Essays in Human Capital Development and Gender

A Thesis Submitted to Trinity College Dublin, the University of Dublin, for the Degree of Doctor of Philosophy in Economics

By

Elijah Kipkech Kipchumba

Supervised By

Selim Gulesci

Dublin, 2026

Declaration, Online Access and the General Data Protection Regulation

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university, and it is entirely my own work.

Chapter 3 of this thesis is co-authored with Catherine Porter, Danila Serra, and Munshi Sulaiman. Chapter 4 of this thesis is co-authored with Rachel Cassidy, Smita Das, Clara Delavallade and Munshi Sulaiman.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Elijah Kipkech Kipchumba
December 17, 2025

Summary

This thesis comprises three self-contained empirical essays that collectively explore the constraints to human capital development among youth in Sub-Saharan Africa. The unifying thread across the three essays is twofold. First, all three interventions target individuals during critical periods of human capital formation—from school-age children to young adults. Second, each chapter adopts a gender lens, either by explicitly evaluating gendered effects or by focusing on outcomes with precise gender dimensions. The essays are organised to reflect an intentional progression—from structural access to education, to aspirational influences on schooling decisions, and finally to the role of self-perception in influencing economic outcomes.

The first essay examines the long-term and intergenerational effects of Kenya's 2008 tuition-free secondary education reform. Exploiting a regression kink design and recent Demographic and Health Survey data, the chapter estimates the causal effects of the reform on women's schooling, fertility, household wealth, and child health outcomes. While the policy significantly increased women's education, it had no measurable impact on the health of their offspring. This null finding motivates subsequent chapters by highlighting the limits of standalone structural reforms and the potential role of complementary interventions in realising the full benefits of education investments.

The second essay examines whether behavioural nudges can influence educational attitudes and gender norms among school-age children. The study evaluates a low-cost role-model intervention in Somalia in which male and female college students visited elementary schools to share their personal educational journeys. The intervention, randomised at the school and grade levels, improved gender-equitable attitudes among students, especially when delivered by female role models. In the long term, the role model intervention increased high school enrollment and reduced early marriage. This chapter highlights the potential of psychosocial inputs, such as relatable role models, to influence gendered norms and beliefs, thereby shaping educational trajectories.

The final essay focuses on young adults who have left school and are transitioning into the labour market. It evaluates the effects of socio-emotional skills (SES) training on employment outcomes in urban Tanzania through a randomised controlled trial. Participants received training in either awareness skills (e.g., empathy, listening), management skills (e.g., self-control, negotiation), or both. While the training had limited impacts on objectively measured SES, it significantly improved self-perceptions, particularly among men. Men who overestimated their SES showed better labour market outcomes, suggesting that an inflated perception of self may act as a behavioural signal in labour markets. This chapter extends the thesis's core themes by demonstrating that not only educational attainment, but also perceived ability, can enhance youth labour outcomes.

Acknowledgements

I am deeply grateful to my supervisor, Selim Gulesci, for his unwavering support and guidance throughout my PhD journey. His thoughtful advice, patience, and encouragement have profoundly shaped both my research and my growth as a scholar. Selim believed in me even at times when I doubted myself. His support extended beyond academic matters; he always took the time to ask after my family's well-being back home. I feel truly privileged to have worked under his mentorship.

I owe immense thanks to Munshi Sulaiman, my longtime mentor and friend. This PhD journey has been eight years in the making, beginning with our first experiment in Somalia and continuing throughout our work at BRAC and Save the Children. Munshi has consistently challenged me, believed in my potential, and guided me to become a better researcher. His mentorship has been a constant anchor, and I am excited to reconnect and continue our work together now that this academic chapter has come to a close.

I am sincerely grateful to my PhD committee members, Martina Zanella and Tara Mitchell, for their invaluable feedback, particularly in the formative stages of my research. I also extend my heartfelt thanks to the faculty in the Economics Department, who fostered a supportive academic environment and were always ready to offer advice and encouragement whenever needed.

To my PhD cohort — Bhavya Shrivastava, Flavia Pellegrini, Hannah Ortega McCormack, and Nouf Abushehab — thank you for being far more than colleagues. You have been a true support system, brightening the journey both academically and socially. Our gatherings, filled with laughter and shared stories, made the long PhD days lighter and the memories richer. When life became heavy, you were always there. I could not have wished for a better cohort to share this journey with.

A special thank you to my friend Danish Us Salam, whose friendship made my transition from professional life to full-time study feel not only natural but also exciting. Our long walks around Dublin became spaces of reflection, laughter, and encouragement. And to my dear friend, flatmate, and deskmate, Talent Nesongano — thank you for making the last two years so lively. Our daily walks to and from TRiSS, along with the endless conversations and shared laughter, made even the hardest days feel lighter and softened the ache of being far from our families.

I am also deeply grateful to the wider PhD community at TRiSS — including Aishwarya Deshpande, Doireann O'Brien, Enrico Cavallotti, Giorgia Conte, Libor Melioris, Matteo Pograxha, Michael McRae, and Míde Griffin — for your thoughtful feedback, countless conversations, and the camaraderie that made TRiSS a warm and welcoming place to work.

A heartfelt thank you to Christina Lysaght, whose kindness and dedication made TRiSS feel like a second home. Your support behind the scenes meant more than I can say, and your care for all of us will remain one of my fondest memories of this time. Our TRiSS research corner could not have been such an ideal set-up without you.

My gratitude also extends to the earlier PhD cohorts — Alexis De Saint-Lager and Roland Umanan — who made the early days so joyful. Our excursions and shared adventures gave me a sense of belonging that I will always treasure. I am also thankful to Simone Arrigoni for always being approachable and patient with my many questions over the years.

Above all, I am profoundly grateful to my family. To my wife, Mercy Jemeli, my daughter, Curie Jeruto, and my son, Bardeen Kiptiegoi — your love grounded my academic endeavours. Thank you for braving my long absences with such grace and strength, and for filling our virtual catch-ups with joy, laughter, and love. To my mother, Kobilo Kipraisi, and to my late father, Johana Kipraisi, who instilled in me the values of hard work, perseverance, and the courage to dream — I hope I have carried your aspirations with grace. Dad, I hope you are proud, wherever you are. To my siblings — David, Ivene, Kevin, Leonard, Moses, Naomi, Nicholas, Richard, Scholar, Shadrack, and Shantel — your unwavering encouragement has been a quiet force behind every step of this journey. This achievement belongs to you as much as it does to me.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Africa is home to the world's youngest population. Up to a third of its population is aged 15-25, while half of the population is under the age of 18. Therefore, the region's growth and development hinge on its ability to equip its young people with the skills necessary to thrive. Subsequently, interventions aimed at enhancing human capital are a central lever in the region, particularly during adolescence and early adulthood, when salient life decisions are made. However, empirical literature on the long-term effectiveness of these investments is limited. This dissertation brings together three studies contributing to this debate by evaluating the impacts of distinct human capital interventions in three East African countries. Each study targets a different mechanism but shares a common analytical concern: to what extent do these interventions shape economic and social outcomes?

In Chapter 2, *Universal Secondary Education for Intergenerational Health? Evidence from Kenya*, I explore the intergenerational impacts of Kenya's tuition-free secondary education reform. Literature argues that subsidising secondary schooling tends to be regressive as it benefits inframarginal students from affluent backgrounds (Crawfurd and Ali, 2022). Nonetheless, free secondary education (FSE) is a popular policy lever in SSA as a means to achieve universal basic education. On the other hand, FSE can be progressive if 1) it induces additional transition to secondary school by those from poorer backgrounds and 2) it pays for itself, through private returns or otherwise, in the long term. This chapter examines the latter case. Here, if FSE creates positive externalities in the form of improved child health, then investment to subsidise secondary schooling can be justified as a means to realise the wider social benefits, including the well-being of the next generation, and thus possibly offer higher returns on investment beyond the private returns.

To identify the intergenerational effects of FSE, I take advantage of a natural experiment induced by the exogenous rollout of FSE in Kenya. The government initiated capitation grant transfers to all public secondary schools to cover tuition fees for all secondary school students enrolled from January 2008. I draw on data from the 2022 Demographic and Health Survey, collected approximately 15 years after the policy rollout, and I estimate the effects of this reform on mother-child dyads. Using a regression discontinuity design with a kink, I compare child health outcomes of children whose mothers were exposed to the FSE when they were transitioning to secondary school versus children of mothers who were too old to qualify for the policy. To explore whether additional years of schooling improve child health outcomes, I instrument the mother's educational attainment with her eligibility for the FSE.

I find that the additional years of education induced by the FSE policy had no significant impact on the child's anthropometrics. I additionally find that the additional years of education do not confer any significant advantage to mothers in utilising maternal and child health services. Examining a series of causal mechanisms linking mother's education and child health, I find that FSE neither improved the mother's labour market outcomes nor equipped them with superior human capital. These findings suggest FSE had no discernible spillover effect on the health of the next generation.

This chapter contributes to the literature on potential externalities of women's education. The chapter extends the literature by demonstrating that, in the long term, the intergenerational benefits from additional years of schooling may be attenuated in some contexts. In contexts similar to the Kenyan case, where schools driving secondary education access are at the lower end of the educational quality distribution, the findings suggest that nonmarket returns to higher levels of schooling may be attenuated as these institutions may not equip learners with sufficient learning outcomes. This unexpected result challenges assumptions that expanding access to education automatically yields broad welfare gains and highlights the need to consider the quality of educational expansion.

In Chapter 3, *The Impact of Role Models on Youths' Aspirations, Gender Attitudes, and Education in Somalia* (joint work with Catherine Porter, Danila Serra, and Munshi Sulaiman), we examine how normative constraints influence schooling decisions. By abstracting from the supply-side interventions addressed by the Kenyan FSE policy in Chapter 1, this chapter analyses whether gendered beliefs and aspiration constraints are binding in the demand for higher levels of education. The central hypothesis in this chapter is that in contexts where children, particularly girls, are socialised to marry early or have limited exposure to successful role models, the demand for higher levels of education is lower, thereby increasing the likelihood of early school dropout. We investigate whether exposure to relatable role models affects gender attitudes and influences educational and marriage decisions.

The role model intervention in Somalia involved male and female college students visiting elementary schools to share their personal education journeys. During the visit, the role models delivered unscripted talks during regular classroom sessions, which lasted 40 to 60 minutes. Each talk ended with 5 to 10 minutes of students' questions and role models' answers, facilitated by the class teacher. We collaborated with Save the Children International (SCI), a non-governmental organisation, to identify college students enrolled in local universities who had experienced similar challenges to those faced by the primary school children being visited. We then randomly assigned SCI-supported schools to be visited in person by role models or to act as control schools. Within each treatment school, we further randomly selected two grades to be visited by a male role model and two grades to be visited by a female role model.

We find that visits from female role models significantly improved gender-equal attitudes among boys and girls six months after the intervention. However, there was no effect on students' college aspirations. Two- and four-year follow-up data reveal small and imprecisely estimated effects on these attitudinal measures. Nevertheless, teacher-reported data indicate positive impacts on high school enrollment and a reduced likelihood of early marriage. These findings suggest that the role model intervention—and, in particular, exposure to female role models—not only altered gender attitudes in the short term but also led to improved educational outcomes for girls, potentially impacting their life trajectories.

This chapter contributes to the literature on the role of aspirations and gendered beliefs constraints in educational outcomes. By randomly varying exposure to male and female role models, the study demonstrates that low-cost behavioural psychosocial interventions can influence educational pursuits, particularly by reducing early school exits. Our light-touch and cost-effective intervention, costing approximately 27 USD per classroom, was delivered in a unique and challenging setting: Somalia. Subsequently, this chapter contributes to the economic literature, an area that has been under-researched..

In Chapter 4, *Returns to Socio-emotional Skills Training in the Labour Market: Evidence from Youth in Tanzania* (joint work with Rachel Cassidy, Smita Das, Clara Delavallade and Munshi Sulaiman), we examine the relative importance of socio-emotional skills (SES). These skills are increasingly in demand in the labour market, with firms often indicating difficulties in finding workers with the appropriate levels of these skills. As a result, efforts to improve these skills have become an increasing part of youth employment programs. Different types of SES can shape economic outcomes through different channels. Self-awareness fosters adaptive goal pursuit by enabling individuals to learn from setbacks and capitalise on opportunities. At the same time, management skills such as negotiation and persuasion enable better navigation of social and institutional barriers.

We identified 14 SES delineated into two domains: four awareness-related and ten management-related skills. We then developed two novel training curricula, each comprising 14 one- to two-hour sessions, totalling 27 hours of content. We next conducted a randomised control trial involving 4,728 young men and women not in full-time education, employment, or training (NEET) drawn from three peri-urban regions of Tanzania. We experimentally vary whether an individual is trained in awareness-related skills, management-related skills, or a combination of both. The weeklong SES trainings were conducted in a classroom setting at local community venues, facilitated by trained and gender-matched facilitators who were tertiary college graduates.

Three months post-intervention, treatment groups reported significant improvements in self-assessed SES competencies. We also find small and imprecise improvements in SES when assessed via behavioural measures. The effects on skills, however, were not sustained at one-year follow-up. On average, no group experienced meaningful improvements in employment or earnings, either in the short term or at the one-year follow-up. Regardless of the outcome and time horizon, we do not find any differential impacts based on training type. And therefore, combining the two trainings yields no additional gains in skills or labour market outcomes for an average participant.

Our two-way skill elicitation suggests a nuanced pattern in the returns to SES training. First, by inducing a substantial increase in self-assessed SES but with limited improvements in behaviourally-elicited SES, the SES training increases the levels of skills overestimation. Secondly, this induced skill overestimation disproportionately benefits men in the labour market. We find that young men who overestimated their socio-emotional skills at baseline were more likely to experience improved labour market outcomes. In the one-year follow-up, these men were more likely to be engaged in income-generating activities. On the other hand, young women do not accrue any significant benefit from the overconfidence induced by the training.

This chapter makes two contributions to the literature. First, it attempts to disentangle and identify which SES training yields the highest returns in the labour market and beyond. Our study reveals that SES elements are interrelated; thus, efforts to develop skills in one domain can have a positive impact on other domains. This could rationalise the focus on a few skills. Our study identifies the pure effects of specifically targeted training, unlike the preceding studies, which often isolate the marginal impact of these skills. Subsequently, we document that SES skills training alone (either in awareness, management, or both) may not be sufficient to improve the labour market outcomes of the youths, especially women. Secondly, we introduce a new consequence of upskilling programs. In addition to actual skill acquisition, these programs can lead to an upwardly biased self-perception, which may serve as a behavioural signal, a feature favourably considered in the labour market.

Chapter 6 briefly summarises the thesis's main findings and highlights the policy implications.

Chapter 2

Universal Secondary Education for Intergenerational Health? Evidence from Kenya

2.1 Introduction

Growth in early childhood has been shown to shape later-life outcomes such as health, education, and labour-market outcomes (Bhalotra and Rawlings, 2011; Currie and Vogl, 2013; Almond, Currie, and Duque, 2018). Conversely, growth faltering, such as child stunting, is costly—accounting for up to half of under-five deaths globally (Black, Victora, et al., 2013; Galasso and Wagstaff, 2019). Given the high returns to early investments (Heckman, Stixrud, and Urzua, 2006; Cunha, Heckman, and Schennach, 2010), improving early health is a development priority. In many low- and middle-income countries (LMICs), expanding maternal education is viewed as a key pathway, based on the assumption that educated mothers make more informed health decisions, invest more in nutrition, and utilise health services more effectively (World Bank, 2018).

Building on the success of free primary education (FPE), many Sub-Saharan African (SSA) countries are now prioritising universal access to secondary education. Free secondary education (FSE) is often a popular policy tool toward achieving this goal, though its cost-effectiveness remains debatable (Gruijters, Abango, and Casely-Hayford, 2024). The FSE is frequently criticised for 1) benefiting inframarginal students who would have transitioned to secondary education regardless of the policy, 2) the inability of free schools to generate the requisite learning outcomes, and 3) the oversupply of secondary education vis-à-vis available jobs (Pritchett, 2001; Duflo, 2004; Alderman and Headey, 2017). Yet, if FSE leads to improvements in child health, it may yield broader social returns that justify the investment. These intergenerational benefits however hinge on whether the additional years of schooling induced by FSE enhance maternal productivity in health production — particularly in LMICs where young children face heightened risks of illness and malnutrition .

Kenya was among the first countries in the SSA to implement free secondary education (FSE), introducing the policy in January 2008, following the introduction of free primary education in 2003. Under FSE, the government provided capitation grants of KES 10,265 (about USD 164) per student annually to public secondary schools, covering tuition but not the full cost of schooling. I utilise this nationwide, exogenous policy rollout to investigate whether expanded access to secondary education has improved the health of the next generation. Specifically, I compare child health outcomes between children whose mothers were age-eligible for FSE at the time of secondary school transition and those whose mothers were too old to benefit. Using a regression discontinuity design with a kink—reflecting the policy's gradual rather than abrupt impact on schooling — I estimate the intergenerational effects of the reform on mother–child dyads using data from the 2022 Kenya Demographic and Health Survey, collected 15 years post-reform.

I document three main findings. First, Kenya's FSE policy significantly increased women's educational attainment. On average, women eligible for the policy attained nearly two additional years of schooling compared to those born too early to benefit. This is a substantial gain, given that household fees declined by 58% in local day schools and 31% in boarding schools, suggesting high price elasticity in the demand for education. The probability of completing some secondary education rose by 13 percentage points (pp). However, gains in full secondary completion were more modest, suggesting that the policy primarily facilitated transition from primary to secondary school rather than ensuring completion of 12 years of basic education.

Second, despite these gains in women's educational attainment, I find no meaningful improvements in child health. Neither the reduced-form estimates nor those instrumented with additional years of schooling show significant effects on children's anthropometrics or malnutrition risk. These null results are consistent across subgroups defined by child gender and birth order. At the same time, the FSE reform lowered maternal fertility by about 7% fewer live births, and treated women partnered with more educated—and plausibly higher-earning—men. Because the spousal education gap remained unchanged, these shifts did not weaken women's relative bargaining position within households. In principle, however, lower fertility combined with higher household resources should relax constraints and allow greater investments in child health. Yet the data show no such improvements.

Third, the null effects extend to healthcare utilisation and infant and young child feeding practices. Here, I find that the additional years of education did not improve mothers' use of maternal or child health services. The muted effects may reflect the diminished marginal value of formal schooling in contexts where primary healthcare is free and widely available—making health worker-provided knowledge a more relevant substitute than school-acquired information (Karlsson, Neve, and Subramanian, 2019). Using the geospatial distribution of health facilities, I find limited reduced-form effects of the FSE, even in contexts where substitution effects are less likely, that is, regions with low health facility coverage. Instead, I conjecture that the muted intergenerational returns to maternal education likely stem from the FSE's impact on the quality of learning. By expanding access primarily at the lower end of the school quality distribution, the FSE predispose an even larger segment of primary school graduates to lower learning outcomes in secondary school, with limited impacts on labour market outcomes and spillovers to the next generation.

This paper primarily contributes to the literature on the externalities of women's education in LMICs, particularly its intergenerational impacts. The conventional view holds that increased maternal education improves child health, either by enhancing household income (the income channel) or by increasing the efficiency of health input use (the human capital channel) (Grossman, 1972). However, I find no robust evidence supporting the human capital channel: additional schooling did not translate into better health-specific knowledge or improved capacity to absorb health-related information. While there is some evidence of an income effect, primarily through assortative mating rather than increased female labour force participation, this too did not improve child health outcomes. Notably, FSE appears to have reinforced educational hegemony within couples, thereby consolidating men's decision-making, with attendant inefficiencies in household production of child health.

My null findings align with evidence from developed countries, where expansions of secondary education have yielded limited intergenerational health effects (e.g., Lindeboom, Llena-Nozal, and Klaauw, 2009; McCrary and Royer, 2011; Arendt, Christensen, and Hjorth-Trolle, 2021). They contrast, however, with studies in LMICs examining merit-based education expansions, which tend to find substantial intergenerational gains (e.g., Duflo, Dupas, Spelke, et al., 2024). This divergence likely stems from differences in the marginal beneficiaries. Merit-based programs attract high-performing students with greater potential for learning gains, whereas universal programs like Kenya's FSE are non-selective; thus, the marginal student may be at the lower end of the distribution. Indeed, recent evidence shows that merit-based scholarships yield larger cognitive and labour-market returns than poverty-targeted scholarships (Barrera-Osorio, Barros, and Filmer, 2024). My results also diverge from findings in Indonesia and Zimbabwe, which link school construction to improved child health (e.g., Breierova and Duflo, 2004; Hasan, Nakajima, and Rangel, 2020; Grépin and Bharadwaj, 2015). However, they are consistent with Neve and Subramanian (2018), who found no such gains in Zimbabwe's secondary school expansion.

This paper further contributes to the emerging literature on free secondary education policies. While the FSE led to significant reductions in fertility—consistent with prior research from Kenya and other LMICs (e.g., Ozier, 2018; Brudevold-Newman, 2021; Muchiri, 2021; Kazibwe and Li, 2025)—its effects on women's economic outcomes appear limited in the long term. The null impact on women's economic outcomes contrasts with short-run studies, such as Brudevold-Newman (2021), which document shifts from agriculture to skilled employment. One plausible explanation is that early gains fade over time as labour markets become saturated with FSE beneficiaries (Fujimoto, Lagakos, and Vanvuren, 2023; Duflo, Dupas, and Kremer, 2024). This paper extends this literature by showing that, in the long term, the aggregate gains from additional years of education induced by FSE may be attenuated. Additionally, the findings suggest that the limited learning outcomes from FSE do not confer discernible benefits to the next generation in terms of their early childhood health.

2.2 Theory of Change

Maternal education can influence child outcomes through various channels. A key pathway is that education may affect child outcomes via human capital and health knowledge. Here, schooling is argued to enhance literacy, numeracy, and information-processing skills, which influence fertility choices, nutrition practices, hygiene behaviours, and healthcare utilisation. This mechanism is supported by empirical research. For example, in Ghana, Duflo, Dupas, Spelke, et al. (2024) demonstrate that secondary schooling improved women's ability to protect child health, lowering child mortality and boosting cognitive development. Importantly, this channel does not depend on income increases: via allocative efficiency, mothers with stronger information-processing skills make better health input decisions (see Grossman, 1972). Under this mechanism, cohorts exposed to Kenya's FSE reform could theoretically demonstrate greater knowledge and more effective child health investments.

