relation to children's health and schooling in developing countries: conceptual links, empirical evidence, and policies," *Cornell Food and Nutrition Policy Program Working Paper*, 2002, (131).

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko, "The Homecoming of American College Women: The Reversal of the College Gender Gap," Working Paper 12139, National Bureau of Economic Research April 2006.

Goldman, Amy D. and Andrew M. Penner, "Exploring international gender differences in mathematics self-concept," *International Journal of Adolescence and Youth*, 2016, *21* (4), 403–418.

Hammond, Alicia, Eliana Rubiano Matulevich, Kathleen Beegle, and Sai Krishna Kumaraswamy, "The Equality Equation: Advancing the Participation of Women and Girls in STEM," 2020.

Härmä, Joanna, "Low cost private schooling in India: Is it pro poor and equitable?," *International journal of educational development*, 2011, *31* (4), 350–356.

Heath, Rachel and A. Mushfiq Mobarak, "Manufacturing growth and the lives of Bangladeshi women," *Journal of Development Economics*, 2015, *115*, 1 – 15.

Heuveline, Patrick and Savet Hong, "One-parent families in contemporary Cambodia," *Marriage & family review*, 2016, *52* (1-2), 216–242.

Hewett, Paul and Sajeda Amin, "Assessing the impact of garment work on quality of life measures," *New York: Population Council. Mimeo*, 2000.

Hoxby, Caroline Minter, "School choice and school productivity. Could school choice be a tide that lifts all boats?," in "The economics of school choice," University of Chicago Press, 2003, pp. 287–342.

Hsieh, Chang-Tai and Miguel Urquiola, "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of public Economics*, 2006, *90* (8-9), 1477–1503.

Hunt, Frances Mary, “Dropping out from school: A cross country review of literature,” 2008.

International Labour Organization, “Wages and productivity in garment sector in Asia and the Pacific and the Arab States,” Technical Report, International Labour Organization 2015.

_ , “Cambodian Garment and Footwear Sector Bulletin: What explains strong export and weak employment figures in the Cambodian garment sector?,” Technical Report Issue 6, International Labour Organization May 2017.

_ , “Cambodia Garment and Footwear Sector Bulletin: Living conditions of garment and footwear sector workers in Cambodia,” Technical Report Dec 2018.

_ , “Cambodian Garment and Footwear Sector Bulletin: How has garment workers’ pay changed in recent years?,” Technical Report Issue 7, International Labour Organization June 2018.

International Labour Organization and National Institute of Statistics, “Cambodia Labour Force and Child Labour Survey 2012: Child Labour Report,” Technical Report, Cambodia 2013.

Jayachandran, Seema and Rohini Pande, “Why are Indian children so short? The role of birth order and son preference,” *American Economic Review*, 2017, *107* (9), 2600–2629.

Jensen, Robert T, “Economic opportunities and gender differences in human capital: Experimental evidence for India,” *National Bureau of Economic Research*, 2010.

Joensen, Juanna Schrøter and Helena Skyt Nielsen, “Is there a Causal Effect of High School Math on Labor Market Outcomes?,” *Journal of Human Resources*, 2009, *44* (1).

_ and Helena Skyt Nielsen, “Mathematics and gender: Heterogeneity in causes and consequences,” *The Economic Journal*, 2016, *126* (593), 1129–1163.

Kabeer, Naila, *The power to choose: Bangladeshi women and labor market decisions in London and Dhaka*, Verso, 2002.

Kamas, Linda and Anne Preston, “The importance of being confident; gender, career choice, and willingness to compete,” *Journal of Economic Behavior & Organization*, 2012, *83* (1), 82–97.

Kessels, Ursula, "Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects," *European journal of psychology of education*, 2005, *20* (3), 309–323.

Kheam, They and Emily Treleaven, "Women and migration in Cambodia: A further analysis of the Cambodian Rural-Urban Migration Project (CRUMP)," Technical Report, UNFPA and National Institute of Statistics 2013.

Kingdon, Geeta Gandhi, "The private schooling phenomenon in India: A review," *The Journal of Development Studies*, 2020, pp. 1–23.

Lavecchia, A.M., H. Liu, and P. Oreopoulos, "Chapter 1 - Behavioral Economics of Education: Progress and Possibilities," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds.*, Vol. 5 of *Handbook of the Economics of Education*, Elsevier, 2016, pp. 1 – 74.

Levine, Phillip and David Zimmerman, "The Benefit of Additional High-School Math and Science Classes for Young Men and Women," *Journal of Business Economic Statistics*, 1995, *13* (2), 137–49.