Education may also influence child outcomes through fertility timing and life-course trajectories. In this pathway, schooling is argued to increase the opportunity cost of early marriage and early pregnancy. Previous studies find that improving access to education delays sexual debut, reduces teenage fertility, and shifts first births to later ages (Black, Devereux, and Salvanes, 2008; Osili and Long, 2008; Ozier, 2018). Later fertility enhances maternal biological readiness and lowers health risks for both mother and child. Furthermore, delaying childbirth allows births to happen when households are more financially stable. Consequently, this pathway predicts positive intergenerational effects.

A third pathway is intrahousehold bargaining. Education can enhance women's bargaining position by increasing expected earnings or improving fallback options (Duflo, 2003; Lundberg and Pollak, 1993). More broadly, in line with household bargaining literature, a narrower spousal education gap may also influence household decision-making more in favour of women (Doss, 2013). With greater bargaining power, mothers may allocate more resources to children's nutrition, health, and early learning (Duflo, 2003). At the same time, recent evidence suggests that empowerment may lead to more gender-equal allocations rather than to direct increases in child investments (Ringdal and Sjursen, 2021). Nonetheless, this channel predicts a generally positive effect.

Education can further influence child health outcomes by increasing income. More educated women may improve access to higher-productivity jobs, which, in turn, boost household resources (Duflo, Dupas, and Kremer, 2024). A higher resource endowment implies that households can increase investment in their children's nutrition and healthcare. This is further sharpened by positive assortative matching - the tendency of educated women to partner with more educated, higher-earning men (Chiappori, Iyigun, and Weiss, 2009). Although these pathways can benefit child outcomes, they involve trade-offs in time allocation. Increased maternal labour supply may reduce the time available for breastfeeding, supervision, or clinic visits (Agüero and Marks, 2008, 2011; Del Boca, Flinn, and Wiswall, 2014). Additionally, these channels depend on labour-market returns to education—a condition that might not hold in environments where formal-sector opportunities remain limited (Duflo, 2004; Duflo, Dupas, and Kremer, 2024).

A final qualifying condition is that years of schooling matter only if they lead to genuine learning. It is widely recognised that returns to education come from skill acquisition, not merely attendance (Hanushek and Woessmann, 2008; Pritchett, 2013). Therefore, when enrolment increases rapidly, outpacing investments in teachers, infrastructure, or instructional quality, the additional year of schooling yields little human capital (Pritchett, 2001, 2013). This issue is common in Sub-

Saharan Africa, where expansion often raises pupil–teacher ratios and diminishes instructional effectiveness (Bold et al., 2017; Spaull, 2013; World Bank, 2018). Furthermore, in places such as Kenya, where schools are hierarchically organised, shifting marginal students into already resource-strapped schools likely results in smaller gains from an extra year of schooling. Under such circumstances, both the human-capital and income effects weaken, leading to reduced returns to FSE-induced schooling and limited intergenerational impacts.

2.3 The Kenyan Context

2.3.1 Early Childhood Health

Kenya has made significant progress in reducing malnutrition over the last two decades. For example, in 2022, about 18% of children under five are stunted (chronic malnutrition) compared to 36% in 2003. The prevalence of underweight children has also halved in the last two decades, to about 10% in 2022, while wasting (acute malnutrition) rates remained relatively stable, affecting 5% of children in 2022 (KNBS, 2023b). Despite the nationwide progress, substantial sub-national variation exists in child malnutrition, with arid and semi-arid regions persistently recording high levels of malnutrition. Counties such as West Pokot, Kitui, and Kilifi often have the highest rates of stunting, while Turkana, Wajir, Marsabit, and Mandera have the highest proportions of wasted children. These malnutrition hotspots are clustered in northern Kenya.

The Kenyan government has addressed these regional disparities in child malnutrition through several interventions. First is the Hunger Safety Net Programme (HSNP), a regular unconditional cash transfer programme targeting the four counties[1] with the highest burden of child wasting. Secondly, it expanded primary healthcare by integrating nutrition services, maternal care, routine immunisation, and community outreach. Despite these targeted interventions[2], these counties continue to face weaker health infrastructure and health worker shortages (Ministry of Health, 2015; Macharia et al., 2021; Okoroafor et al., 2022). Additionally, these counties have higher poverty levels[3] and poorer public infrastructure following a history of marginalisation. Furthermore, these counties

[1] HSNP coverage has been expanded to cover eight poorest and arid counties, namely: Turkana, Wajir, Mandera, Marsabit, Garissa, Tana River, Isiolo, and Samburu.

[2] There have been other universal programs targeting maternal and child healthcare. A free maternity policy was instituted in 2013, and subsequently expanded in 2017, and covers all costs related to pregnancy throughout the country (Oyugi et al., 2024).

[3] Child malnutrition is not perfectly associated with poverty levels in Kenya. Although counties with higher malnutrition rates also have higher poverty rates, not all counties with higher poverty have higher rates of child malnutrition (Eberwein et al., 2016).

additionally face recurrent droughts, food insecurity, and conflicts. Therefore, the subnational variation in child malnutrition is due to a complex mix of inadequate diets and inadequate access to healthcare.

2.3.2 Education Expansion

Kenya's basic education comprises eight years of primary education and four years of secondary education. Children begin primary schooling at age six and sit for the end-of-primary national exam at ages 13-14. The student's performance in the primary school national exam, the Kenyan Certificate of Primary Education (KCPE), determines their secondary school placement. Secondary education, on the other hand, caters to children aged 14-17. At the end of secondary school, students take a national exam, the Kenyan Certificate of Secondary Education (KCSE), which determines their transition to tertiary education, including universities and colleges. KCPE and KCSE are high-stakes exams as they determine transition to the next level. Public primary education has been tuition-free since 2003, resulting in near-universal primary school enrollment and completion. KNBS (2023a) estimated that 84% of children complete secondary education as of 2022; however, recent efforts have been geared towards attaining 100% transition from primary to secondary and subsequent universal basic education.

Public secondary schools in Kenya exist in a quadripartite hierarchy (Oketch and Somerset, 2010). The top 1% are national schools, the second tier are extra-county (or provincial) schools (4%), the third tier are county (or district) schools (13%), and the substantial base are sub-county schools (83%). Students' placement within school tiers depends on their performance in the Kenya Certificate of Primary Education (KCPE), a standardised national exam administered at the end of primary school. Subsequently, the two top-tier schools are competitive institutions that admit the best-performing students nationwide. The third tier admits the best-performing students within their host sub-national region, remaining after placement by the top-tier schools. The schools in the bottom tier are barely selective in admissions and generally admit students from their localities. Another defining feature of this school hierarchy is the fees charged. For example, in 2008, national schools charged between KES 50,000 and 60,000 (USD 750 to 950) per annum, while extra-county schools charged between KES 35,000 and 50,000 (USD 550 to 750). Before the free secondary policy, the national guideline indicated that sub-county schools would charge up to KES 11,000 (USD 185) for day scholars and KES 23,400 (USD 185) for boarders. It is also important to note that if a student performs well in primary school exit exams but cannot afford the fees charged by top-tier schools, they often have no choice but to enrol in the sub-county schools[4].

Kenya's free secondary education policy has been in effect since January 2008, providing a capitation grant to all public secondary schools. Each year, the central government transfers KES 10,265 (approximately USD 164) per student enrolled in these schools, regardless of the school's hierarchy. The government set the capitation grant to cover all the fees required of day scholars in sub-county day schools. However, within the first year of the free secondary school policy, Ohba (2009) estimated that the capitation grant reduced fees charged by sub-county day schools by only 58%, while fees charged to boarders in these sub-county schools declined by 31%. Fees charged in other school tiers also fell, albeit by a much smaller margin. The effective reduction in fees was smaller than anticipated, as schools increased other discretionary fees, such as lunch and infrastructure development project fees. Furthermore, the capitation grant did not include significant expenses such as school uniforms and boarding fees. Households were then responsible for the shortfall. Therefore, although the free secondary capitation grant made schools more affordable by some margin, the costs borne by households were still substantial, not only in higher-tier schools but also in bottom-tier schools. Subsequently, free secondary education did not result in a drastic increase in secondary school enrolment, unlike the case of Free Primary Education, but enrolment gradually increased over the years.

Kenya has progressively expanded secondary school capacity by building new schools and expanding existing schools. However, it is the sub-county schools that drove the majority of the expansion, with most of these schools set up as community efforts, primarily funded by the Constituency Development Fund (CDF[5]) established in 2003. As shown in Figure 2.1(a), most schools in 2007 were sub-county schools. The number of these sub-county schools has considerably increased over the years. For example, by 2014, CDF had increased the number of secondary schools nationwide by nearly half (Figure 2.1(b)). As of 2007, it was estimated that sub-county schools cumulatively accounted for 44%

[4]Therefore, the hierarchical nature of schools inadvertently engenders inequality. For example, Lucas and Mbiti (2012) found that higher socioeconomic households strategically enrol their children in better quality private primary schools to secure placement in good quality public secondary schools, thus out-competing children from poorer backgrounds enrolled on public primary schools.

[5]CDF is a decentralised fund with an annual budgetary allocation equivalent to 3% of the government's ordinary revenue. The fund is disbursed to each electoral constituency in the country, with 75% allocated equally and 25% based on each constituency's poverty index. Communities then decide on the projects to fund. Harris and Posner (2019) estimated that 56% of the CDF was allocated to school infrastructure. Although some higher-tier schools, such as extra-county and county schools, were allocated CDF funds for infrastructure, most of the funds directed to secondary schools prioritised sub-county schools.

of students enrolled in secondary schools nationwide. However, by 2015, these schools accounted for 60% of students enrolled in secondary schools nationwide (Nicolai, Prizzon, and Hine, 2014). On the other hand, the central government only responded to regional inequalities in the distribution of national schools by upgrading[6] two extra-county schools to national schools in every county between 2011 and 2014. Therefore, the efforts to achieve universal secondary education in Kenya predominantly address supply constraints by making secondary education more accessible, both in terms of physical accessibility and affordability, and by relaxing the qualification threshold needed to enrol in secondary school.

From a human-capital perspective, the secondary expansion driven by low-resource schools is equivalent to a negative input shock for the marginal student. This lowers the marginal product of an extra year of schooling: each additional year in an under-resourced school yields less learning than a year in a well-resourced, selective school. [7] Cross-country evidence from Sub-Saharan Africa and beyond shows that without commensurate investment in quality, expanded access to schooling often leads to stagnant or declining test scores (Bold et al., 2017; Lucas and Mbiti, 2014; Pritchett, 2013; World Bank, 2018). In this sense, the FSE reform may have shifted marginal students into the lowest-quality segment of the system, where the learning yield per year is low. Subsequently, the marginal returns to FSE-induced education – and thus the intergenerational effects on child health – may be weaker.

[6]The upgrade was in the form of a large grant of up to USD 300,000 for school infrastructure, increased priority in teacher allocation and permitting them to charge higher fees. Brudevold-Newman (2025) found that this upgrade increased students' KCSE performance by 0.2 standard deviations and their subsequent university placement by eight percentage points.

[7]Lucas and Mbiti (2014) shows that learning and later-life earnings differ along this school hierarchy, disadvantaging learners at the bottom of the quality distribution.

2.4 Empirical Strategy

2.4.1 Data

This research mainly utilises survey data from the 2022 Kenya Demographic and Health Survey (KDHS). The 2022 KDHS survey was designed to provide national and subnational statistical estimates. The survey collects data on women's health and health-seeking behaviours, infant and child health, and feeding practices, among other health outcomes. The survey also records anthropometric measurements (height, weight, and age) for all children under five, alongside their mothers' characteristics. For this study, I formed mother-child dyads. Subsequently, I extract the mother's background characteristics, including date and place of birth, educational attainment, marital status, current occupation, and fertility history (the number of children ever born). Additionally, I extracted information on women's knowledge of fertility and reproductive health. From the children's dataset, I extract the anthropometric measures to assess their nutritional status, including height-for-age, weight-for-height, and weight-for-age, alongside their ages, sex, and birth order.

The key outcome of this study is child health, as indicated by child anthropometrics. I focus on health in early childhood for three main reasons. First, extensive research has shown that health in early childhood is a key determinant of future outcomes (Bhalotra and Rawlings, 2011; Currie and Vogl, 2013; Almond, Currie, and Duque, 2018). Secondly, the accumulated health stock for a child at a given point is a function of the household's investments, which are directly associated with maternal knowledge, resource allocation, and decision-making. Thirdly, anthropometric indicators are more sensitive and continuous measures of child wellbeing, unlike mortality indicators, which reflect extreme poor health and are less prevalent. I further express these child anthropometrics in three common indicators of child malnutrition: stunting, underweight, and wasting. Stunting (height-for-age less than two standard deviations below the median) reflects linear growth faltering, or chronic malnutrition, resulting from long-term nutritional and health deficiencies. Conversely, wasting (weight-for-height less than two standard deviations below the mean) indicates acute malnutrition resulting from sudden health insults or decreased food intake. Conversely, underweight (weight-for-age less than two standard deviations below the mean) is a composite measure of wasting and stunting. Since stunting indicates long-term growth faltering and is insensitive to short-term changes in nutrition, it is often considered an accurate measure of malnutrition.

I augment the 2022 KDHS survey data with other administrative and census datasets. First, to describe the response to the FSE policy, I extract the number of candidates sitting for the KCSE from annual statistical extracts between 2007 and 2020. Second, I draw on 2007 census data from secondary schools, conducted just before the rollout of the FSE. Third, I further augment this school census data with administrative data from the Kenya National Examination Council to extract further school characteristics. To characterise the health facility coverage at the time of birth for children in this study, I draw on the geospatial inventory of all public health facilities in Kenya as of 2019, as detailed by Maina et al. (2019), and combine it with perturbed geospatial information from the KDHS 2022 clusters.

2.4.2 Identification Strategy

One must address potential biases, including omitted-variable bias and simultaneity bias, to estimate the causal effect of a mother's education on her children's health outcomes. In this study, pre-existing household endowments and regional heterogeneity in public services could induce omitted-variable bias. For example, household socioeconomic status influences a girl's likelihood of pursuing further education and shapes her children's health outcomes. Regarding simultaneity, it is plausible that a mother's education and her children's health outcomes are jointly determined. For example, if healthier girls are more likely to attain more years of education, the girls' health would then confound the effect of schooling on the offspring's wellbeing. In this study, I address both concerns using a regression discontinuity design (RDD), which leverages quasi-random variation in mothers' educational attainment induced by the free secondary education policy qualification threshold.

I identify the effects of the policy by exploiting the fact that the mother's age at the time of the policy change is plausibly exogenous to it. The Kenyan FSE policy did not induce a discrete jump in years of schooling at the cutoff; instead, it could cause changes in the rate at which education attainment increases by birth cohort[8]. I, therefore, implement an RDD with a kink design (RKD) introduced by Card et al. (2015). Since the policy is not deterministic in inducing additional years of education, I instrument the mother's years of schooling with the policy qualification threshold to identify the local average treatment effect (LATE) for compliers — those whose schooling decisions were affected by the policy.

The empirical strategy proceeds in two stages. Equation 2.1 specifies the first-stage regression-kink model, which captures how mothers' years of schooling

[8]Figure 2.A.2 shows the year-on-year trend in the number of candidates taking the national exam temporarily increased among cohorts who transitioned to secondary school under the FSE policy before regressing to the long-term trend.

vary with birth-cohort exposure to the FSE policy. Equation 2.2 presents the second-stage instrumental-variable specification, where the predicted education from the first stage is used to estimate causal effects on child health outcomes. These equations formalise the graphical intuition already discussed and provide the foundation for the reduced-form estimates that follow.

$$D_j = \gamma_0 + \gamma_1(X_j - c) + \gamma_2 \mathbb{1}(X_j \geq c)(X_j - c) + \phi' Z_j + u_j \tag{2.1}$$

$$Y_{ij} = \alpha + \beta \hat{D}_j + \theta' Z_j + \eta' W_{ij} + f(X_j) + \epsilon_{ij} \tag{2.2}$$

Where Y_{ij} is the child health outcome for a child i born to a mother j. $\hat{D}_j$ is the mother's education instrumented by the policy change at the threshold c, in this case, January 1992. $X_j - c$ indicates the mother's age in months, centred on the discontinuity cutoff, January 1992. $\mathbb{1}(X_j \geq c)$ is a dummy taking the value of 1 for mothers born in January 1992 or later and zero otherwise[9]. To estimate the effects of the policy at the mother's level, I control for the mother's covariates, Z_j, such as ethnicity and religion, to improve efficiency. Similarly, I include child-level covariates, W_{ij}, such as birth order, sex and age at the time of the survey. It is also important to note that the specification in Equation 2.1 does not include a dummy variable indicating a jump at the cut-off, which helps improve precision (Card et al., 2015). Instead, the estimand relies on the difference in the slope.

I adopt a data-driven approach to select the optimal bandwidth using the Mean Squared Error (MSE), following Calonico, Cattaneo, and Titiunik (2014), for each regression separately. I allow asymmetric bandwidths on each side of the cutoff, following Jales, Ma, and Yu (2017), to accommodate potential differences in observation density. For the headline results, the optimal bandwidth varies from 5.5 years (on the left) to 4.0 years (on the right) on mother's education while for children's health outcomes regressions it ranges from 6.5 years (on the left) to 5.0 years (on the right). I further check for the robustness of this optimal bandwidth choice by iteratively adding and trimming observations within the neighbourhood of the cut-off. For each iteration, I vary the optimal bandwidth by 10%. Furthermore, I report bias-corrected standard errors associated with the MSE-optimal bandwidth choice throughout the estimates. The coefficient of interest, β, in the second-stage regression identifies the effects of additional years of schooling induced by the free secondary education policy.

[9]This cutoff implies that girls who were at most 16 years in 2008 were deemed eligible for FSE. Although the national policy indicates that students should sit for their KCPE at ages 13-14, many students delay enrolling for primary school or repeat grades in primary. A re-analysis of the 2009 KDHS in Figure 2.A.1 shows that 70% of female candidates sitting for the grade eight exams in 2007 were aged 16 years and below.

I fit a local nonparametric regression with a quadratic polynomial. I choose the local quadratic form rather than the local linear form, as the binned averages of educational attainment along the running variable in Figure 2.A.8 suggest a quadratic form. To give more weight to data points closer to the threshold, I use a triangular kernel function that reduces the influence of observations farther from the threshold as their distance from it increases.

The reduced form estimates are equally important in this study, as they reveal the full effect of the FSE policy, not just its impact through additional years of education. This would capture the total effect of policy exposure, including any indirect channels beyond educational attainment, such as spillovers. Therefore, throughout this analysis, I report the reduced-form and instrumented effects of additional schooling generated by the free secondary education policy. The reduced form effects are estimated by substituting maternal years of education with the outcome of interest in Equation 2.1. Subsequently, the differential slope, γ_2, identifies FSE's reduced form effects as shown in Equation 2.3.

$$Y_{ij} = \gamma_0 + \gamma_1(X_j - c) + \gamma_2 \mathbb{1}(X_j \geq c)(X_j - c) + \phi' Z_j + \eta' W_{ij} + \epsilon_{ij} \tag{2.3}$$

2.4.3 Identification Assumptions

The key underlying assumption of RDD is that potential outcomes are continuous in the absence of treatment. In other words, potential outcomes should be a smooth function of the running variable at the cutoff; therefore, the only thing that causes the outcome to change abruptly is the treatment. In the context of RKD, the gradient between the running variable and potential outcomes would not have exhibited any kink at the threshold (Card et al., 2015). Therefore, in the absence of treatment manipulation at the cutoff, any difference in the gradient at the cutoff should be attributable to the free secondary policy. I argue that this assumption is plausible since the qualifying threshold was exogenously determined, and it is unlikely that grandparents would have strategically planned their daughter's month of birth in anticipation of this policy.

The RKD identifying assumption is not directly testable; however, two proxy tests—density and placebo tests—can provide some persuasion. The density tests provide suggestive evidence for or against treatment manipulation under the null hypothesis that the running variable's density is continuous at the cutoff point. In contrast, the alternative hypothesis is that a kink at the cutoff would suggest endogenous sorting against the free secondary qualifying threshold. Plots in Figures 2.A.3(a) and 2.A.3(b) suggest significant overlap in the density plots at the cutoff for mothers with children under the age of 5 and the entire women's sample, respectively. Moreover, following Cattaneo, Jansson, and Ma (2020), I cannot reject the null hypothesis of no discontinuities of the running variable at the cutoff, both for mothers of under-fives (p-value=0.3986) and the entire women sample (p-value=0.4865). Even though the formal tests show no evidence of discontinuity of the running variable at the cutoff, the plots in Figures 2.A.3(a) and 2.A.3(b) are not entirely smooth at the cutoff, which might suggest some evidence of the heaping of birth dates, rather than strategic sorting[10] around the qualifying threshold. Subsequently, I apply the donut-hole RKD approach, excluding observations close to the threshold, as part of the robustness checks.

Under the placebo tests, pretreatment characteristics should be invariant to the free secondary policy change; there should be no discontinuity at the cutoff for these background characteristics. I conduct these placebo tests by estimating Equation 2.1 on a series of mothers' characteristics. Indeed, the results in Table 2.B.2 assure no discontinuity in select background characteristics of the mothers included in this study. Subsequently, it is plausible that any omitted, slowly moving variables that would have biased the estimates should affect both groups similarly. I can, therefore, conclude that mothers on either side of the cutoff were comparable in observable and unobservable characteristics. Thus, the RKD estimates are possibly credible causal estimates of the free secondary education policy.

The classic instrumental regression assumptions, including exclusion, relevance, and monotonicity (Card et al., 2015) are needed to identify the effect of additional years of education due to FSE on child health outcomes, that is β in Equation 2.2. By design, the continuity assumption demonstrates that the exclusion restriction holds. Thus, any observed changes in outcomes should be attributed to the policy change in education, rather than to other confounding factors. Relevance directly relates to the first-stage regression and shows how much the policy-induced changes in mothers' levels of education. Lastly, regarding monotonicity, I argue that the policy shift does not induce non-compliance; the free secondary policy increases schooling for all individuals and does not discourage them from pursuing additional years of education.

[10] Endogenous sorting is unlikely to be a significant concern in this setting. For strategic sorting to occur, parents needed to have planned their children's dates of birth 16 years earlier, anticipating the policy, which is improbable. Another plausibly strategic behaviour is to alter children's grade progression so that they sit for the KCPE no earlier than 2007. Although grade repetition was common at the time of the FSE policy announcement, it would, at most, affect candidates within a one-year window. Additionally, even within this one-year window, the FSE policy was unanticipated as it was announced in August 2007 as part of the presidential campaign, approximately six months after the 2007 KCPE candidates' registration.

2.5 Results

2.5.1 Effect of FSE on Education Attainment

Free secondary education increased educational attainment for mothers at both primary and secondary levels. Results in Panel B of Table 2.1 show that at the cutoff, a mother is likely to have 1.6 years more of education (Column 1), which is composed of 1.4 years in the 8-year primary education (Column 4) and 0.4 years in a 4-year secondary education (Column 7). Although, in relative terms, the policy increased years of education by approximately 23%, the absolute terms indicate the gains were not large enough to increase the likelihood that a mother completed primary education. The fact that free secondary education induces a larger effect on primary school education attainment might seem counterintuitive. However, it is plausible that the free secondary education policy induces households to adopt forward-looking responses, thereby reducing dropouts at the primary school level. The null effects on secondary school completion, on the other hand, suggest that while the free secondary education policy increased the transition to secondary school, it was not enough to keep students throughout their four years of secondary education. It is important to note that households are responsible for non-tuition education expenses, which can be a significant cost barrier for families. These costs could be among the key reasons why only 63% of mothers in the control group who completed primary school transitioned to secondary school.

Women with children under five are a subsample (43%) of women interviewed in the 2022 Kenya Demographic and Health Survey. An immediate concern is whether the effects estimated among the sub-sample are similar to those of the entire women's sample. The graphical presentation in Figure 2.A.8 indicates that the effects realised by mothers with children under 5 years mainly reflect the impact on the entire sample of women. However, the gradients in the top panel, which covers mothers under 5 years old, suggest a slightly smaller effect size than in the bottom panel, which covers the entire women's sample. Additionally, the gradient shift in years of primary education is suggestively larger, as women (and mothers) in the control group exhibited a nearly flat gradient before the introduction of free secondary education. Still, after introducing the policy, we see a positive kink. On the other hand, years of secondary education exhibit lower effects due to an upward trajectory even before the policy and thus slightly lower changes in gradients after the introduction of free secondary education. Panel A of Table 2.1 formally re-estimates the effects of the free secondary education policy on the entire women's sample and shows that the policy significantly increased women's years of education at both primary and secondary levels. Nonetheless, both panels of Table 2.1 generally indicate that the effects realised by the entire women sample are identical to those realised by the mother's sub-sample.

2.5.2 Effect of Policy on Child Health

It is hypothesised that raising women's educational levels results in a favourable allocation of resources to children and, hence, better child health outcomes. This study, however, finds limited evidence to support this hypothesis. Specifically, the results in Table 2.2 indicate that the health outcomes of children whose mothers were exposed to free secondary education were, on average, similar to those of children whose mothers were not eligible for the program. Using both the continuous anthropometric measurements (Columns 1-3) and the malnutrition status dummies (Columns 4-6), we do not observe statistical differences among these children. Moreover, the instrumented gains in their mothers' education levels in the second panel do not significantly correlate with child health outcomes. Although the effect sizes are large for continuous anthropometric measures, they do not show a consistent direction.

Unlike contexts such as Asia, boys in Sub-Saharan Africa are generally at increased risk of poorer child health outcomes relative to girls (Thurstans et al., 2020). This male disadvantage stems from a combination of biological and social factors that place them at a disadvantage relative to girls. In line with these previous findings, boys in the study sample were at least 5-pp more likely to experience either form of malnutrition. More importantly, the results in Figures 2.2(a) and 2.2(b) indicate that additional years of education induced by the free secondary education policy neither widened nor narrowed this gender gap in children's health outcomes. Since health investment per child is often negatively correlated with household size, one might expect that children born later will be at a disadvantage. However, the results in Figure 2.2(c) and 2.2(d) indicate no significant differences in child health effects among children in lower birth order relative to those in higher birth order. We can, therefore, conclude that, uniformly, there were hardly any significant gains in child health outcomes realised from the additional years of education induced by free secondary education.

2.5.3 Robustness

2.5.3.1 Sensitivity to Bandwidth Selection

I examine the sensitivity of the estimates to the selection of data points at the end of the bandwidth. I proceed in two steps. Starting with the MSE-optimal bandwidth, I iteratively inflate it by 10% to examine how the estimated effects change as I add data points at the endpoints. Conversely, to study the sensitivity to removing data points at the endpoints, or shrinking the optimal bandwidth, I iteratively deflate the optimal bandwidth by 10%. In both steps, I fit the same local polynomial given by Equation 2.1 and subsequently plot the corresponding coefficients. The results plotted in Figures 2.A.4 and 2.A.9 show broadly consistent point estimates across various bandwidth choices, and as expected, there are reductions in standard errors as more observations are included. These results provide additional evidence that the RKD estimates in this paper are robust and not sensitive to data points at the endpoints of the optimal bandwidth.

2.5.3.2 Running Variable Heaping

Irrespective of the birth year, there is a higher tendency to indicate the month of birth as either January, June or December. These non-random heaps of the running variable may violate the continuity assumptions, thereby likely biasing the treatment effect estimates. This is a key concern in this study since the cutoff points fall between two heaping points. I explore an alternative specification - donut RKD - which drops observations within the neighbourhood of the cutoff. Although this approach provides valuable checks against potential misspecification and potentially yields unbiased treatment effect estimates for non-heaped values, Barreca, Lindo, and Waddell (2016) demonstrate that it does not eliminate the bias from non-random heaping.

I incrementally exclude observations within the neighbourhood of the cutoff, one month at a time. Specifically, I fitted the preferred specification, Equation 2.1, to the entire sample, then dropped one month on either side of the cutoff and subsequently increased the excluded sample by one month in each subsequent iteration. I plot the estimated coefficients for each iteration in Figure 2.A.5. I follow a similar approach for the reduced form effects on child health outcomes, as shown in Figure 2.A.10. The results on women's education show that excluding the FSE effects persists even after excluding observations within a year of the cutoff (Figure 2.A.5). We also observe that the null effects on child health outcomes are insensitive to observations within the neighbourhood of the cutoff (Figure 2.A.10).

2.5.3.3 Placebo Cutoffs

I also examine whether there are other outcome discontinuities at arbitrarily chosen cutoffs. Here, I introduce an artificial cutoff on either side of the qualifying threshold such that the probability of treatment mimics the actual probability of treatment (50% for women-level regressions and 55% for child-level regressions) in the actual FSE policy. I focus on regions just outside the MSE-optimal bandwidth of the actual estimates. Within this region, I introduce an artificial cutoff for each quarter. Only treated observations are included for artificial cutoffs above the actual cutoff, and only control observations are included for cutoffs below the actual cutoff to avoid contamination from the actual treatment effects. For each artificial FSE cutoff, on either side of the actual FSE cutoff, I re-estimate Equation 2.1 using the same local-polynomial methods within an optimally chosen bandwidth around the artificial cutoff and plot the respective coefficients. These artificial cutoffs reflect changes induced by placebo FSE; therefore, significant discontinuities in either the mother's level of education or her children's health outcomes would falsify the findings in this paper. The results in Figure 2.A.6 indicate the general absence of discontinuity in women's education at all artificial FSE cutoffs examined. Although the results on the reduced-form effects of the artificial FSE on child health outcomes in Figure 2.A.11 do not show a consistent pattern, we observe a majority vote in favour of the null effects, consistent with our expectation.

2.5.3.4 Potential FSE Cutoffs

The official age for sitting for KCPE, the primary exit exam at grade eight, is between 13 and 14 years. Complete adherence to this official age-specific cutoff implies that the first cohort exposed to the FSE policy while transitioning to secondary school was born between 1993 and 1994. In practice, however, only a few students sit for the KCPE within this official age cutoff due to delayed primary school enrollment, grade repetition, and temporary school withdrawals. I plot the cumulative distribution of age profiles of the 2007 KCPE candidates, the first cohort exposed to the FSE, in Figure 2.A.1(a). It shows that approximately 70% of KCPE candidates were 16 years and below, while the rest were overaged, aged 17 years and above. Examining the age profiles of the second cohort of KCPE candidates exposed to FSE, as shown in Figure 2.A.1(b), I also find that the composition did not substantially change in subsequent years. This is consistent with the 29% prevalence of overage students in Kenyan primary schools as of 2009 (KNBS, 2012).

In the main specification, I consider cohorts born before January 1992 to be ineligible for the FSE. On the other hand, I consider those born on or after January 1992 eligible for the FSE. This is a notable deviation from Brudevold-Newman (2021), who considered January 1991 as the FSE eligibility cutoff. Following Ozier (2018), I adopt a data-driven approach in identifying the optimal cutoff.

Specifically, I examine candidate FSE discontinuities, focusing on birth years between 1990 and 1994. I assume there will be a discontinuity in every quarter. I re-estimate Equation 2.1 one potential quarterly discontinuity at a time. The results are shown in Figure 2.A.7. They suggest that FSE cutoffs corresponding to cohorts born between the second half of 1991 and the first half of 1992 would yield the largest gains in years of education, with a peak at the preferred cutoff, January 1992. I also derive the child health effects of these potential FSE cutoffs. The results in Figure 2.A.12 broadly indicate that the reduced form effects of FSE on child anthropometrics are insensitive to the choice of FSE cutoffs.

2.5.4 Effect of FSE on Child Health Outcomes' Channels

It is anticipated that a mother's higher levels of education will enhance her economic opportunities, thereby improving her children's health through two channels. Firstly, a higher level of education increases the opportunity cost of childbearing, thereby lowering her overall fertility. Secondly, with more resources at her disposal, she could afford to invest more in child health. These two channels suggest an aggregate increase in average investment per child, thereby improving child health outcomes. On the other hand, if higher education levels induce a mother to trade off more time to pursue market opportunities at the cost of childcare, it could result in poorer child health outcomes. Supporting the first argument, the results in Column 2 of Table 2.3 indicate that an additional year of education resulting from the expansion of secondary education was associated with a 7% decline in the number of children a mother has given birth to. Although these mothers desired to have at least 4% fewer children, the estimated effect is not statistically significant at conventional levels. Furthermore, the changes in a mother's fertility do not extend to other demographic outcomes, such as the initiation of childbearing or the spacing of subsequent births.

We do not observe mothers' earnings; however, we can infer that a mother's engagement in specific occupations, such as the professional or services sector, correlates with higher incomes. The reduced form effects in Column 2 of Table 2.4 indicate that free secondary education did not increase the probability of a mother engaging in income-generating activities; we, however, observe a 9-percentage point increase in the likelihood of a mother's engagement in a professional occupation following exposure to the free secondary education policy. The findings in the second stage, however, indicate that the additional years of education induced by the free secondary education policy were not correlated with changes in the mother's occupation[11]. The absence of significant

[11]While we do not find a significant correlation between a mother's education and their

effects on the mother's occupation suggests limited demand and supply of the mother's labour for livelihood opportunities outside the home. In the context of this study, gender norms disproportionately place the burden of childcare on women. These gender norms, along with limited childcare services, imply that mothers face significant time constraints. Nonetheless, the results in Column 1 of Table 2.4 rule out the concerns that higher maternal education attainment could potentially worsen child health outcomes by substituting time away from childcare.

Higher women's education can improve a mother's household's material well-being through the marriage market. Indeed, results in Column 3 of Table 2.5 show that an additional year of mothers' education was associated with a similar increase in their spouse's education. This increase in spousal education resulting from the reform is best interpreted as a mechanical consequence of cohort exposure rather than as evidence of deliberate marital sorting. Specifically, since men of comparable birth cohorts were also eligible for the policy, the pool of potential partners possessed higher average schooling levels. Additionally, results in Column 5 of Table 2.5 show that spousal educational differences remained unchanged. Finally, results in Column 6 indicate that the mother's additional years of education are significantly associated with higher material well-being, as noted in the household wealth index. The gains in household wealth likely reflect the labour-market returns to men's education rather than to women's, as the earlier results showed no change in women's occupation following the FSE policy. Taken together with the results in Columns 3 and 6 of Table 2.3, the effect size realised on child health outcomes could be interpreted as an upper-bound estimate, including changes in household composition (fewer children in wealthier households). Yet, there are no statistically significant gains in child health outcomes.

Higher levels of education could enhance child health outcomes through human capital effects, independent of the income channel. Here, it is anticipated that education equips women with health-specific knowledge and improves their ability to access and absorb information. In both cases, one would expect mothers with higher education to interact more effectively with healthcare services, adopt healthier behaviours, and translate the healthcare inputs into better health outcomes for their children. This will be the case even in the context where healthcare cost is not a significant binding constraint[12]. The results in

husband's occupation, it does not necessarily indicate the absence of effects of higher education on their spouse's occupation. To test this correlation, one needs to link the husband's education levels with their occupation, which is beyond the scope of this study.

[12]Kenya has made significant progress in expanding maternal and child healthcare. It instituted

Table 2.6 show that increased years of education did not significantly increase the likelihood of seeking maternal and child healthcare. The ceiling effect is unlikely to explain these null results; we find suboptimal uptake of common maternal and child healthcare, even among those with higher education, such as post-secondary education. For example, at most 80% of mothers with post-secondary education completed the required four antenatal care visits, and a similar proportion attended postnatal care.

Even when education does not equip mothers with health-specific knowledge through the curriculum, it should enhance their ability to comprehend and absorb new information. This information absorption capacity, rather than explicit knowledge gain, has indeed been argued to be the most important mechanism through which education influences child health outcomes (Glewwe, 1999). Alternatively, the socialisation skills gained from additional years of schooling build mothers' communication confidence, especially when interacting with healthcare professionals in institutional settings, thus increasing the likelihood that they will extract better-quality services for themselves and their children (LeVine et al., 2012). These skills could be essential in the strained healthcare workforce accompanying expanded healthcare provision. It is, therefore, intuitive to expect the higher levels of education induced by the policy change to enhance the quality of health services received by mothers and their children. The results in Column 2 of Table 2.6 do not show any significant relationship between higher education attainment and quality of health services, as proxied by the count of services received during antenatal care.

Beyond utilising health services, we can infer a mother's health-specific knowledge along two dimensions directly related to women's fertility. The first relates to her ability to identify the fertile window in the menstrual cycle. The second is the awareness of a mix of modern contraceptives. It is important to note that in Kenya, human reproduction is covered in both primary and secondary curricula[13]. In both cases, it would be intuitive to expect mothers with higher levels of education to possess greater fertility knowledge. However, the results in Columns 5 and 6 of Table 2.3 do not show any significant knowledge advantage in mothers exposed to the secondary expansion policy. This is surprising, given that only 40% of mothers not eligible for the FSE policy could correctly identify

a free maternity services policy in 2013 while simultaneously abolishing user fees for all primary health facilities. The free maternity policy was expanded in 2017 to cover antenatal care, pregnancy complications, delivery services, postnatal care, emergency referrals, and care for newborns.

[13]Although the primary curriculum coverage is more elementary, it introduces topics related to adolescence and the menstrual cycle. On the other hand, the secondary education curriculum details the menstrual cycle, including hormonal changes and pregnancy, but fails to link these two concepts.

the phase of the ovulatory cycle when a woman is most fertile. Additionally, these control mothers self-reported that they were aware of only 59% of modern contraceptives available in the country.

2.5.5 Why Null Effects?

2.5.5.1 Effects on Women's Empowerment?

Higher women's bargaining power within a household is often associated with positive externalities on their children's health (Doss, 2013). The implicit assumption[14] is that placing additional resources under women's control results in an overall increase in the share of resources devoted to child health production. This implies that if a woman's bargaining position in the household deteriorates relative to her husband, one would intuitively expect lower child health outcomes. The results in Column 4 of Table 2.7 show that the additional years of education did not worsen women's decision-making power in the household. We additionally do not find significant changes in intra-household relations, including the likelihood of intimate partner violence and husband controlling behaviours. This directly corresponds to contemporaneous structural shifts in educational attainment for both men and women, and thus FSE is not likely to upset the pre-existing husband-wife hegemony. Indeed, the results in Columns 4 and 5 of Table 2.5 suggest FSE did not change the relative education and age differences between spouses; thus, no changes occurred in intrahousehold relations. Therefore, the plausible weakening of the mother's intra-household bargaining power is likely not a consistent explanation behind the muted effects of additional years of the mother's education on their children's health outcomes.

2.5.5.2 Did Primary Healthcare Expansion Mute the Education Effects?

Formal schooling is just one way mothers acquire knowledge about health and nutrition, and it may not be the most effective. The generic information learned in school may be less actionable than the context-specific, practical knowledge healthcare workers provide. Therefore, expanding primary healthcare can have positive externalities beyond direct medical benefits, ensuring that even mothers without formal education gain access to well-targeted health and nutrition guidance. Primary healthcare, particularly maternal and child healthcare, has

[14] This assumes that spouses have heterogeneous preferences. And that more resources under the mother's control would have resulted in more favourable allocation towards her children. Therefore, efforts to promote women's education are often pursued for their instrumental value in providing women with greater bargaining power relative to their husbands. Ringdal and Sjursen (2021), however, showed that increased women's bargaining power within a household results in more gender egalitarian allocation rather than increased allocation for children.

expanded significantly in Kenya. For instance, in 2013, Kenya introduced a free maternity policy, which was later expanded in 2017 to cover all pregnancy-related costs, including antenatal and postnatal care, in private and public health facilities. The increased uptake of maternal healthcare associated with this expanded policy might have incentivised healthcare workers to delegate health promotion to community health volunteers (CHVs)[15], hence narrowing the child health gap between mothers with high education and those with lower education levels.

Suppose health facility-based promotion provides more accessible and actionable knowledge. Then, we would observe better child health outcomes in regions with high concentrations of health facilities, even among mothers with little or no formal education. This suggests that interactions with health facilities may substitute for schooling as a source of maternal health knowledge. Here, it is intuitive to expect that regions with higher concentrations of public health facilities have higher health knowledge saturation, thus limiting the value of formal education in shaping child health outcomes. In contrast, in regions with lower health facility coverage, the value of health knowledge derived from formal schooling is even more critical. Subsequently, in settings where health services are limited, mothers are likely to rely on their health-specific knowledge to enhance or correct their children's health. Therefore, if health facility information substitutes for health-specific knowledge from formal schooling, it is intuitive to expect higher child health outcomes with additional years of education in contexts with limited health coverage, but muted returns in contexts with higher coverage of health facilities.

I next examine whether the reduced form effects of FSE on child health outcomes vary by health facility coverage. Using the 2019 geospatial inventory of all public health facilities in Kenya (Maina et al., 2019), I classify DHS clusters into low and high health facility coverage [16]. As expected, descriptive analysis indicates that mothers residing in clusters with higher concentrations of health facilities tended to have children with better health outcomes than those living in areas with fewer health facilities. Regarding the FSE effects, the results in both panels of Table 2.B.3 indicate that reduced form effects were muted irrespective

[15]CHVs are a cadre of health workers trained to create demand for preventive services. In Uganda, for example, Nyqvist et al. (2019) found that the CHV's delivery approach successfully made health and nutrition information more accessible to mothers with even lower education levels.

[16]This classification is based on a median split of the number of health facilities within a five-kilometre radius of each cluster centroid. Health facility placement may be endogenous to child health outcomes, as it may correlate with other factors that determine child health, such as community infrastructure. Table 2.B.2, however, suggests that DHS cluster health facility density was not discontinuous at the FSE cutoff; this alleviates these endogeneity concerns for the heterogeneity analysis.

of health facility coverage. Therefore, these findings do not support the argument that health facility-based knowledge is a substitute for health knowledge from formal schooling. This suggests that the expansion of primary healthcare did not significantly mute the effects of additional years of education.

2.5.5.3 Education Quality

I document that a consistent hierarchy of schools exists, and that the FSE policy may have reinforced it. Table 2.B.1 presents the characteristics of schools before the implementation of the FSE policy. Here, we observe a significant resourcing gap between sub-county schools and the other tiers, with the former often under-resourced and understaffed. For example, we find significant differences in key inputs to students' learning outcomes, specifically, teaching and non-teaching staff. National schools have nearly three times as many teachers as sub-county schools. Due to the shortage of teaching staff, these sub-county schools heavily rely on teachers employed through the Parent-Teacher Association. While sub-county schools often have a smaller student population than national schools, the average of ten teachers per school is barely enough to cover the 11 subjects[17] often offered by these sub-county schools. Additionally, while all national schools have boarding facilities, only 16% of sub-county schools do. This implies that the majority (84%) of sub-county schools are exclusively day schools[18]. It is important to note that FSE was rolled out during a period characterised by expansion in the number of these sub-county secondary schools[19]. For example, by 2015, these sub-county schools accounted for upto 60% of students enrolled in secondary schools nationwide compared to 44% in 2007 (Nicolai, Prizzon, and Hine, 2014).

This study cannot directly assess the impacts of FSE on learning outcomes due to a lack of suitable data[20]. However, circumstantial evidence suggests

[17] As a prerequisite for teacher registration, a secondary school teacher specialises in two teaching subjects based on combinations approved by the Teachers Service Commission (TSC). Therefore, it does not necessarily mean that ten teachers are sufficient to teach the 11 subjects on offer.

[18] These schools do not offer residential facilities; their students return home at the end of the day.

[19] The expansion of sub-county secondary schools was primarily through the Constituency Development Fund (CDF), a decentralised fund disbursed annually to all constituencies since 2003. Local communities often allocated these funds towards the establishment or expansion of local schools. Subsequently, the number of sub-county schools considerably increased over the years. For example, by 2014, CDF had increased the number of secondary schools nationwide by nearly half (Figure 2.1(b)). This expansion, alongside gradual growth in the number of students transitioning from primary to secondary school, (see Figure 2.A.2), implies that congestion externality was unlikely in the Kenyan FSE policy.