Lewin, KM, "Access, age and grade," *CREATE Policy Brief*, 2008, *2*.

Mack, John and Rachel Wilson, "Trends in mathematics and science subject combinations in the NSW HSC 2001-2014 by gender," 2016.

MacLeod, W Bentley and Miguel Urquiola, "Competition and educational productivity: incentives writ large," *Education policy in developing countries*, 2013, p. 243.

_ and _ , "Reputation and school competition," *American Economic Review*, 2015, *105* (11), 3471–88.

Maitra, Pushkar, Sarmistha Pal, and Anurag Sharma, "Absence of altruism? female disadvantage in private school enrollment in india," *World Development*, 2016, *85*, 105–125.

Mallick, Lindsay, Courtney Allen, and Rathavuth Hong, "Trends in Maternal and Child Health in Cambodia, 2000-2014.Further Analysis of the 2000, 2005, 2010, and 2014 Cambodia Demographic and Health Surveys," DHS Further Analysis Reports 106 2018.

McGaw, Barry, *Shaping their future : recommendations for reform of the Higher School Certificate / Barry McGaw*, Dept. of Training and Education Coordination N.S.W Sydney, 1997.

McMillan, Robert, "Competition, incentives, and public school productivity," *Journal of Public Economics*, 2004, *88* (9-10), 1871–1892.

Meyer, Bruce D and James X Sullivan, "Measuring the well-being of the poor using income and consumption," Technical Report, National Bureau of Economic Research 2003.

Ministry of Planning, "Migration in Cambodia: Report of the Cambodian Rural Urban Migration Project," Technical Report August 2012.

Morgan, Stephen L, Dafna Gelbgiser, and Kim A Weeden, "Feeding the pipeline: Gender, occupational plans, and college major selection," *Social Science Research*, 2013, *42* (4), 989–1005.

Morrow, Virginia and Emma Wilson, "Parents' Perspectives on Quality of Schooling in Andhra Pradesh India," 2013.

Munshi, Kaivan and Mark Rosenzweig, "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy," *American Economic Review*, 2006, *96* (4), 1225–1252.

Muralidharan, Karthik, "Priorities for primary education policy in India's 12th five-year plan," in "India Policy Forum," Vol. 9 National Council of Applied Economic Research 2013, pp. 1–61.

_ and Michael Kremer, "Public and private schools in rural India," *Harvard University, Department of Economics, Cambridge, MA*, 2006.

_ and Venkatesh Sundararaman, "The aggregate effect of school choice: Evidence from a two-stage experiment in India," *The Quarterly Journal of Economics*, 2015, *130* (3), 1011–1066.

National Institute of Statistics, "Cambodia Demographic and Health Survey 2014," Technical Report, National Institute of Statistics, Directorate General for Health, and ICF International, Phnom Penh, Cambodia, and Rockville, Maryland, USA 2015.

National Sample Survey Office, "Education in India," Technical Report, Ministry of Statistics and Programme Implementation 2014.

Niederle, Muriel and Lise Vesterlund, "Explaining the Gender Gap in Math Test Scores: The Role of Competition," *Journal of Economic Perspectives*, June 2010, *24* (2), 129–44.

Pedro, F, G Leroux, and M Watanabe, "The Privatization of Education in Developing Countries: Evidence and Policy Implications.[Online] United Nations Educational, Scientific and Cultural Organization, Education Sector," 2015.

Pope, Devin G. and Justin R. Sydnor, "Geographic Variation in the Gender Differences in Test Scores," *Journal of Economic Perspectives*, June 2010, *24* (2), 95–108.

Ravallion, Martin, "A comparative perspective on poverty reduction in Brazil, China, and India," *The World Bank Research Observer*, 2011, *26* (1), 71–104.

Riegle-Crumb, Catherine and Melissa Humphries, "Exploring Bias in Math Teachers' Perceptions of Students' Ability by Gender and Race/Ethnicity," *Gender & Society*, 2012, *26* (2), 290–322.

Rodríguez-Planas, Núria and Natalia Nollenberger, "Let the girls learn! It is not only about math… it's about gender social norms," *Economics of Education Review*, 2018, *62*, 230–253.

Rolleston, Caine and Rhiannon Moore, "Young Lives School Survey, 2016-17: Value-added Analysis in India," 2018.

_ , Zoe James, and Elisabetta Aurino, "Exploring the effect of educational opportunity and inequality on learning outcomes in Ethiopia, Peru, India, and Vietnam," Technical Report, UNESCO 2013.