[20] Aggregate students' performance in KCSE is available before and after the introduction of FSE. However, descriptive analysis suggests that KCSE is graded on a curve. Therefore, the student's annual performance is barely informative on year-on-year changes in learning outcomes. Thus, the proportion of students achieving the university qualifying grade, as shown in Figure 2.A.2, reflects

this lower-tier expansion yielded limited learning gains. First, consistent with institutional reports (e.g., Oketch and Somerset, 2010; Glennerster et al., 2011; Usawa, 2024) results in Table 2.B.1 indicate that these lower-tier schools have consistently performed poorly in the KCSE[21]. For example, in 2008, when FSE was introduced, only 6% of students in the sub-county schools attained the grade needed to advance to the university, that is, at least C+, while 54% of students in national schools achieved this grade. Secondly, while relying on proxy indicators of learning outcomes, such as fertility knowledge and family planning, the results in Columns 5 and 6 of Table 2.3 revealed no significant difference in these knowledge domains between mothers eligible for free secondary education and those who were not. Consequently, one could argue that the expansion of bottom-tier schools plausibly resulted in poor learning outcomes and, thus, a poor accumulation of health-specific knowledge. Therefore, additional years of education with limited learning gains had limited effects on their offspring's health outcomes.

2.6 Summary and Conclusions

Free secondary education (FSE) is being increasingly adopted in Sub-Saharan Africa as countries strive to ensure universal access to secondary education for their citizens. Guaranteeing free secondary education is a debatable issue, especially in limited-resource settings. Besides costing up to five times the cost of free primary education, subsidising secondary education benefits relatively privileged households. Nonetheless, this policy could be progressive if 1) it induces enrollment from poorer households, who are price-sensitive, and 2) private returns for additional years of education remain positive (Crawfurd and Ali, 2022). This study examines non-market returns to FSE, specifically the possibility of positive intergenerational health spillovers. I study the Kenyan FSE to determine whether it enhances the next generation's health outcomes by stimulating early health investment.

Kenya introduced FSE in January 2008 through capitation grant transfers to all public secondary schools, covering a portion of tuition fees for all enrolled

the percentile distribution of candidate scores rather than the actual scores scored by a specific cohort.

[21]The poorer learning outcomes among bottom-tier schools could be due to multiple factors such as resource gaps, behavioural responses among admitted students and adverse selection (Lavy, Silva, and Weinhardt, 2012; Pop-Eleches and Urquiola, 2013). On the other hand, even when top-tier schools do not significantly value-add student test scores (Lucas and Mbiti, 2014), their selective admission at the upper tail of the distribution still increases the likelihood of deriving significant value from these schools even among their marginal students (Jackson, 2010; Pop-Eleches and Urquiola, 2013; Brudevold-Newman, 2025).

students. Using this FSE reform as a natural experiment within a regression discontinuity framework, I compare the health outcomes of children born to mothers exposed to the policy versus those born to mothers who were just too old to qualify. I find that the additional years of education induced by the FSE policy had no significant impact on the child's anthropometrics or the likelihood of experiencing malnutrition. Examining a series of causal mechanisms linking a mother's education and her children's health, I find that FSE neither improved the mother's labour market outcomes nor equipped them with superior human capital. Subsequently, I argue that the null effects on child health outcomes were likely due to the supply-side expansion of schools at the lower end of the distribution. The FSE made local sub-county schools more accessible. However, the significant resourcing gap in these schools implied that their graduates realise limited learning outcomes and accrued limited private and intergenerational returns for the additional years of education.

FSE, as implemented in Kenya, did not generate intergenerational health benefits. The findings suggest that increasing the level of schooling at the bottom of the educational distribution might not equip learners with sufficient learning outcomes, thereby attenuating the nonmarket returns to higher levels of schooling. Therefore, expanding maternal education through FSE, as implemented in Kenya, may favourably compare with alternative policy tools, such as investments in health, sanitation, and nutrition, in improving child health outcomes. Thus, if the goal is to improve child health, direct health interventions may be more cost-effective than subsidising secondary education.

The conventional prediction is that education raises a person's health knowledge, thereby increasing the efficiency of health inputs, independent of the income effect. However, suppose the intergenerational externality of FSE is convex, so that the returns to schooling are highest at the secondary and post-secondary levels. Additionally, consider that this convexity is amplified by local spillovers, such that mothers in regions with more educated neighbourhoods not only derive greater value from their higher education levels but also experience more substantial spillover effects from more educated neighbourhoods. The analysis in this paper, however, cannot speak to these dimensions, as FSE increased the probability of attending secondary school but not the likelihood of completing it.

2.7 Tables

Table 2.1: Effects of FSE on Women's Education

	(1) Years of education	(2) Some Primary	(3) Completed Primary	(4) Years Primary	(5) Some Secondary	(6) Completed Secondary	(7) Years Secondary	(8) Has Child Under Five
				Panel A: All Women				
FSE	1.879 (0.496)***	0.133 (0.034)***	0.107 (0.038)***	1.126 (0.285)***	0.216 (0.061)***	0.139 (0.056)**	0.696 (0.221)***	-0.017 (0.035)
Left BW	3.977	3.771	4.760	3.883	3.746	4.555	4.039	7.193
Right BW	3.472	3.722	4.426	3.581	3.450	3.678	3.634	4.776
Control Mean	7.906	0.834	0.620	6.162	0.374	0.285	1.318	0.433
Observations	28,549	28,549	28,549	28,549	28,549	28,549	28,549	28,549
				Panel B: Mothers of Under Five				
FSE	1.638 (0.553)***	0.175 (0.045)***	0.106 (0.051)**	1.392 (0.380)***	0.132 (0.058)**	0.056 (0.052)	0.359 (0.218)	
Left BW	4.173	3.569	4.373	3.656	4.451	5.529	5.063	
Right BW	3.832	3.242	4.189	3.232	4.105	3.950	3.957	
Control Mean	7.255	0.746	0.562	5.521	0.355	0.282	1.270	
Observations	14,049	14,049	14,049	14,049	14,049	14,049	14,049	

Note: Table 2.1 illustrates the effects of FSE on various measures of a mother's or woman's education. The dependent variables in this table are dummy variables, except for columns 1, 4, and 7. Column 1 is a continuous variable representing the total years of education a mother or woman attained. I break down these years of education into the total years of primary education (top-coded to a maximum of eight years) in Column 4 and the total years of secondary education (top-coded to a maximum of four years) in Column 4. The dummy in Column 2 indicates whether a mother or woman attended primary school, while the dummy in Column 3 shows whether she completed eight years of primary education. Similarly, Column 5 indicates whether a mother or woman attended secondary school, and the dummy in Column 6 indicates whether she completed four years of secondary education. The last column is the sample selection dummy, indicating whether a woman has a child under five years old. Thus, Panel A includes the full sample of women, regardless of whether they have a child under five. Panel B, however, is limited to mothers with at least one child under five. I fit a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable as specified in Equation 2.1 for each panel. For each RKD, I used a triangular kernel with a quadratic polynomial within optimal bandwidths, following Calonico, Cattaneo, and Titiunik, 2014, while allowing different bandwidths on either side of the cutoff, following Jales, Ma, and Yu, 2017. Each regression includes fixed effects such as her county of birth, ethnicity, and religion. The standard errors, shown in parentheses, are clustered at the county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.2: Effects of FSE on Child Health Outcomes

	(1) Height for age	(2) Weight for height	(3) Weight for age	(4) Stunted child	(5) Wasted child	(6) Underweight child
			Reduced Form			
FSE	0.102 (0.134)	-0.187 (0.129)	-0.198 (0.139)	-0.053 (0.040)	0.023 (0.029)	-0.016 (0.036)
Left BW	4.963	5.455	4.757	5.591	7.090	5.841
Right BW	4.500	3.527	2.994	4.306	3.855	4.697
			Second Stage			
Maternal education (Years)	0.129 (0.143)	-0.078 (0.119)	0.010 (0.109)	-0.029 (0.049)	0.018 (0.040)	-0.012 (0.044)
Left BW	5.717	6.498	5.580	5.335	5.970	5.727
Right BW	5.036	4.721	5.000	3.887	4.341	4.747
Control Mean	-0.831	-0.407	-0.738	0.169	0.084	0.132
Observations	16,861	16,861	16,861	16,861	16,861	16,861

Note: Table 2.2 shows the reduced form effects of FSE on child health outcomes and the effect of additional years of a mother's education on her children's health. The dependent variables in columns 1 - 3 are continuous, while those in columns 4 - 6 are binary. Height-for-age, weight-for-height, and weight-for-age are standardised anthropometric measurements relative to the WHO Child Growth Standards median for all children under 59 months. The binary variables in columns 4- 6 are derived from the anthropometrics in columns 1 - 3. A child is classified as stunted, wasted, or underweight if their height-for-age, weight-for-height, and weight-for-age z-scores, respectively, fall below negative two standard deviations. In the first panel, I fit a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable as specified in Equation 2.3. In the second panel, I fit a separate fuzzy RKD for each dependent variable as specified in Equation 2.2, using the FSE policy eligibility to instrument the mother's years of education. For each RKD, I utilise a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014, allowing different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes the mother's county of birth, ethnicity, and religion fixed effects, child-level fixed effects (including sex and birth order), and adjusts for the child's age. The standard errors, indicated in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: Effects of FSE on Mother's Fertility

	(1) Age at First birth	(2) Number Children	(3) Birth Spacing	(4) Ideal Children	(5) Fertility knowledge	(6) Contraceptive knowledge
			Reduced Form			
FSE	0.665 (0.471)	-0.597 (0.225)***	0.019 (0.064)	-0.478 (0.344)	-0.017 (0.073)	0.018 (0.021)
Left BW	4.021	3.807	4.881	5.005	5.924	4.640
Right BW	4.298	4.704	3.847	3.694	4.184	4.574
			Second Stage			
Education (Years)	0.293 (0.299)	-0.332 (0.133)**	0.013 (0.049)	-0.250 (0.241)	-0.038 (0.077)	0.012 (0.015)
Left BW	5.649	4.602	4.967	4.676	4.904	5.210
Right BW	4.207	4.023	3.842	5.245	4.531	4.450
Control Mean	21.365	4.649	0.766	5.181	0.402	0.586
Observations	14,049	14,049	10,448	7,174	7,372	14,049

Note: Table 2.3 shows the reduced form effects of FSE on mothers' fertility outcomes and the effect of additional years of education on a mother's fertility. This analysis sample is restricted to mothers with at least one child under five. The dependent variables presented in this table are continuous, except for Columns 3 and 6. Columns 1 and 2 are based on the mother's birth history, indicating the age at which she had her first live birth and the total number of live births before the survey. Column 3 is a dummy variable that indicates whether the interval between a mother's first and second births was at least two years, applicable only to mothers with a minimum of two live births. Column 4 reflects the ideal number of children a mother wishes to have. Column 5 is a dummy variable indicating whether a mother can recognise when a woman is most fertile, which occurs halfway between two menstrual periods. Column 6 is a continuous variable denoting the proportion of modern contraceptive methods a mother can name or describe spontaneously without prompting. In the first panel, I fit a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable, as specified in Equation 2.1. In the second panel, I fit a separate fuzzy RKD for each dependent variable, as specified in Equation 2.2, using FSE policy eligibility to instrument the mother's years of education. For each RKD, I utilise a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014, allowing for different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes fixed effects for the mother's county of birth, ethnicity, and religion. The standard errors, shown in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Effects of FSE on Mother's Occupation

	(1) Employed	(2) Professional	(3) Agriculture	(4) Services	(5) Domestic	(6) Manual
			Reduced Form			
FSE	0.068 (0.058)	0.092 (0.041)**	0.017 (0.033)	0.041 (0.046)	-0.034 (0.022)	-0.040 (0.029)
Left BW	4.767	5.259	5.675	6.091	5.124	6.743
Right BW	4.282	5.094	5.090	3.373	4.154	4.032
			Second Stage			
Education (Years)	0.040 (0.044)	0.038 (0.044)	0.025 (0.034)	0.032 (0.038)	-0.024 (0.017)	-0.017 (0.027)
Left BW	5.069	5.224	4.557	4.814	4.463	4.531
Right BW	4.630	3.717	4.230	3.258	5.012	5.227
Control Mean	0.551	0.190	0.166	0.110	0.039	0.092
Observations	14,049	14,049	14,049	14,049	14,049	14,049

Note: Table 2.4 shows the effects of FSE on mothers' occupations in a reduced form and the effects of additional years of education on a mother's occupation. This analysis sample is restricted to mothers with at least one child under five. The dependent variables in this table are dummy variables. Each dummy in Columns 2 - 6 indicates whether a mother was engaged in a specific occupation listed in those columns at the time of the survey. In contrast, Column 1 is an aggregate variable that indicates whether a mother was engaged in any of the occupations listed in Columns 2 - 6. In the first panel, I perform a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable, as specified in Equation 2.1. In the second panel, I conduct a separate fuzzy RKD for each dependent variable, as specified in Equation 2.2, using FSE policy eligibility to instrument the mother's years of education. For each RKD, I utilise a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014, allowing for different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes fixed effects for the mother's county of birth, ethnicity, and religion. The standard errors, shown in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: Effects of FSE on Marriage Market

	(1) Currently Married	(2) Spouse Age	(3) Spouse Education	(4) Difference Spouse age	(5) Difference Spouse education	(6) Household Wealth
			Reduced Form			
FSE	0.046 (0.053)	-1.526 (0.941)	1.228 (0.571)**	-1.688 (0.945)*	-0.096 (0.453)	0.247 (0.110)**
Left BW	4.049	4.864	5.121	4.781	5.107	4.678
Right BW	3.305	4.469	4.456	4.458	4.643	4.927
			Second Stage			
Education (Years)	0.013 (0.037)	-1.394 (0.885)	0.985 (0.465)**	-1.423 (0.964)	-0.015 (0.465)	0.174 (0.072)**
Left BW	4.615	5.443	5.241	5.529	5.241	5.194
Right BW	3.302	4.694	4.285	4.558	4.285	5.161
Control Mean	0.859	42.366	7.946	6.381	0.755	-0.125
Observations	14,049	11,252	11,252	11,252	11,252	11,252

Note: Table 2.5 displays the effects of FSE on mothers' marriage outcomes in a reduced form and the effects of additional years of education on these outcomes. The sample analysed in this table is limited to mothers with at least one child under five. Columns 2 to 6 sample is further restricted to married or partnered mothers. The dependent variables in this table are continuous, except for Column 1. This first column is a dummy variable indicating whether a mother was married or partnered during the survey. Column 2 represents the recall value of a mother entering her first marital union. Columns 3 and 4 indicate the age and years of education the mothers' spouses attained under study. Meanwhile, Columns 5 and 6 show the differences between a mother's age and education and her partner's age and education levels, respectively. In the first panel, I perform a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable, as specified in Equation 2.1. In the second panel, I conduct a separate fuzzy RKD for each dependent variable, as specified in Equation 2.2, using FSE policy eligibility to instrument the mother's years of education. For each RKD, I utilise a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014, allowing for different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes fixed effects for the mother's county of birth, ethnicity, and religion. The standard errors, shown in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Effects of FSE on Nutrition Practices and Healthcare Utilisation

	(1) ANC 4 visits	(2) ANC quality	(3) Skilled birth	(4) Postnatal care	(5) Healthcare sick child	(6) Exclusive breastfeeding	(7) Acceptable child diet
			Reduced Form				
FSE	-0.000 (0.070)	0.093 (0.083)	0.050 (0.037)	0.068 (0.044)	-0.195 (0.170)	0.386 (0.197)*	-0.032 (0.106)
Left BW	5.134	4.815	6.095	5.529	5.025	3.908	3.887
Right BW	3.374	4.724	5.043	4.308	3.039	3.149	4.742
			Second Stage				
Maternal education (Years)	0.018 (0.067)	0.116 (0.103)	0.062 (0.076)	0.086 (0.085)	-0.213 (0.191)	-0.208 (0.260)	-0.027 (0.079)
Left BW	5.810	5.406	5.967	5.643	7.200	4.157	4.992
Right BW	4.080	3.996	5.148	4.310	3.532	4.041	3.838
Control Mean	0.609	0.176	0.780	0.629	0.317	0.294	0.245
Observations	9,364	9,364	10,327	10,327	3,154	1,613	2,842

Note: Table 2.6 illustrates the reduced-form effects of FSE on healthcare utilisation and nutrition practices. It also estimates the impact of additional years of a mother's education on these outcomes. A mother reports these outcomes for her youngest child, who is 36 months old. The dependent variables in this table are dummy variables, except for Column 2. Column 1 indicates whether the mother completed the four recommended antenatal care (ANC) visits during her pregnancy with this child. Column 2 shows the proportion of recommended services received during the ANC visits. The dummy in Column 3 pertains to whether the mother gave birth with the assistance of a health professional. Column 4 indicates whether the mother attended postnatal care after giving birth to the child. Column 5 is a dummy variable representing whether the mother sought healthcare for her sick child within a day of illness. Columns 6 and 7 relate to the feeding practices adopted by the mother for the child in question, indicating whether the mother exclusively breastfed the infant under six months (Column 6) or provided the infant with the recommended number of diets based on the infant's age. In the first panel, I fit a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable as specified in Equation 2.3. In the second panel, I fit a separate fuzzy RKD for each dependent variable as specified in Equation 2.2, using the FSE policy eligibility to instrument the mother's years of education. For each RKD, I utilise a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014, allowing different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes the mother's county of birth, ethnicity, religion fixed effects, and child-level fixed effects (including sex and birth order) and adjusts for the child's age. The standard errors, indicated in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Effects of FSE on Intrahousehold Relations

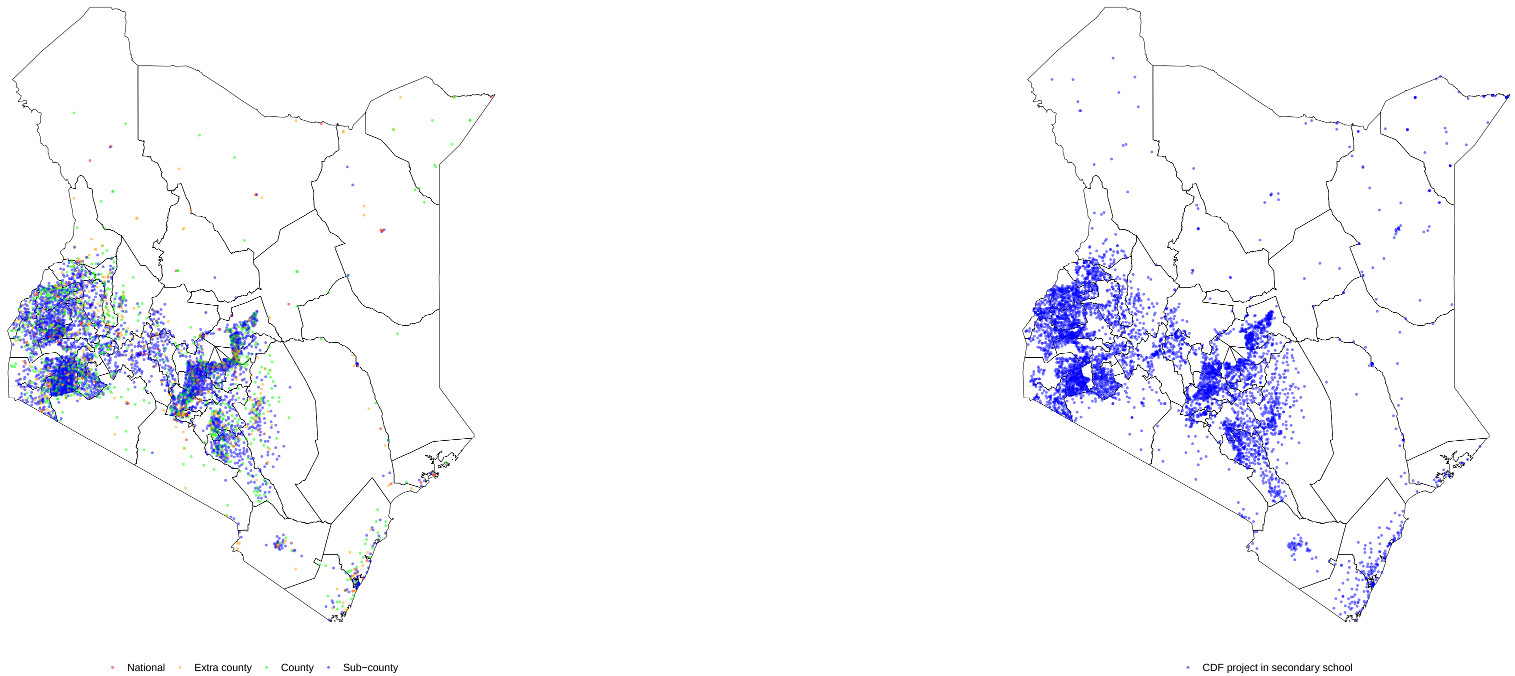
	(1) Decision Own health	(2) Decision HH purchases	(3) Decision visits	(4) Decision index	(5) Justifies Wife-beating	(6) Experienced IPV	(7) Husband has control behaviors
				Reduced Form			
FSE	0.041 (0.054)	-0.047 (0.042)	0.015 (0.054)	0.006 (0.086)	0.016 (0.052)	-0.035 (0.066)	0.018 (0.066)
Left BW	5.051	7.045	5.969	6.592	5.156	5.743	6.461
Right BW	4.665	4.550	4.033	4.273	4.860	4.583	4.610
				Second Stage			
Education (Years)	0.025 (0.048)	-0.046 (0.058)	0.021 (0.057)	0.012 (0.100)	0.006 (0.049)	-0.004 (0.056)	-0.023 (0.060)
Left BW	5.213	4.821	4.759	4.799	5.962	4.348	4.347
Right BW	5.031	4.246	4.177	4.434	4.816	4.537	4.705
Control Mean	0.833	0.764	0.810	0.521	0.398	0.395	0.229
Observations	11,252	11,252	11,252	11,252	11,252	6,987	6,987

Note: Table 2.7 shows the effects of FSE on mothers' decision-making power and their relationships with partners in reduced form. It also estimates the impact of additional years of education on these outcomes. The sample analysed in this table is limited to married or partnered mothers. The dependent variables in this table are dummy variables, except for Column 4. In Columns 1-3, mothers report whether they are involved in decisions regarding their health, major household purchases, and visits to relatives. Column 4 aggregates these decision dummies using factor analysis, with the first principal component extracted. Column 5 is a dummy variable indicating whether a mother believes that a husband is justified in beating his wife under any circumstances. Columns 6 and 7 are further restricted to mothers who also responded to the domestic violence module. These columns indicate whether a mother has ever experienced intimate partner violence and whether her partner exhibits controlling behaviour. In the first panel, I perform a separate local nonparametric regression discontinuity with a kink (RKD) for each dependent variable, as specified in Equation 2.1. In the second panel, I conduct a separate fuzzy RKD for each dependent variable, as specified in Equation 2.2, using FSE policy eligibility to instrument the mother's years of education. For each RKD, I utilise a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014, allowing for different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes fixed effects for the mother's county of birth, ethnicity, and religion. The standard errors, shown in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.8 Figures

Figure 2.1: Secondary School Distribution and Expansion

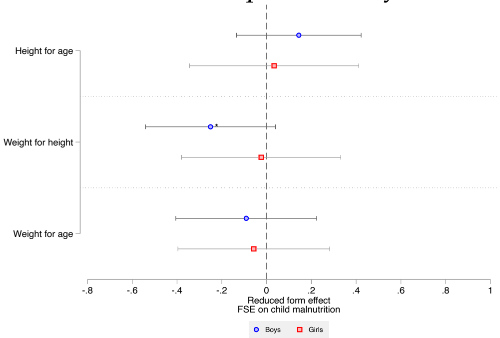
(a) Secondary schools (2007) **(b)** CDF in secondary schools (2008-2014)



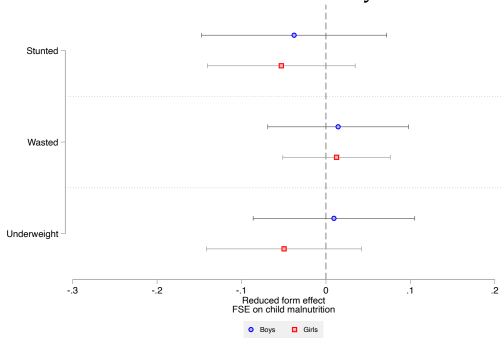
Note: Figure 2.1(a) shows the geospatial distribution of all secondary schools nationwide as of 2007. Each point corresponds to a single school. It distinguishes the four school tiers: national, extra-county, county, and sub-county. The distribution of schools is derived from school census data collected in 2007, just before the government introduced its free secondary education policy. On the other hand, Figure 2.1(b) shows the geospatial distribution of CDF projects that were indicated to be secondary school projects. I extracted the underlying data from Harris and Posner (2019), who mapped all CDF projects across the country between 2003 and 2014. The data in this panel are a subset of projects implemented in secondary schools between 2008 and 2014. Both panels in Figure 2.1 overlay county boundaries. In summary, both panels show significant supply-side expansion of secondary education nationwide, both before the FSE and after the introduction of FSE.

Figure 2.2: FSE Effects on Child Health Outcomes by Birth Order and Sex

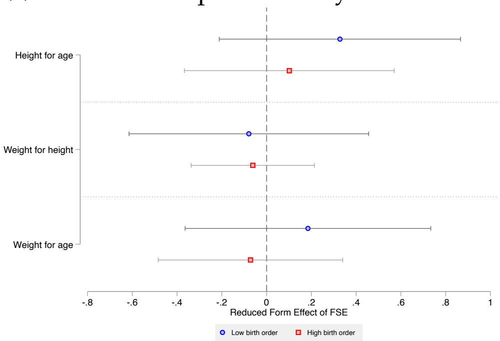
(a) Child Anthropometrics by Sex



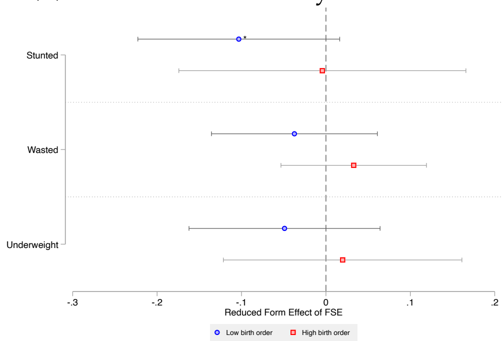
(b) Child Malnutrition by Sex



(c) Child Anthropometrics by Birth Order



(d) Child Malnutrition by Birth Order

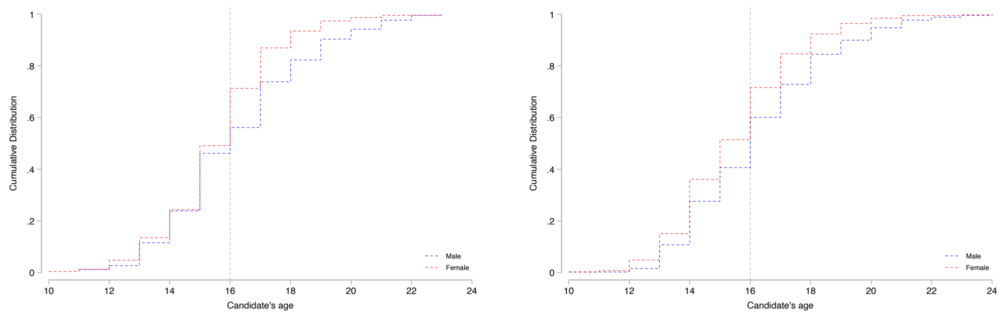


Note: These figures show the heterogeneous reduced form effects of FSE on child health outcomes. I follow the subsample approach to examine heterogeneity in the reduced-form effects by child's sex and birth order. I distinguish FSE effects by child sex in the top panels. In its reduced form, I distinguish the FSE effects among children in lower birth order (the first and the second born) and children in higher birth order in the bottom panels. I re-estimate Equation 2.1 for each subsample using various child health measures. The coefficient plots are derived from separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths, as in Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff, as in Jales, Ma, and Yu (2017). The right panels (Figure 2.2(a) and Figure 2.2(c)) relate to continuous measures of child anthropometrics, while the left panel (Figure 2.2(b) and Figure 2.2(d)) concerns the corresponding child malnutrition dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 2.A Appendix Figures

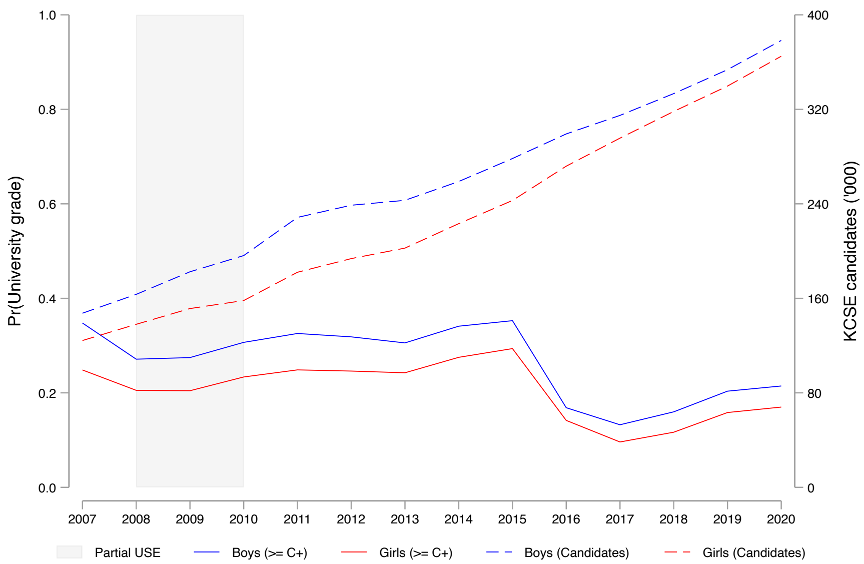
Figure 2.A.1: Age Profiles of Grade 8 Candidates

(a) Grade 8 Candidates (2007) **(b)** Grade 8 Candidates (2008)



Note: Figure shows age profiles of grade 8 students for two academic years: 2007 and 2008. It indicates the proportion of candidates who were of the specified age. The 2007 candidates (Figure 2.A.1(a)) were the first cohort to be exposed to Free Secondary Education upon transitioning to secondary school, while the 2008 candidates (Figure 2.A.1(b)) were the second cohort. For the two consecutive years, the results show that approximately 70% of female candidates were aged 16 years and below, while only about a third were within the official age range of 13-14 years. The underlying data are drawn from the 2008/2009 KDHS and are restricted to household members reported to have been in grade 8 in the respective years.

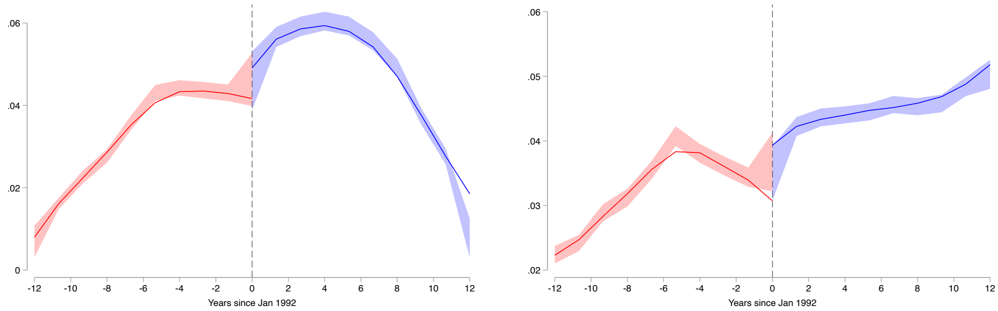
Figure 2.A.2: Trends in Secondary School Candidates and Performance in KCSE



Note: At the end of secondary education, students in the grade sit for the KCSE. This figure displays the proportion of KCSE candidates who attained the university entry grade of C+ or higher over the years, indicated by the solid line on the first Y-axis. The second Y-axis shows the absolute number of students who sat for the KCSE in a given year, indicated by a dotted line. The figure distinguishes between the trends of female and male candidates for each indicator. It also highlights the cohorts already in secondary school when the government instituted the FSE policy, and therefore, these cohorts were partially exposed to FSE. Candidates who sat for the KCSE from 2011 onwards were under the policy throughout their secondary schooling, whereas those who sat for the KCSE in 2007 or earlier were never exposed to this policy. I extracted the underlying aggregated data from the Kenya Statistical Abstracts.

Figure 2.A.3: Treatment Manipulation

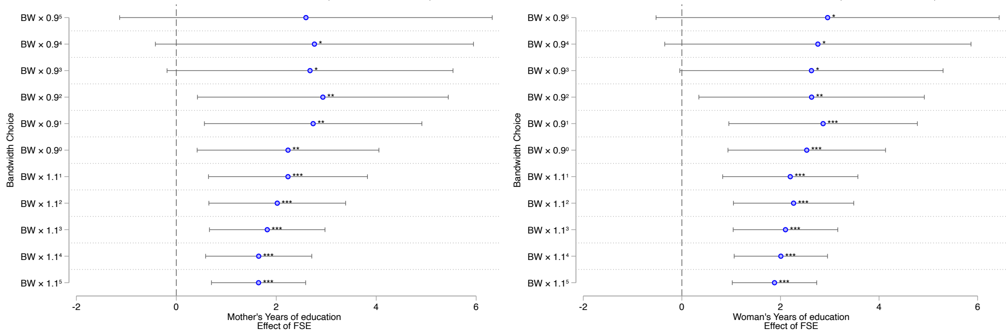
(a) Mothers of Under 5 **(b)** All Women



Note: These figures show the distribution of the running variable, centred around the FSE qualifying threshold, that is, January 1992. The x-axis shows the distance from the cutoff in years. The underlying running variable was the number of months since January 1992, subsequently expressed in years. I allow for discontinuity of the distribution of the running variable at the cutoff, between treated and control states, and test if this discontinuity is statistically significant within an MSE-optimal bandwidth following Cattaneo, Jansson, and Ma (2020). The first panel (Figure 2.A.3(a)) shows the distribution of the running variable among mothers of under-fives. On the other hand, the second panel (Figure 2.A.3(b)) shows the distribution of the running variable for the full sample of women, regardless of whether they had a child under 5 years old.

Figure 2.A.4: Bandwidth Sensitivity: Years of Education

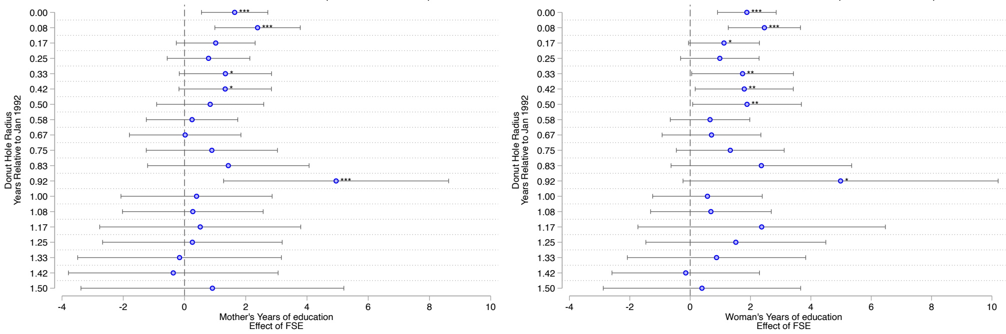
(a) Years of education (Mothers) **(b)** Years of education (Women)



Note: These figures show how bandwidth modification changes the FSE effects on the mother's education. Starting with the MSE-optimal bandwidth (BW), I either narrow or widen it. For every subsequent iteration, I modify the bandwidth by a tenth and re-estimate Equation 2.1 for various measures of the mother's education levels with a separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within a modified bandwidth. The initial MSE-optimal bandwidth was selected as in Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff as in Jales, Ma, and Yu (2017). The first panel (Figure 2.A.4(a)) shows changes in mothers' total years of education due to the FSE policy across different bandwidth choices. On the other hand, the second panel (Figure 2.A.4(b)) shows the corresponding changes in total years of education for the full sample of women, unconditional on having a child under five years old. The coefficient plots with the unmodified MSE-optimal bandwidth correspond to the preferred specifications in this study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.A.5: Donut RKD: Effect of FSE on Attained Years of Education

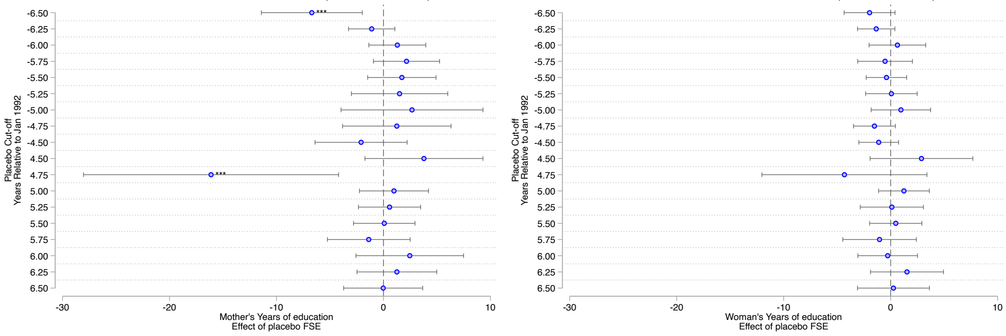
(a) Years of education (Mothers) **(b)** Years of education (Women)



Note: These figures show how estimated FSE effects on mothers' education are sensitive to changes in the neighbourhood of the qualifying threshold. Starting with the full sample of mothers, I progressively exclude mothers born closest to the January 1992 threshold on either side of the cutoff, one month at a time. For every subsequent iteration, I re-estimate the donut of Equation 2.1 with a separate local nonparametric regression-discontinuity with a kink (RKD). For each donut RKD, I utilised a triangular kernel with a quadratic polynomial within an MSE-optimal bandwidth for the sub-sample. For each donut RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths of the sub-sample following Calonico, Cattaneo, and Titiunik (2014) while allowing different bandwidths on either side of the cutoff following Jales, Ma, and Yu (2017). The first panel (Figure 2.A.5(a)) shows changes in mothers' total years of education due to the FSE policy for a specific donut. On the other hand, the second panel (Figure 2.A.5(b)) shows the corresponding changes in total years of education for the full sample of women, unconditional on having a child under five years old. The first coefficient plot in each sub-graph corresponds to the preferred specifications in this study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.A.6: Effect of Placebo FSE on Attained Years of Education

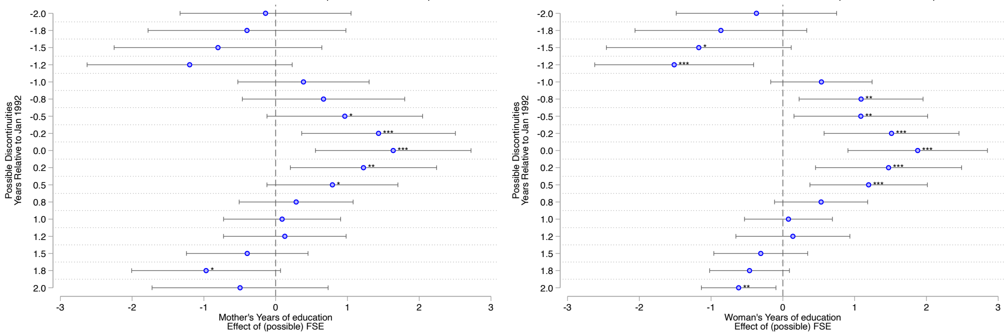
(a) Years of education (Mothers) **(b)** Years of education (Women)



Note: These figures show the effects of placebo FSE on the mother's education. The figures vary the qualifying threshold within a three-year window of the preferred cutoff. For each quarterly cutoff, normalised to January 1992 as indicated on the vertical axis, I re-estimate Equation 2.1 for various measures of mothers' education levels. The coefficient plots are derived from separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths, following Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff, following Jales, Ma, and Yu (2017). The first panel (Figure 2.A.6(a)) shows changes in mothers' total years of education due to the placebo FSE policy for a specific placebo choice. On the other hand, the second panel (Figure 2.A.6(b)) shows the corresponding changes in total years of education due to a placebo FSE for the full sample of women, unconditional on having a child under five years old. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.A.7: Effect of Potential FSE on Attained Years of Education

(a) Years of education (Mothers) **(b)** Years of education (Women)

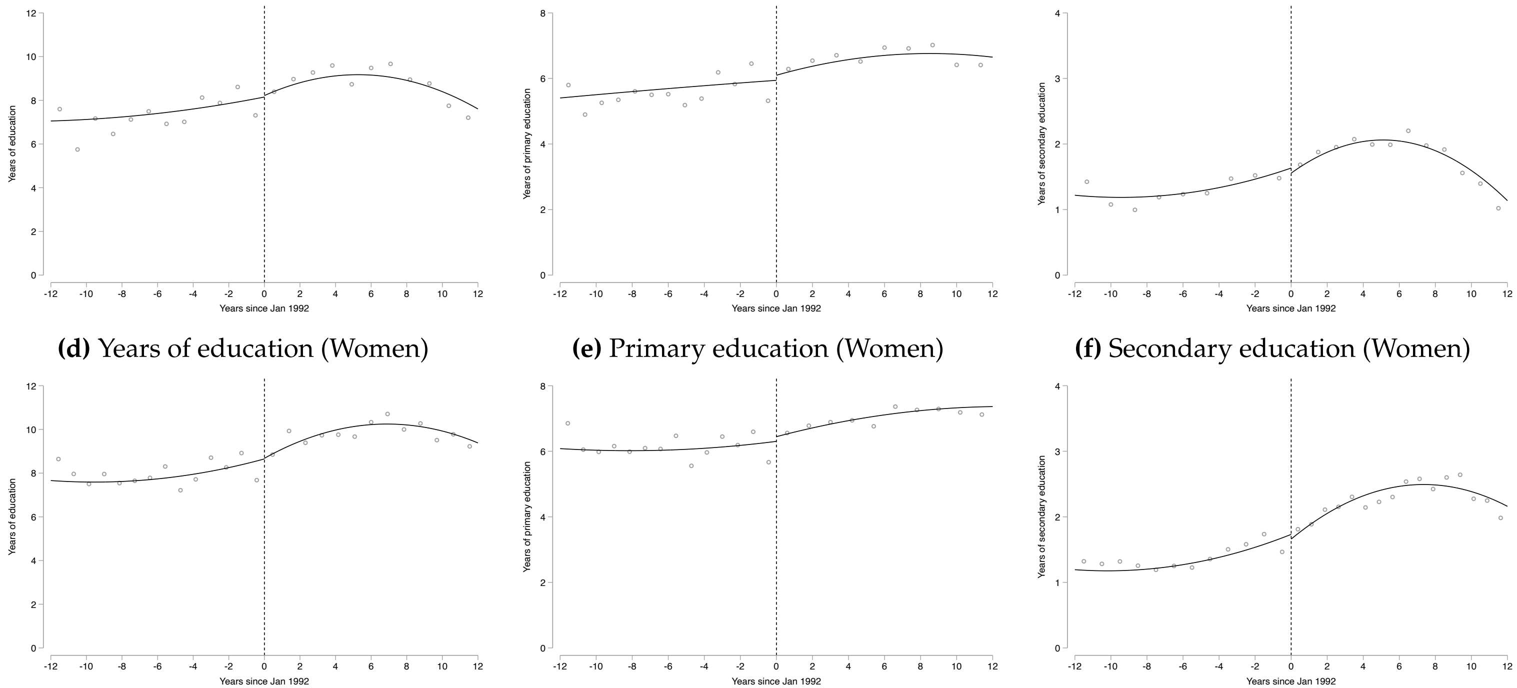


Note: These figures show the effects of potential FSE on the mother's education. The figures vary the qualifying threshold within a four-year window of the preferred cutoff. I introduce a potential cutoff for each quarter within this four-year window and re-estimate Equation 2.1 for various measures of maternal education level. Each coefficient plot corresponds to the potential FSE effect derived from a separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths, following Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff, following Jales, Ma, and Yu (2017). The first panel (Figure 2.A.7(a)) shows changes in mothers' total years of education resulting from the FSE policy for a given cutoff. On the other hand, the second panel (Figure 2.A.7(b)) shows the corresponding changes in total years of education for the full sample of women, unconditional on having a child under five years old. The coefficient plots at the cutoff of zero correspond to the preferred specifications in this study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.A.8: Effect of FSE on Years of Education Attained