Rose, Elaina, "Consumption smoothing and excess female mortality in rural India," *Review of Economics and statistics*, 1999, *81* (1), 41–49.

Rose, Heather and Julian R. Betts, "The Effect of High School Courses on Earnings," *The Review of Economics and Statistics*, 2004, *86* (2), 497–513.

Rose, Pauline, Ricardo Sabates, Benjamin Alcott, and Ioana Ilie, "Overcoming inequalities within countries to achieve global convergence in learning," 2017.

Rothman, Sheldon et al., "Estimating attrition bias in the Year 9 cohorts of the Longitudinal Surveys of Australian Youth," *Australian Council for Education Research (ACER) Technical report*, 2009, (48).

Salinger, L, "Cambodia garment industry workforce assessment: Identifying skill needs and sources of supply," *The Cambodia Garment Industry Productivity Center (GIPC)-Phnom Penh: USAID/Cambodia*, 2006.

Schoellman, Todd, "Early childhood human capital and development," *American Economic Journal: Macroeconomics*, 2016, *8* (3), 145–74.

Singh, Abhijeet, "Private school effects in urban and rural India: Panel estimates at primary and secondary school ages," *Journal of Development Economics*, 2015, *113*, 16–32.

Sivasankaran, Anitha, "Work and Women's Marriage, Fertility, and Empowerment: Evidence from Textile Mill Employment in India," *Job Market Paper, Harvard University, Cambridge, MA*, 2014.

Smith, James P, "The impact of childhood health on adult labor market outcomes," *The review of economics and statistics*, 2009, *91* (3), 478–489.

Spears, Dean, "Height and cognitive achievement among Indian children," *Economics & Human Biology*, 2012, *10* (2), 210–219.

Standing, Guy et al., *Global labour flexibility: Seeking distributive justice*, Vol. 287, Macmillan London, 1999.

Steckel, Richard H, "Heights and human welfare: Recent developments and new directions," *Explorations in Economic History*, 2009, *46* (1), 1–23.

Strauss, John and Duncan Thomas, "Human resources: Empirical modeling of household and family decisions," *Handbook of development economics*, 1995, *3*, 1883–2023.

Streuli, Natalia, Uma Vennam, and Martin Woodhead, *Increasing choice or inequality? Pathways through early education in Andhra Pradesh, India*, The Bernard van Leer Foundation, 2011.

Sydney Morning Herald, "NSW to make maths mandatory for students in years 11 and 12," October 2019.

Tabarrok, Alexander, "Private education in India: A novel test of cream skimming," *Contemporary Economic Policy*, 2013, *31* (1), 1–12.

Thomas, Duncan, "Incomes, expenditures, and health outcomes: Evidence on intrahousehold resource allocation," *Intrahousehold resource allocation in developing countries*, 1997, pp. 142–64.

Tooley, James, "Challenging educational injustice:'Grassroots' privatisation in South Asia and sub-Saharan Africa," *Oxford Review of Education*, 2013, *39* (4), 446–463.

United Nations Children's Fund, "The ready-made garment sector and children in Bangladesh," Technical Report, United Nations Children's Fund 2015.

United Nations Refugee Agency, "2017 Findings on the Worst Forms of Child Labor: Cambodia," Technical Report 2017.

Universities Admissions Centre, "University Entry Requirements 2020 for Year 10 Students," 2017.

U.S. Department of Education, "Paths Through Mathematics and Science: Patterns and Relationships in High School Coursetaking," 2018.

Verhoogen, Eric A, "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

Woodhead, Martin, Melanie Frost, and Zoe James, "Does growth in private schooling contribute to Education for All? Evidence from a longitudinal, two cohort study in Andhra Pradesh, India," *International Journal of Educational Development*, 2013, *33* (1), 65–73.

World Bank, "Improving trade competitiveness: Cambodian garment exports," Policy Note, World Bank Group 2013.

_ , "Cambodia trade corridor performance assessment," Technical Report, World Bank Group 2014.

_ , "Cambodia economic update: recent economic developments and outlook," Technical Report, World Bank Group October 2018.

_ , "Cambodia - Secondary Education Improvement Project," Technical Report, World Bank Group, Washington, D.C. 2019.

_ , "Doing Business 2019: Training for Reform - Cambodia," Technical Report, World Bank Group 2019.

World Trade Organization, "World Trade Statistical Review," Technical Report, World Trade Organization 2018.

Younus, M and T Yamagata, "Dynamics of the Garment Industry in Low-Income Countries: Experience of Asia and Africa (Interim Report)," *IDE-JETRO. Japan: IDE-JETRO*, 2012.