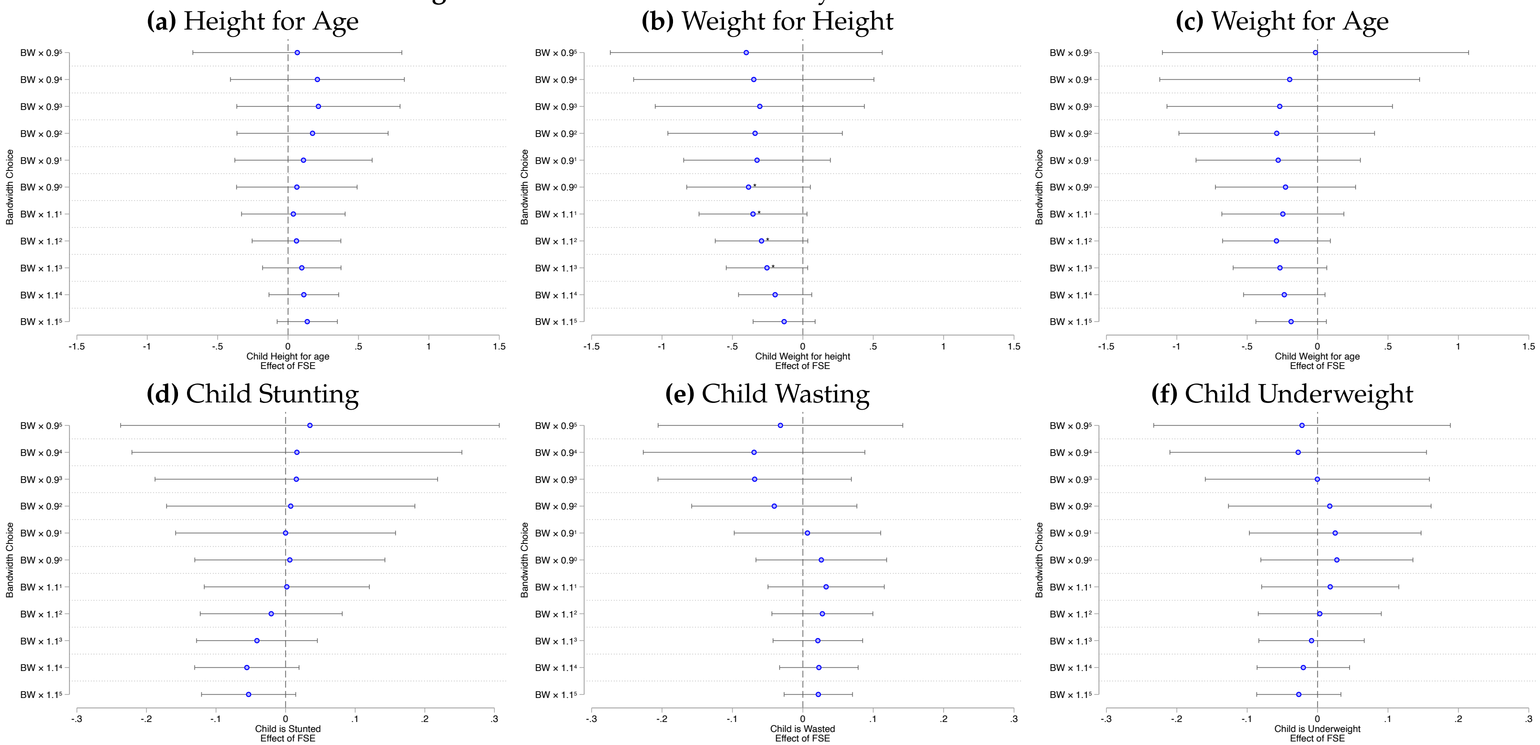
(a) Years of education (Mothers) **(b)** Primary education (Mothers) **(c)** Secondary education (Mothers)

(d) Years of education (Women) **(e)** Primary education (Women) **(f)** Secondary education (Women)



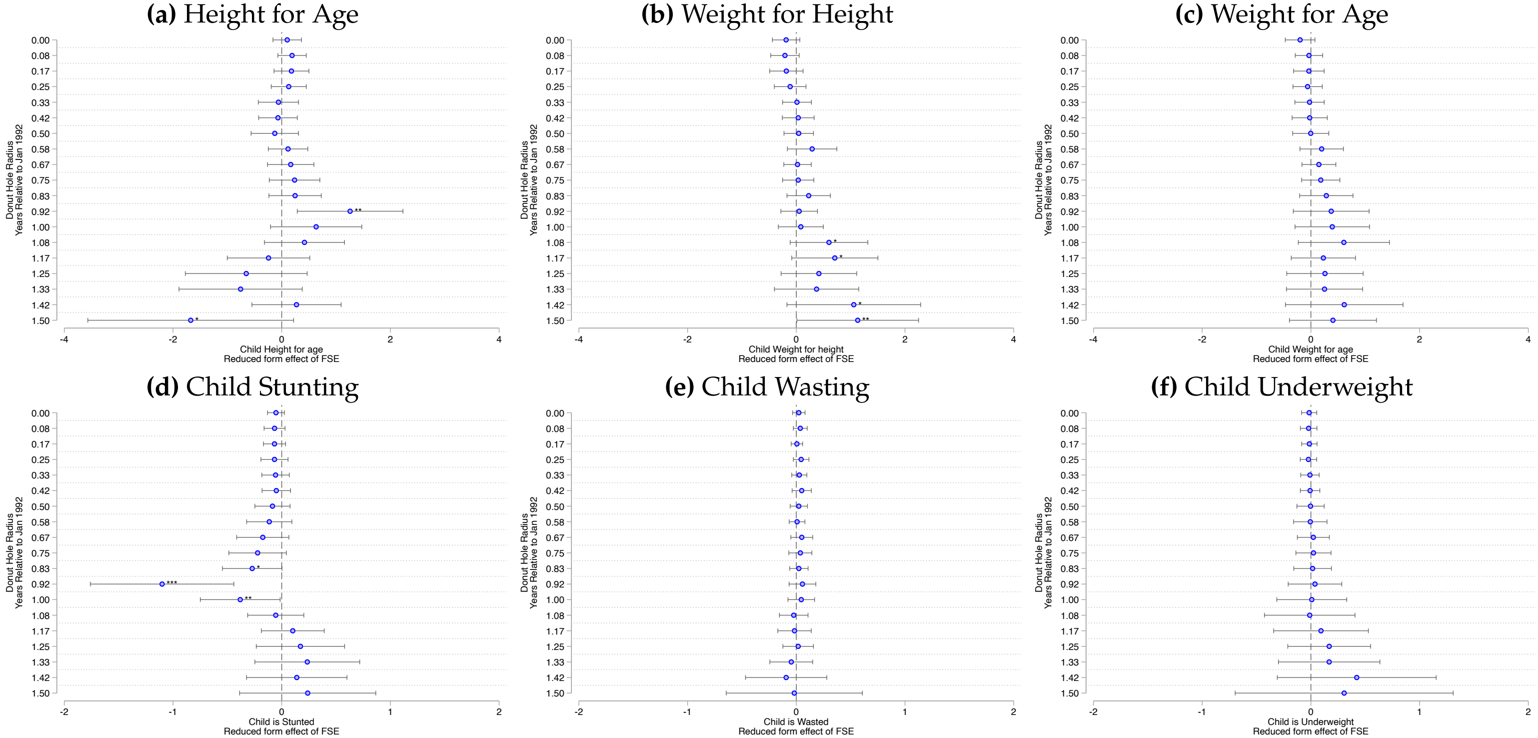
Note: Figures are regression discontinuity plots and thus show the relationship between the mother's age, relative to the qualifying threshold, and her years of education. I select the optimal bins for each plot as in Calonico, Cattaneo, and Titiunik (2015). The dotted line is the FSE cutoff: women born after December 1991. The plot segment before the cutoff represents trends in educational attainment before the policy, while the segment after the cutoff represents educational attainment for those born after the cutoff. The first panel illustrates the changes in mothers' total years of education resulting from the FSE policy for a specific cutoff. The second and third panels decompose these years of education into total years of primary education (top-coded at 8) and total years of secondary education (top-coded at 4). The coefficient plots at the zero cut-off correspond to the preferred specifications in this study. The top panels focus on the restricted sample of mothers with children under five, while the bottom panels focus on all women of reproductive age included in the 2022 KDHS sample.

Figure 2.A.9: Bandwidth Sensitivity: Child Health Outcomes



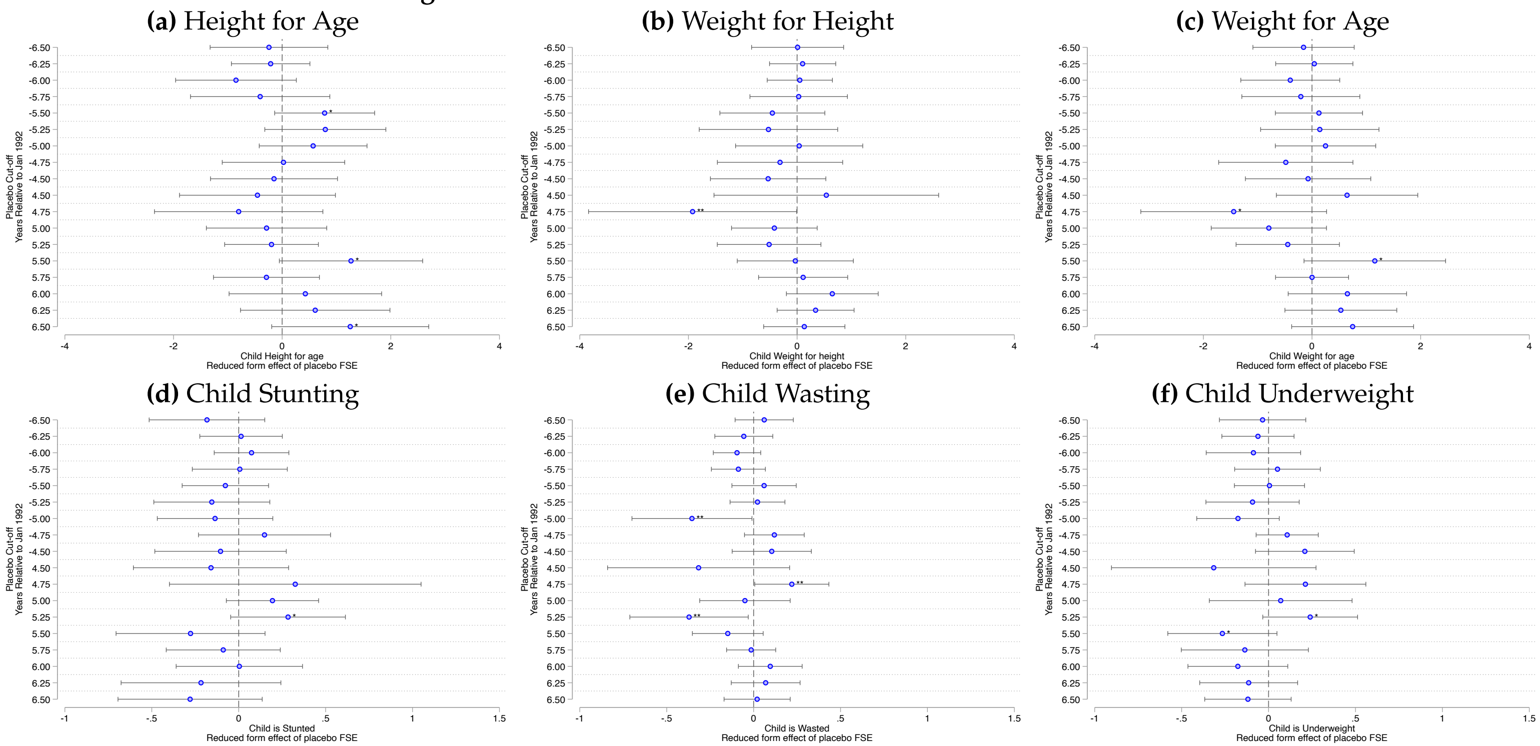
Note: These figures show how bandwidth modification changes the reduced form FSE effects on child health outcomes. Starting with the MSE-optimal bandwidth (BW), I either narrow or widen it. For every subsequent iteration, I modify the bandwidth by a tenth and re-estimate Equation 2.1 for various measures of child health outcomes with a separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within a modified bandwidth. The initial MSE-optimal bandwidth was selected as in Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff as in Jales, Ma, and Yu (2017). The top panels (Figure 2.A.9(a) - Figure 2.A.9(c)) relate to continuous measures of child anthropometrics, while the bottom panels (Figure 2.A.9(d) - Figure 2.A.9(f)) concern the corresponding child malnutrition dummies. The coefficient plots with the unmodified MSE-optimal bandwidth correspond to the preferred specifications in this study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.A.10: Donut RKD: Effect of FSE on Child Health Outcomes



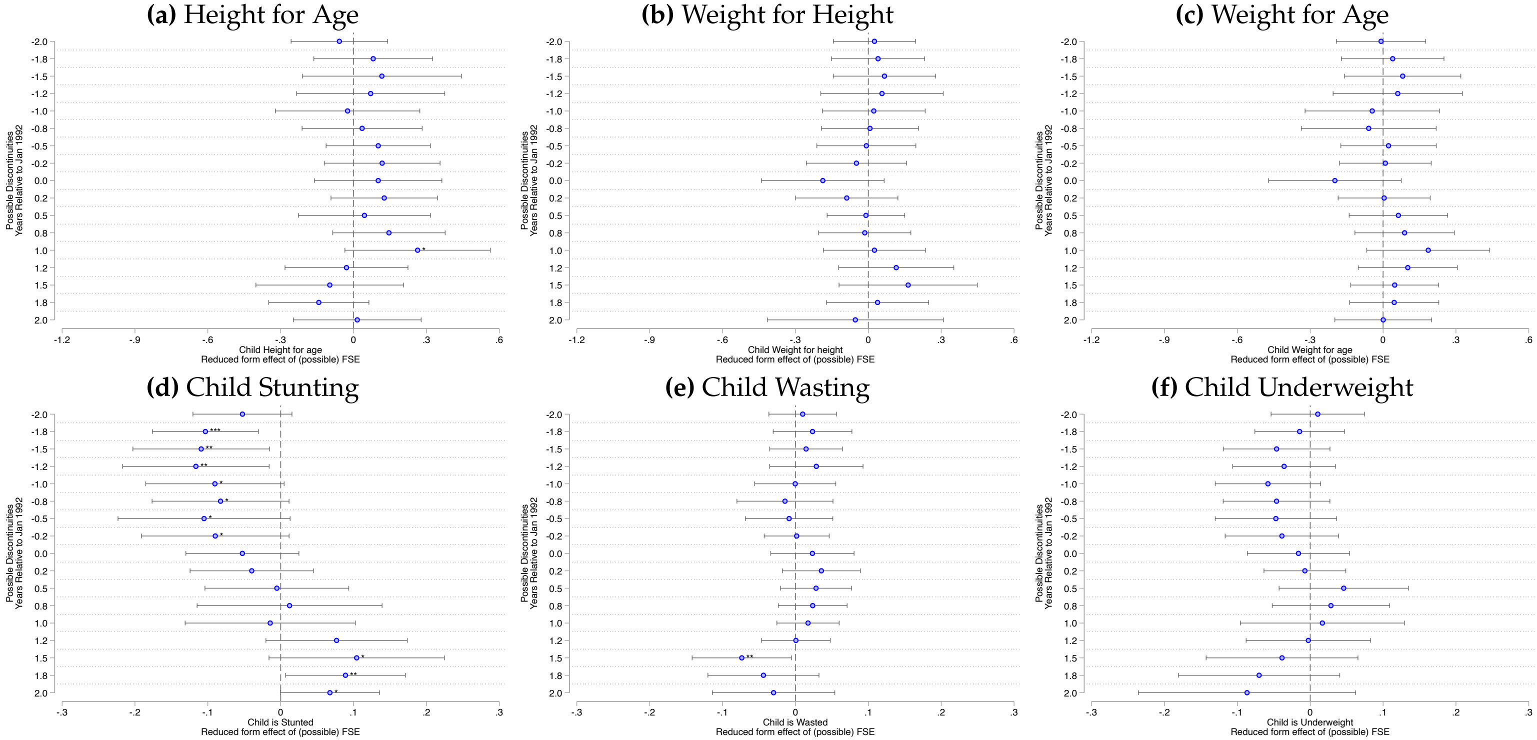
Note: These figures show how the reduced form FSE effects on child health outcomes are sensitive to changes in the neighbourhood of the qualifying threshold. Starting with the full sample of children, I progressively exclude children whose mothers were born closest to the January 1992 threshold on either side of the cutoff, one month at a time. For every subsequent iteration, I re-estimate the donut of Equation 2.1 with a separate local nonparametric regression-discontinuity with a kink (RKD). For each donut RKD, I utilised a triangular kernel with a quadratic polynomial within an MSE-optimal bandwidth for the sub-sample following Calonico, Cattaneo, and Titiunik (2014) while allowing different bandwidths on either side of the cutoff following Jales, Ma, and Yu (2017). The top panels (Figure 2.A.10(a) - Figure 2.A.10(c)) relate to continuous measures of child anthropometrics, while the bottom panels (Figure 2.A.10(d) - Figure 2.A.10(f)) concern the corresponding child malnutrition dummies. The first coefficient plot in each sub-graph corresponds to the preferred specifications in this study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.A.11: Effect of Placebo FSE on Child Health Outcomes



Note: The figures show the reduced form effects of placebo FSE on child health outcomes. The figures vary the qualifying threshold within a three-year window of the preferred cutoff. For each placebo FSE cutoff, normalised to January 1992 as indicated on the vertical axis, I re-estimate Equation 2.1 for various measures of mother's education level. The coefficient plots are derived from separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths, following Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff, following Jales, Ma, and Yu (2017). The top panels (Figure 2.A.11(a) - Figure 2.A.11(c)) relate to continuous measures of child anthropometrics, while the bottom panels (Figure 2.A.11(d) - Figure 2.A.11(f)) concern the corresponding child malnutrition dummies. The coefficient plots at the cutoff of zero correspond to the preferred specifications in this study.

Figure 2.A.12: Effect of Potential FSE on Child Health Outcomes



Note: The figures show the reduced form effects of potential FSE on child health outcomes. The figures vary the qualifying threshold within a four-year window of the preferred cutoff. For each quarter within this four-year window, I introduce a potential cutoff and re-estimate Equation 2.1 for various measures of child health outcomes. Each coefficient plot corresponds to the reduced-form effect of the potential FSE, derived from a separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths, following Calonico, Cattaneo, and Titiunik (2014), while allowing different bandwidths on either side of the cutoff, following Jales, Ma, and Yu (2017). The top panels (Figure 2.A.12(a) - Figure 2.A.12(c)) relate to continuous measures of child anthropometrics, while the bottom panels (Figure 2.A.12(d) - Figure 2.A.12(f)) concern the corresponding child malnutrition dummies. The coefficient plots at the cutoff of zero correspond to the preferred specifications in this study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 2.B Appendix Tables

Table 2.B.1: Baseline School Characteristics

	National (1)	Extra County (2)	County (3)	Sub-county (4)	Total (5)
Single-sex school (Boys)	0.41	0.33	0.10	0.02	0.08
Single-sex school (Girls)	0.44	0.39	0.23	0.03	0.13
Mixed sex school	0.15	0.28	0.67	0.95	0.79
Day school only	0.00	0.17	0.35	0.84	0.64
Boarding school only	1.00	0.67	0.30	0.03	0.19
Day and boarding school	0.00	0.16	0.36	0.12	0.17
Total teachers	33.43	23.88	13.78	9.64	12.90
Government teachers	29.45	20.17	10.96	6.45	9.69
PTA teachers	3.98	3.71	2.83	3.20	3.21
Total non-teaching staff	26.42	17.79	9.54	4.99	8.08
Total students	642.23	464.90	240.30	148.93	219.62
Teacher-student ratio	26.55	19.38	17.41	14.41	15.92
Female candidates KCSE pass rate (2006)	0.53	0.29	0.10	0.07	0.11
Female candidates KCSE pass rate (2007)	0.53	0.31	0.11	0.07	0.11
Female candidates KCSE pass rate (2008)	0.54	0.31	0.10	0.06	0.11
Female candidates KCSE pass rate (2009)	0.54	0.31	0.10	0.06	0.11
Female candidates KCSE pass rate (2010)	0.59	0.35	0.12	0.08	0.11
Observations	100	632	898	3,029	4,659

Note: Table 2.B.1 shows the mean values of school characteristics at baseline. I distinguish between the four types of schools. Each column, therefore, reports the mean values of the variable indicated in the corresponding row among the schools falling under the category specified by the column. For a nuanced characterisation of the teaching staff, I classify them into two groups: those remunerated by the government and those remunerated by the parents through their Parent-Teacher Association (PTA). I extracted the underlying data from the school census conducted in 2007, a year before the FSE policy rollout. KCSE results are extracted from the Kenya National Examinations Council, and the aggregate performance of candidates taking the national exam for each school is reported. The school-level KCSE performance data were available only for two untreated cohorts (2006 and 2007) and three partially treated cohorts (2008, 2009 and 2010). These partially treated cohorts were already in secondary school when the government instituted the FSE policy. For each cohort in each school, I calculate the proportion of female candidates achieving a university entry grade of at least C+ in KCSE.

Table 2.B.2: Covariate Balance at Cutoff

	FSE (1)	Left BW (2)	Right BW (3)	Control Mean (4)	Observations (5)
Mother height	0.005 (0.012)	4.795	3.233	1.611	7,299
Catholic	0.085 (0.043)**	4.728	3.726	0.161	14,049
Protestant	0.020 (0.040)	6.736	4.880	0.315	14,049
Muslim	-0.064 (0.040)	5.283	5.416	0.208	14,049
Other religion	-0.019 (0.033)	4.707	5.068	0.098	14,049
Rural residence	-0.049 (0.068)	4.648	3.818	0.653	14,049
Migrated	-0.035 (0.043)	6.908	4.017	0.312	14,049
Health facility within 30 minutes	0.019 (0.058)	4.376	4.196	0.236	14,049
High health facility density	-0.021 (0.051)	5.936	4.066	0.419	14,049

Note: Table 2.B.2 shows the differences in predetermined covariates at the FSE cutoff, specifically, January 1992. The dependent variable in the corresponding row is a dummy variable. I fitted a local nonparametric regression discontinuity with a kink (RKD) as specified in Equation 2.1, separately for each covariate listed in the row. For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths, following Calonico, Cattaneo, and Titiunik, 2014, while allowing different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes mother-level fixed effects, such as county of birth, ethnicity, and religion. The standard errors, presented in parentheses, are clustered at the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.B.3: Effects of USE on Child Health Outcomes by Health Facility Density

	(1) Height for age	(2) Weight for height	(3) Weight for age	(4) Stunted child	(5) Wasted child	(6) Underweight child
			Panel A: Low Health Facility Density			
FSE	0.168 (0.199)	-0.042 (0.111)	-0.022 (0.165)	-0.042 (0.055)	0.043 (0.041)	0.049 (0.059)
Left BW	4.514	6.330	4.860	4.726	6.506	6.113
Right BW	4.559	4.919	4.348	5.097	4.568	3.921
Control Mean	-0.919	-0.630	-0.938	0.193	0.110	0.169
Observations	10,173	10,173	10,173	10,173	10,173	10,173
			Panel B: High Health Facility Density			
FSE	0.094 (0.161)	-0.453 (0.251)*	-0.260 (0.250)	-0.103 (0.074)	-0.031 (0.030)	-0.065 (0.053)
Left BW	5.534	4.299	4.587	5.374	5.201	4.789
Right BW	5.318	2.715	2.783	3.399	5.056	3.442
Control Mean	-0.693	-0.056	-0.421	0.132	0.042	0.074
Observations	6,688	6,688	6,688	6,688	6,688	6,688

Note: Table 2.B.3 shows the heterogeneous reduced form effects of FSE on child health outcomes. The dependent variables in columns 1 - 3 are continuous, while those in columns 4 - 6 are binary. Height-for-age, weight-for-height, and weight-for-age are standardised anthropometric measurements relative to the WHO Child Growth Standards median for all children under 59 months. The binary variables in columns 4-6 are derived from the anthropometric data in columns 1-3. A child is classified as stunted, wasted, or underweight if their height-for-age, weight-for-height, and weight-for-age z-scores, respectively, fall below negative two standard deviations. I follow the subsample approach to examine the heterogeneity of the reduced-form effects by health facility density at the current residence. Panel A is a sub-sample of children residing in clusters with low health facility density. At the same time, Panel B is restricted to a sub-sample of children who resided in clusters with high health facility density. I re-estimate Equation 2.3 for various child health measures for each subsample (sub-panels). The coefficient plots are derived from separate local nonparametric regression-discontinuity with a kink (RKD). For each RKD, I utilised a triangular kernel with a quadratic polynomial within optimal bandwidths following Calonico, Cattaneo, and Titiunik, 2014 while allowing different bandwidths on either side of the cutoff following Jales, Ma, and Yu, 2017. Each regression includes the mother's county of birth, ethnicity, and religion fixed effects, child-level fixed effects (including sex and birth order), and adjusts for the child's age. The standard errors, indicated in parentheses, are clustered by the mother's county of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

The Impact of Role Models on Youths' Aspirations, Gender Attitudes and Education in Somalia[1]

with Catherine Porter, Danila Serra and Munshi Sulaiman

3.1 Introduction

Achieving an education in the poorest countries is extremely challenging. On the supply side, major constraints include the lack of infrastructure, such as school buildings, well-functioning classrooms, and sanitation facilities (Burde and Linden, 2013; Duflo, 2001; Kazianga et al., 2013), as well as the unavailability of qualified and motivated teachers (Beg et al., 2022; Mbiti et al., 2019; Leaver et al., 2021). At the same time, recent research also shows that demand-side factors can play an important role in explaining low enrollment and high dropout rates. This may be due to a failure to recognise the returns to education (Jensen, 2010; Nguyen, 2008), or low aspirations of parents and/or children (Dalton, Ghosal, and Mani, 2015; Genicot and Ray, 2017; Lybbert and Wydick, 2018; Ray, 2006). In addition, prevailing gender norms within traditional communities often place a lower value on girls' education, creating additional barriers for them. As a result, girls are more likely to be excluded from schooling and less able to access higher-return jobs in adulthood (Dhar, Jain, and Jayachandran, 2019; Heath and

[1]We thank Save the Children for allowing us to combine our role model intervention with their NORAD school program. The first wave of data collection received ethical clearance from the Institutional Review Board (IRB) at Southern Methodist University; the second wave from the IRB at Texas A&M University, and the third wave from the IRBs at Texas A&M University and Lancaster University. We acknowledge financial support from the Spencer Foundation through grant number 202200248. AEA RCT Registry ID: AEARCTR-0006991

Jayachandran, 2017; Jayachandran, 2021).

We investigate whether exposure to role models can significantly impact youths' gender attitudes, aspirations, and their ability to enrol in secondary school and, ultimately, pursue higher education. Following Morgenroth, Ryan, and Peters (2015), we define role models as individuals who can "influence role aspirants' achievements, motivation, and goals by acting as behavioural models, representations of the possible, and/or inspirations" (p468). We conduct a field experiment in Somalia, a very low-income and fragile context that has historically been understudied by the economics profession (Porteous, 2020). Literacy rates are among the lowest in the world; very few children are enrolled in school, and even fewer complete their education (UNICEF, 2017). For example, the net enrollment rate among secondary-school-aged girls is 15%, and among boys, 19% (World Bank, 2019). The scarcity of college-educated individuals means that Somali children rarely encounter role models—particularly women—who could inspire them to pursue higher education. Exposure to women who have graduated from high school, are enrolled in college, and aspire to non-traditional careers can reshape children's views on gender roles and equality while strengthening their self-efficacy, growth mindset, and perseverance in the face of school-related challenges.

Our study was rolled out in collaboration with Save the Children International (SCI). SCI was already implementing an intervention to address supply-side constraints to educational infrastructure and teaching in 46 primary schools (primarily grades 1 to 8) across two regions of Somalia. We randomly assigned these schools to be visited in person by current college students with backgrounds similar to those of the target youths or to serve as control schools. The role model visits took place between April and May 2018. Within each treatment school, we randomly selected two grades to be visited by a male role model and two by a female role model. When designing the study, we hypothesised that exposure to role models would influence gender attitudes and aspirations, thereby improving educational outcomes. We predicted a greater aspiration effect for female students exposed to female role models (Porter and Serra, 2020), and a stronger gender attitude effect for both male and female students exposed to female role models.

We assess the impact of the role model intervention using survey data and exam results, after conducting three distinct waves of data collection. The initial wave, or First Endline, took place about 6 months after the intervention and involved students enrolled in grades 4-7 at the time of the role model visits (we refer to these initial grades as *baseline grades*). We administered a short survey inside the schools and elicited students' desired levels of education and plans to attend high school and college, as well as a set of questions aimed at recording attitudes toward men's and women's education and labour market participation. The Second Endline data collection took place two years after the role model visits, in Spring 2020, at the onset of the COVID-19 pandemic, and to expedite fieldwork, only involved the oldest student cohort who were in grade 6 at the time of the role model visits, and who were about to graduate from 8th grade. We conducted the Third Endline data collection in the spring and summer of 2022, four years after the intervention. This involved baseline grade 4 students who were about to graduate from 8th grade at the time of data collection. Additionally, we obtained the standardised grade 8 exit exam scores for all surveyed students from the graduating classes of 2020 and 2022 (baseline grades 6 and 4), as this exam is required for elementary school graduation.

The extremely fragile environment of our study made tracking students after they left elementary school infeasible due to instability, widespread displacement, and the lack of mobile phones. To gather information on secondary school enrollment, as well as marriage and fertility outcomes, we surveyed corresponding class teachers during the Third Endline in the summer of 2022. We asked them for information on each student who graduated in 2020 and 2022. Specifically, we asked whether each sampled student was currently enrolled in high school (and the name of the high school) and whether s/he was married and/or had children.

Our impact evaluation, therefore, relies on three primary data sources. First, the survey-generated measures of education aspirations and gender attitudes. We can assess the impact of the intervention on these outcomes for students in baseline grades 4 to 7 six months after the intervention, for students in baseline grade 6 two years after the intervention, and for students in baseline grade 4 four years after the intervention. Second, we test the intervention's impact on academic performance using the exam data for grade 4 and grade 6 students. Third, we evaluate whether the intervention influenced post-primary educational choices and fertility outcomes, such as delayed marriage, using teacher-reported data, which we can interpret as "revealed" gender attitudes.

The implementation of this study was challenging for several reasons. At the outset, the absence of reliable student rosters meant that we could not precisely identify which students were present during the 2018 role model visits or track individuals across waves. Instead, we infer treatment status from students' grades at each follow-up, which limits us to estimating intent-to-treat effects, while acknowledging that students' retention and school transfers are likely to bias our estimates toward the null. Despite these challenges, our study provides rare and valuable evidence on the short- and medium-term impacts of a role model intervention in a fragile context. By combining survey-based measures of aspirations and gender attitudes with standardised exam scores and teacher-reported data on post-primary enrollment, marriage, and fertility, we can assess attitudinal and behavioural outcomes across multiple student cohorts.

Six months after the intervention, we find evidence of large and significant positive impacts of the female role models' visits on both boys and girls' attitudes toward gender equality. The effects are large (and equal for boys and girls): 0.37 standard deviations from the means observed in the control schools, and enough to close the observed initial gap in gender attitudes. When surveying baseline grade 6 and 4 students, two and four years after the intervention (respectively), we see that the impact of female role models on gender attitudes remains positive but is much smaller and no longer statistically significant. However, the number of schools in our sample declines over time because not all schools served grades 6-8, thereby reducing our statistical power.

We find no significant short- or medium-run impacts of role models on educational aspirations; virtually all students expect to enrol in secondary school, and about 70% aspire to attend college, with no significant differences by gender or treatment status. This lack of treatment effects may also reflect the difficulty of measuring children's educational aspirations. Across contexts, educational aspirations tend to be high (see, e.g., Dhar, Jain, and Jayachandran, 2022), and may capture unattainable "dreams" or youths' over-optimism rather than realistic educational goals. Aspirations that start high are harder to raise. Moreover, aspirations may evolve gradually over time (Appadurai, 2004), so meaningful changes can be observed only in the longer run. Consistent with this interpretation, our data show marginally significant positive effects of female role models on girls' aspirations four years after the intervention.

Our analysis of education outcomes for baseline grades 4 and 6 students indicates that role models, especially female role models, increase the likelihood of enrolling in secondary school among boys and girls by 6-10percentage points. Although not statistically significant, we also see a negative impact on marriage and fertility outcomes of boys and girls. These findings suggest that the role model intervention not only altered gender attitudes in the short term but also improved educational outcomes and possibly influenced their life trajectories. We are cautious about claiming external validity for our point estimates, as they are likely larger than those found elsewhere, given the especially disadvantaged target population and the lack of exposure to role models through in-person or media interactions. We also reiterate the importance of implementing role-model interventions alongside efforts to alleviate supply-side barriers to education, especially in fragile, underresourced countries such as Somalia.

This study contributes to a growing literature investigating role model effects by testing the impact of movies and TV shows on education outcomes (Di, 2024; Riley, 2024), women's empowerment and well-being (Jensen and Oster, 2009; La Ferrara, Chong, and Duryea, 2012; Lubega et al., 2021), entrepreneurial aspirations (Batista and Seither, 2019; Bjorvatn et al., 2019), and economic and educational investments of poor rural households (Bernard et al., 2023; Leight et al., 2024). In high-income contexts, Porter and Serra (2020) and subsequent replications and extensions by Patnaik et al. (2024) and Karszen, Gitter, and Groves (2025) have shown that exposing young women to successful and inspiring women through brief in-person classroom visits may impact their aspirations and educational choices.[2] Similarly to Porter and Serra (2020), our experimental design relies on the direct interaction between the target sample (primary school students) and real-life role models who come from a similar background and have experienced similar challenges in life.[3] Contrary to Porter and Serra (2020), we randomised the gender of the role model students were exposed to to assess the differential impacts of male and female role models by the gender of the target students. While attrition at the school level limits longer-term comparisons, our design provides new evidence on role-model effects on children in a fragile, low-income setting, a topic largely absent from the literature.

We also contribute to the literature on gender norms and women's labour-market participation, focusing on a highly gender-unequal setting. Prior work has highlighted the role of family and societal characteristics in shaping female labour force participation (Fernández, Fogli, and Olivetti, 2004; Olivetti, Patacchini, and Zenou, 2020), but studies on how individual attitudes toward gender equality are formed and can change remain scarce. Notable exceptions include a field experiment in Norway showing that male recruits' gender attitudes shifted when assigned to live and work with female peers (Dahl, Kotsadam, and Rooth, 2021), and a school-based intervention in India that used classroom discussions on gender roles over two years to shift adolescents' gender attitudes and behaviours, particularly among boys (Dhar, Jain, and Jayachandran, 2022). Our low-cost and easily scalable intervention, consisting of brief classroom visits by relatable college students, contributes to this evidence by demonstrating that light-touch

[2] See Serra (2025) for a review of role model interventions in low-income countries. Other related studies have examined the impact that female teachers may have on women's performance in math (Muralidharan and Sheth, 2016) and their field of study in higher education (Lim and Meer, 2019).

[3] Our work is also related to the unpublished study of Nguyen (2008), which exposed pupils in schools in Madagascar to educated individuals from the same school districts, employing a total of 72 role models.

programs can have a meaningful impact on students' gender attitudes and key life choices, such as secondary school enrollment and marriage.

More broadly, this study adds to the literature on the socio-economic impacts of interventions targeting psychological attributes. Related work has evaluated programs to strengthen self-efficacy (McKelway, 2025), foster a proactive mindset (Campos et al., 2017), improve mental health (Angelucci and Bennett, 2024), and cultivate grit and patience among children (Alan, Boneva, and Ertac, 2019; Alan and Ertac, 2018). By focusing on aspirations, gender attitudes, and educational outcomes, our findings highlight the promise of brief, scalable interventions that operate through the demand side of education in fragile contexts. While the extremely disadvantaged nature of our study population and the scarcity of role models in their environment may have amplified the effects, these same features underscore the importance of demand-side interventions to support critical schooling decisions where change is most needed.

3.2 Context

Ever since the devastating civil war of 1988-91, the vast majority of school-age children in Somalia have had limited access to basic services in general and basic education in particular. Conflict, combined with recurring droughts and floods, has been a daily reality for communities and their children. Over 1.5 million people have been living as Internally Displaced Persons (IDPs) for decades across the three regions of Somaliland, Puntland and South Central Somalia (SCS). The government of Somalia faces serious challenges in governing and financing public education due to a lack of financial, human and managerial resources. Parents are required to pay fees to enrol their children in public schools, which contributes to the marginalisation of children from poor and rural families. Moreover, capacity constraints have caused most education services, including examinations and certification, to be provided by private Education Umbrellas rather than the public education system, resulting in a lack of regulation and standardisation.

Unsurprisingly, literacy rates in Somalia are among the lowest in the world. World Bank (2019) estimated that in 2018, only half of the adult population could read and write. Although primary education (grades 1 to 8) has been mandatory since 1975, only 24% (22% for girls versus 25% for boys) of children of primary school age (6-13) are enrolled in primary schools (Ministry of Education, 2022). Girls are equally likely to be enrolled in primary school; however, a significant gap emerges at the secondary level (World Bank, 2019). Among children of secondary school age (14-17), according to the most recent data, net secondary school enrollment is just 15% among girls and 19% among boys.

Gender norms are severely biased against women (UNDP, 2019). The 2022 UNDP Gender Inequality Index, a composite score that assesses the status of women's reproductive health, empowerment, and labour market participation, ranked Somalia third-to-last globally due to extreme gender inequality (UNDP, 2023). Entrenched traditional views assign women and girls secondary status in Somali society. This perpetuates narrow gender-based roles and inequalities. Participation of women and girls in decision-making spheres is limited. Female teachers are unlikely to serve as role models for young women, as only 14% of teachers in primary and 4% in secondary schools are women (Ministry of Education, Culture and Higher Education, 2017). Our paper is one of only a handful of field experiments conducted in Somalia, due to the challenges of conducting research there.[4] A recent review of economics research on African countries in the past 20 years by Porteous (2020) shows that over 40% of published papers are about five countries accounting for less than 20% of the African population.[5] Somalia is one of the "scarce 7" countries,[6] which have equivalent population as the "frequent 5" but are rarely studied and account for less than 5% of economics publications. This makes our data collection efforts in Somalia particularly relevant.

3.3 The Role Model Intervention

We conducted the role model intervention in partnership with Save the Children International (SCI). SCI was already implementing a set of education programs in 46 primary schools in two Somali regions, Puntland and South-Central, in collaboration with the Ministry of Education. The programs focused on increasing access to education for marginalised children (girls, boys, pastoralists, minority groups and children with mild disabilities) in targeted areas.[7] They included upgrades and rehabilitation of classrooms and school facilities to create safe and child-friendly learning environments. SCI also assisted schools in developing and implementing School Improvement Plans, training teachers in basic pedagogy

[4]Other notable recently published or working papers are: Abdullahi et al. (2023), Brar et al. (2023) and Gulesci et al. (2025).

[5]The most commonly studied African countries are Ghana, Kenya, Malawi, South Africa and Uganda.

[6]The other countries are Angola, Chad, D.R. Congo, Guinea, South Sudan and Sudan.

[7]The target regions and schools were chosen based on three criteria: a) the needs of children and the extent to which their rights are violated in the region; b) the absence or limited interventions by other NGOs or agencies in addressing the plights and rights of children in a holistic and integrated manner; c) operational and technical capacity of SCI and its partners to successfully implement similar intervention in the region.

and a code of conduct, enhancing school governance, and boosting student literacy, among others (Save the Children, 2018).

We randomly selected 23 of the 46 primary schools[8] in which SCI was active to receive a role model intervention, using a pairwise randomisation scheme.[9] Given the small number of SCI schools (46), we used pairwise randomisation to select the schools that would receive the role model intervention. We first stratified schools by region and sorted them by the number of grades offered and the total number of enrolled students. We then conducted pairwise matching (Imbens and Rubin, 2015). Within each pair, we randomly selected one school to be treated. Abadie and Imbens (2011) show that matched-pairs randomisation (and stratified randomisation) can increase power in small samples. Simulation evidence presented in Bruhn and McKenzie (2009) supports these findings, though they note that, for large samples, there is little gain from different methods of randomisation over a single draw. Imai, King, Nall, et al. (2009) also derive properties of matched-pair cluster randomisation estimators and demonstrate large efficiency gains relative to pure simple cluster randomisation. The randomisation was successful with no significant pre-intervention differences at the school level in number of classrooms, number of students enrolled (and boy/girl ratio), number of available grades, school fees, length of time operational, girl friendly space availability, number of female teachers (see Table 3.1, which we discuss in Section 3.4).[10]

Each of the treated schools received visits from one male and one female college student (our role models) on the *same day*. Each role model visited two grades, for a total of four treated grades per treated school. Since schools had differing numbers of grades, we decided which grades would receive the role model visits based on the available grades, as follows. In schools with eight grades (46% of the treatment schools), we treated grades 2, 4, 6, and 8. In schools with seven grades (17% of treatment schools), we treated grades 2, 4, 6 and 7. In schools with six grades (9% of treatment schools), we treated grades 2, 4, 5 and 6. In schools with five grades (13% of treatment schools), we treated grades 2, 3, 4 and 5. Finally, in schools with four grades (17 % of the treated schools) we treated all grades. At each treatment school, we randomly assigned two treated grades to be visited by a male role model and two by a female role model.

[8] Figure 3.A.1 in the Appendix shows the locations of the treatment and control schools.

[9] We originally had 47 schools, but one school was excluded as it was an all-girls school.

[10] All the schools in this study were supported by SCI through the Norad program. The majority (37) of the schools were located in urban areas. Of these schools, 14 mostly catered to children from displacement camps, which are located in the periphery of urban centres.

3.3.1 Implementation

The most critical aspect of the field experiment was identifying appropriate role models, who were college students with similar backgrounds and who had experienced challenges similar to those of the primary school children being visited. To recruit them, we worked with head teachers from SCI schools to identify former students currently attending universities in Somalia. From the long list of potential role models, we worked with the SCI program staff to shortlist college students whose life stories were especially inspiring, i.e., students who had achieved academic excellence and pursued university education despite personal difficulties. These role models were invited by SCI program staff to a briefing session in March 2018 to inform them about the role model intervention and to assess their willingness to collaborate with us on the project. A total of 15 role models agreed to participate in the intervention: nine women and six men.

Before the school visits, we invited the role models to a day-long coaching session. In this coaching session, we outlined the structure of the school visits and assisted them in preparing a 30-40-minute speech targeted at primary school students. The role models were invited to include the following items in their speeches: 1) introduction (names, how old they are, which schools they went to including the current university); 2) what challenges they experienced while in school, how did they overcome them; 3) whether they experienced failures and what they would do differently if they could go back in time; 4) future plans and job prospects.[11] At the end of the coaching session, the role models chose the schools they would visit (among those in the treatment group) based on proximity to their place of origin. Each role model visited between one and four schools. Some of the role models were alumni of the sampled schools; even those who were not, had grown up in nearby regions and faced similar challenges as the students they went to speak to, such as growing up in poverty, drought, family migration or having to live away from parents to continue with schooling. The most common fields of study among the role models were medicine (or related fields such as pharmacological sciences), law, and economics.

The intervention took place between April and May 2018. Each role model visited one school per day and gave two speeches, mostly in the morning hours. On all occasions, two role models visited the school. This ensured that they delivered their speeches to all four classes selected to receive the intervention. Each role model was provided with 50 USD per school visit to cover lunch and

[11] Each role model was given 30 minutes to practice their speech in front of fellow role models and their coaches (SCI Norad program staff). We held a separate coaching session for each region where the program operated.

transport costs.

At all treated schools, SCI Norad program staff sought permission for the role model visits from the school administration at least 1 week before the scheduled visit day. The speeches, delivered in the selected classrooms in the presence of the class teacher, lasted 40–60 minutes and concluded with 5–10 minutes of student questions facilitated by the teacher. Since the visits occurred during normal class time, there was no contamination of students in control classes within the treated schools.

3.3.2 Data Collection and Outcome Variables

Due to budget and logistical constraints, we were unable to collect student-level data before the intervention. Therefore, the only pre-intervention baseline data available to us are school-level data from SCI. We discuss these data in Section 3.4.1.

We implemented three separate waves of endline data collection. First, in October and November 2018, we administered a short survey in the sampled schools, involving students in grades 4 to 7 at baseline. This constitutes our First Endline data collection. We implemented the Second Endline data collection in April and May 2020. This was limited to students in grade 8 at the time, who were about to graduate from primary school. We implicitly assume that these students graduating in 2020 were in grade 6 at the time of our intervention. For these students, we also obtained their performance on the standardised exit exam that all grade 8 students must take to progress to secondary school. Finally, between May and August 2022, we implemented the Third Endline data collection. In May 2022, we surveyed grade 8 students who were about to graduate from primary school. We implicitly assume that these students graduating in 2022 were in grade 4 at the time of our intervention. For these students, we also obtained performance in the standardised exit exam. In addition, in August 2022, we implemented a short survey with the grade 8 teacher at each school (at that time). The survey also collected information on high school enrollment and family outcomes (marriage and children) for each of the graduating students surveyed in 2022 and 2020, i.e., those who would have been in grades 4 and 6 at the time of the intervention.

We focus on students from baseline grade 4 and grade 6 in the long run because, following our randomisation, role models visited baseline grades 2, 4, 6, and 8 in all treatment schools with eight grades in Spring 2018 (about 50%

of the sample).[12] This implies that our impact evaluation can rely on a clean comparison of outcomes between (*always treated*) students in the treatment schools and students enrolled in the same grades (and taking the same exit exam) in control schools.[13]

3.3.2.1 First Endline: October-November 2018

The First Endline data collection took place six months after the intervention, between October and November 2018. We collected survey data in 41 control and treated schools.[14] The data collection relied on a self-administered short questionnaire (two pages only), which recorded information on students' family background, aspirations concerning education and future occupation and answers to a set of questions aimed at measuring gender attitudes.

An enumerator went through the questionnaire in class, with the teacher present, explaining the meaning of each question. The students then answered the questions independently. Due to concerns regarding younger students' ability to read and comprehend the questions, we surveyed only students enrolled in grades 5 to 8 at the time of data collection (i.e., baseline grades 4 to 7), provided they were present at school that day. This resulted in a sample of 1,941 students. Recall that baseline grades 4 and 6 were always treated in treatment schools, whereas baseline grades 5 and 7 were treated only in schools with five grades and in schools with seven grades, respectively. This means that our sample includes untreated students in control schools, students in treatment classes in treatment schools, and students in untreated classes in treatment schools. Since the latter group may have still been affected by the role model visits through their peers or teachers, we refer to them as students in *spillover grades*.[15]

As discussed in Section 3.2, gender norms in Somalia are severely biased against women. One of our objectives is to test whether exposing students to role models – especially young women pursuing a college degree – can affect girls'

[12]Grades 2 and 4 were treated in all schools, since all sampled schools had at least four grades. Grade 6 was treated in all schools with at least six grades, and grade 8 was treated in all schools with eight grades.

[13]The students in baseline grade 8 graduated before the First Endline. We had originally planned to survey them just before their graduation. However, due to delays in implementing the intervention, we had to conduct the First Endline data collection at the beginning of the following academic year rather than at the end of the intervention semester.

[14]Because the survey was self-administered, we received mostly blank forms from three schools. Two other schools were part of another education technology evaluation within the wider SCI program; thus, they were excluded from the 2018 data collection.

[15]There are 559 such students for which we have survey data from the First Endline. Our primary results are unaffected by excluding these students from the analysis.

and boys' attitudes toward gender equality in education and job participation. To measure gender attitudes, we asked the following set of agree/disagree questions:

1. "More encouragement in a family should be given to sons than daughters to go to college."
2. "It is more important for boys than girls to do well in school."
3. "Boys are better leaders than girls."
4. "Girls should be more concerned with becoming good wives and mothers than desiring a professional or business career."[16]

We average individual responses to the four questions into an index of attitudes toward gender equality, measured on a scale from 1 to 4, where higher numbers indicate attitudes more strongly in favour of gender equality.

Our other primary outcome variable is students' educational aspirations. In the survey, we asked students whether they intended to enrol in secondary school (and the name of the school they plan to attend). We also ask them to report the highest level of education that they would like to complete "if nothing could stop them." The answers to this question generate an indicator of college aspirations, which is equal to 1 if the student's highest level of education is a college degree (or higher), and 0 otherwise.

While we did not pre-register or pre-specify our two primary survey outcomes (our measure of gender attitudes and our measure of college aspirations) before the First Endline data collection (in 2018), we consistently use the same survey questions and generate the same outcome variables in subsequent data collection waves.

3.3.2.2 Second Endline: April and May 2020

We implemented the Second Endline in April and May 2020, approximately 2 years after the role model visits. Due to time constraints imposed by the emerging COVID-19 risk, we surveyed only students enrolled in grade 8 (who would have been in baseline grade 6 at the time of the intervention). At the time of the Second Endline, these students were about to graduate from primary school. We have

[16] In the First Endline, we also included a fifth statement: "Girls are as smart as boys." However, this question showed a positive item-rest correlation (when a negative one was expected) and lower item-test-retest reliability, suggesting students' misunderstanding. We therefore dropped it from the index, resulting in a Cronbach's alpha of 0.66 for the scale (previously 0.61). The Second and Third Endline survey data only include the four components of the gender attitude index.

data from 32 of the original 46 schools. These are the 21 schools that had all eight grades at the time of the intervention, plus an additional 11 schools that had grown to include all eight grades at the time of the Second Endline. These 32 schools are our target set of schools for subsequent data collection efforts (Third Endline). Half of them are control schools, and half are treatment schools. In 8 of the 16 treatment schools, grade 6 was randomly assigned to the male role model visit, while the remaining eight were assigned to the female role model intervention in Spring 2018.

We have survey data from 829 grade 8 students across 32 schools who were in class on the day of the Second Endline data collection. Importantly, in 2020, for each surveyed student in all the 32 schools, we obtained performance in a standardised exam that all grade 8 students need to take at the end of the school year to be able to graduate from primary school and progress to secondary school.[17] To match survey data and exam data, we had the assistance of the teacher who had taught the grade 8 class in 2020. We presented the teacher with the list of surveyed students and, with the teacher's help, we identified the corresponding student and retrieved his/her exit exam grade from the school records. Since schools do not employ unique student identifiers and many students have very similar (or the same) names, it would have been impossible to attempt this matching independently. Through this teacher-aided matching process, we infer that students whose exit exam results were missing (nearly 30%) had not taken the exam.

Therefore, we have two sets of outcomes for students who were in grade 8 at the time of the Second Endline (2020) and who would consequently have been in baseline grade 6 – hence *always treated in treatment schools* – at the time of the role model visits. The first set of outcomes comprises our survey-generated variables of interest, i.e., gender attitudes and college aspirations; the second outcome is the grade obtained in the standardised exit exam.

3.3.2.3 Third Endline: May-September 2022

The Third Endline took place between May and September 2022. In May 2022, we administered a student survey to grade 8 students shortly before they took their mandatory standardised exit exam. These students would have been the baseline

[17] The country-wide exams are coordinated by the Ministry of Education. Exams are taken across seven subjects, including Islamic studies, Arabic, Somali, Social Science, Science, English and Maths. Scores from each subject are averaged to form a candidate's final score. A candidate is deemed to have passed upon attaining 50% in the final score. In the 2020 exams, results showed that the average national score was 59%, but up to 90% of the students achieved the pass mark with hardly any gender gap (UNESCO, 2022).

grade 4 visited by the role model. We targeted all 32 schools with eight grades in 2022. These were the same schools surveyed in the Second Endline. Of them, 16 are control schools and 16 are treatment schools. However, the within-school randomisation of role model gender did not result in equal numbers of grade 4 classes assigned to male and female role models at the time of the intervention (2018). Purely by chance, in this case, only 4 of the 16 treated schools (with eight grades at the time of our Second Endline) had baseline grade 4 students assigned to a male role model visit. The remaining twelve schools had their baseline grade 4 students visited by a female role model.[18] An additional unforeseen challenge was that the four schools where baseline grade 4 students were assigned to the male role model intervention generated only 35 student surveys in 2022. This small sample size makes it impossible to evaluate the impact of the male role model intervention on survey measures and education outcomes.[19] We therefore have no choice but to drop these four schools from the analysis, leading to a working sample of 28 schools, which were either assigned to the control group or were in the treatment group and had their grade 4 students assigned to female role models in Spring 2018. From these schools, we have survey data from 575 students enrolled in grade 8 at the time of the Third Endline.

We returned to the schools in August 2022 to collect data on student performance on the exit exam. We followed the same procedure as for the Second Endline, i.e., we matched exam and student data for all surveyed students with the help of the grade 8 class teacher. Additionally, as part of the Third Endline, we surveyed the grade 8 class teacher at each school to gather information about the students in our sample who graduated in 2020 (baseline grade 6) and in 2022 (baseline grade 4). Specifically, for each surveyed student in our sample, we asked the corresponding grade 8 class teacher three questions: 1) whether the student enrolled in high school; 2) whether he or she was married, and 3) whether he or she had children. The teachers were able to provide this information unless they had not taught the corresponding grade 8 class, for instance, because they had recently transferred from a different school. When following up the students who graduated in 2020 (baseline grade 6), we encountered this problem in four treatment schools (two assigned to a female role model and two assigned to a male role model) and one control school, leading to a working sample (with no missing data) of 27 schools. When following up on the 2022 graduating class (baseline grade 4), we encountered this problem in only one control school, reducing the

[18] In the original randomisation, an additional five grade 4 classes were assigned to male role models and a further two grade 4 classes were assigned to female role models. However, they were in schools with fewer than eight grades; hence, we are unable to trace the corresponding students.

[19] We have 4, 6, 7 and 18 surveys returned from each of the four schools, respectively.

working sample for this student cohort from 28 to 27 schools.

As a result of the Third Endline, we therefore have three sets of data for the 2022 graduating cohort *(baseline grade 4)*: 1) survey data collected just before graduation from grade 8; 2) performance in the standardised exit exam; 3) enrollment in high school, as reported by the grade 8 teacher; 4) marital status and fertility outcomes, as reported by the grade 8 teacher. These data are available for the surveyed students in 27 schools,[20] of which 15 control, and 12 treatment schools (all assigned to female role models). In addition, the Third Endline also generated teacher data on secondary school enrollment, marriage and fertility for the 2020 graduating cohort *(baseline grade 6)*, for students who graduated from 27 schools,[21] of which 15 control and 12 treatment schools, 6 assigned to male role models and 6 to female role models.

3.3.3 Ethical Considerations and Field Challenges

This study was subject to ethical approval from the Ethics Review Boards at Southern Methodist University (intervention and First Endline), Texas A&M University (Second and Third Endlines) and Lancaster University (Third Endline). As part of the wider Norad Framework evaluation, administrative approval was obtained from the Ministries of Education in Puntland State and South-Central Somalia State, Somalia. Data collection in that program was supervised by local Ministry of Education administration staff, respective school administrators, and SCI Norad program staff.

All data collectors were trained on SCI's Child Safeguarding Policy and were required to append their signatures indicating their understanding of and commitment to abide by it during data collection. Informed consent and assent were obtained in line with SCI precedents. One week before data collection, as part of the mobilisation and preparation efforts, the headteachers of the respective schools informed parents about the study and the data collection (its purpose, the information that would be collected, and whether there would be any benefits or risks to participating). The parents would then inform the head teacher or the school Community Education Committees (CECs) if they did not consent to their child's participation in the study. Additionally, on the day of data collection,

[20] As explained in the text, this is a subset of 28 schools covered in the 2022 student survey after excluding (from the original 32 schools) the four schools that were allocated male role models in grade 4 and for which we received very few survey responses. We further exclude one control school whose grade 8 teacher had changed and could therefore not answer questions about the graduated students.

[21] This is a subset of the 32 schools covered in the 2020 student survey. We exclude five schools where the grade 8 teacher had changed between 2020 (student graduation year) and the time of the survey, leading to missing data on secondary school enrollment and marital status.

students were informed by their class teachers of the purpose and the information to be collected. Then, students were read the consent form aloud and had a chance to read it independently, as it was attached to the survey (on the first page). Students were invited to return the questionnaire "empty" if they did not consent to participate in the study.

Implementing three waves of the survey and complementing them with exam data and teacher-generated student information at the Second and Third Endlines in this fragile context required substantial institutional coordination and innovative data-collection strategies. The absence of reliable student rosters and the varying grade offerings across schools meant we could not track individual students over time or assess the intervention's uptake (based on school attendance during the role model visits). Importantly, in the Second and Third Endlines, we can only identify treatment status based on "current" grade at the time of the follow-up data collection. Subsequently, we assigned students to treatment groups based on their current school and grade, inferring the grade they would have been in at the time of the intervention, assuming they were enrolled in the same school. This may result in some misclassification, particularly in the longer-term follow-ups as students move from schools offering only grades 1 to 5 into schools that offer grade 8. Still, such misclassification would likely attenuate rather than exaggerate estimated intent-to-treat impacts, making our results conservative.

Additionally, the fragile context of our study—where student displacement is common and permanent contact information is scarce—made it infeasible to follow students and re-survey them after primary school graduation. To capture post-graduation outcomes such as high school enrollment, marriage, and fertility, we therefore adopted an innovative strategy of surveying grade 8 teachers, who are typically well acquainted with their students' trajectories. Because this approach depends on teachers' personal knowledge rather than official records, the data were available only when the teacher responsible for the 2020 and 2022 grade 8 cohorts remained at the school at the time of our survey. In practice, this limitation affected five schools from the 2020 cohort and one school from the 2022 cohort.

While we acknowledge the data limitations and the challenges of conducting fieldwork in this fragile setting, our strategies enabled us to assemble a uniquely comprehensive dataset: survey-based measures of aspirations and attitudes collected shortly after the intervention and again two and four years later, standardised exam results at graduation for two student cohorts, and teacher-reported indicators of post-primary enrollment and early-life choices for those same cohorts.

3.3.4 Empirical Strategy

Our main interest is in estimating the impacts of male and female role models on male and female students. When analysing survey data collected during the first endline (Fall 2018), i.e., data from students in baseline grades 4 to 7, we estimate the following equation:

$$\begin{aligned} Y_{igs,j} = \alpha + \beta_1 FRM_{gs,j} + \beta_2 MRM_{gs,j} + \beta_3 F_i * FRM_{gs,j} + \beta_4 F_i * MRM_{gs,j} \\ + \beta_5 F_i + \gamma S_{gs,j} + \delta \mathbf{X}_i + \lambda_j + \zeta_g + \epsilon_{gs} \end{aligned} \tag{3.1}$$

Where Y_{igs} represents the outcome for child i in grade g in school s in pair j; $FRM_{sg,j}$ and $MRM_{sg,j}$ indicate, respectively, the Female Role Model and the Male Role Model treatment assignments for grade g in school s in pair j in Spring 2018. We include a dummy for the gender of student (F_i) and interact it with the two Role model treatment indicators. This allows us to test whether role models of different genders had differential impacts on boys and girls. We also report p-values generated by testing $\beta_1 + \beta_3 = 0$ and $\beta_2 + \beta_4 = 0$, i.e., the total effects of female and male role models on girls, respectively. Since some students were (randomly) in untreated grades in treatment schools, we include an indicator S for being in what we call a "spillover grade" g in school s in pair j.

We control for child characteristics X that may also affect the outcomes of interest. Specifically, we control for student age, a proxy for poverty, i.e., whether the student reports having missed school in the past month due to "lack of money for fees, uniform or books" and a measure of exposure to female role models within the family, i.e., whether a female relative went to college.[22] Due to the pair matching randomisation method, we include pair fixed effects λ_j, as well as grade fixed effects ζ_g. Standard errors are clustered at the unit of randomisation, which is the grade-school level.

For the analysis of treatment effects on the baseline grade 4 and grade 6 cohorts, based on the 2022 and 2020 student surveys, respectively, we estimate a similar model with some modifications. By design, every grade 4 and grade 6 student in Spring 2018 was assigned to a treatment school, yet the gender of the role model

[22] The First Endline survey also elicited additional information on family background, i.e., whether the student lives with both parents, and whether each parent completed secondary school, which we report in Table 3.1. However, these survey questions were not included in the Second Endline due to the need to keep the questionnaire even shorter. Our 2018 results are robust to the inclusion of the additional individual controls in the specification.

was randomised across grades within each school. Since this analysis involves only one cohort per year (either treated or untreated, depending on the school), we cannot exploit grade-level variation in the empirical analysis. The standard errors are clustered at the school level, which is the unit of randomisation in this case. Moreover, since we survey the target students when they were in grade 8 and about to graduate, the sample size is reduced to the set of schools that offered grades 1 to 8, and for which we have survey, exam and teacher-generated data, as explained in Section 3.3.2.3.[23] Subsequently, the number of intact school pairs per cohort decreases, making pairwise comparisons infeasible. We thus substitute school-pair fixed effects for school-level controls.

To assess the impact of the role model intervention on educational, marriage, and fertility outcomes, we make additional modifications to our estimations. First, to hedge against the small sample size and the limited number of clusters, we pool data from the Second and Third Endline. Moreover, these outcomes are elicited at a similar educational stage. Despite this pooling, the number of intact randomisation pairs shrinks to only 12 school pairs. Thus, we substitute pair-fixed effects for school-level controls. Secondly, we focus on estimating the impact of the role model intervention at the school level, i.e., without distinguishing between male and female role models. The latter is necessitated by the lack of observations corresponding to the male role model among the 2022 graduating cohort. This allows us to retain all schools for which we have data, regardless of the gender of the assigned role model at the time of the intervention, thus maximising statistical power. We therefore estimate the following equation:

$$Y_{isg} = \alpha + \beta_1 T_s + +\beta_2 F_i * T_s + \beta_3 F_i + \delta \mathbf{X}_i + \zeta_g + \epsilon_s \tag{3.2}$$

Where T_s is an indicator equal to 1 if school s was assigned to the role model treatment and 0 otherwise. As before, we include grade fixed effects ζ_g to control for student cohort. We cannot, however, include pair-fixed effects λ_j, as some of the treatment schools were initially paired with control schools that did not offer grade 8 and therefore do not appear in the Second and Third Endline samples. We still include school and student controls in the specification. Standard errors are clustered at the school level in this case.

[23] When generating our working sample for the post-graduation outcomes, we are affected by the fact that data on high school graduation, marriage and fertility for the 2020 graduating class are not available for all schools. This leads to a working sample of 27 schools for which we have all student data for the 2020 graduating cohort, and a working sample of 27 schools for the 2022 graduating cohort, as detailed in Section 3.3.2.3.

To account for the small number of clusters, we follow Angrist and Pischke (2008) and Cameron and Miller (2015) and apply the correction for the small number of clusters by using wild cluster bootstrapping (Cameron and Miller, 2015). In our regression tables, we report wild-bootstrapped p-values in addition to clustered standard errors. In addition, given that we have two survey-generated outcomes and four outcomes retrieved from schools and teachers, we correct the p-values associated with individual hypotheses within the same family of outcomes using the step-down multiple testing method developed by Romano and Wolf (2005). Finally, in the Appendix, we also present separate regressions by gender, and confidence intervals for p-values based on randomisation inference (Heß, 2017).

3.4 Results

3.4.1 Balance Tests and Attrition Tests

In Table 3.1, we report descriptive statistics of baseline school characteristics for control and treatment schools. The Spring 2018 (baseline) school-level data show no significant differences between treatment and control schools in any of the available school characteristics, including the number of students, the ratio of boys to girls, student-teacher ratio and the number of female teachers as well as whether the school charges fees and whether it has girl-friendly spaces. At the time of the intervention, the schools were less than 10 years old, had 8 classrooms, and had an average of about 32 students per teacher, with an average of about 230 students and 9 teachers per school. Nearly 35% of schools required school fees, and fewer than 20% had girl-friendly spaces. About 50% of the schools offered grades 1-8, with an average of 6.6 offered grades. Teachers were primarily men, with fewer than 2 of the 9 teachers per school being women.

As discussed in Section 3.3.3, our working sample of schools becomes smaller over time. Our First Endline survey data were generated from 41 schools, and our Second and Third Endline surveys from 32 and 28 schools, respectively.[24] Reassuringly, the school characteristics remain largely balanced between treatment and control arms with the decline in the number of schools.[25] Additionally, the characteristics of the schools retained in subsequent surveys do not drastically change, implying non-selective attrition over our survey waves. Our attrition

[24]As previously discussed, the fall in the number of schools is due to targeting schools offering 8 grades, the exclusion of schools that returned only a handful of student surveys, and the unavailability of grade 8 teachers who could help us match surveyed students with exam data.

[25]See table 3.B.1 in the Appendix.

tests, displayed in Table 3.B.2 in the Appendix, confirm no significant probability of attrition by treatment status. They do, however, show that attrited schools are smaller than schools in the follow-up working sample. This is a direct result of our data collection strategy, which, for the Second and Third Endline, required sampling schools that offered 8 grades to access data on exit exam performance and post-primary education outcomes. As a result, the attrited schools have fewer grades and, hence, fewer students than the non-attrited schools. Among other school characteristics, only the school's age and the charging of school fees consistently predict the likelihood of attrition. Both variables, however, may be correlated with school size; in fact, schools with eight grades are older and more likely to charge school fees.

In Table 3.2 we test for differences in individual characteristics among students assigned to male versus female role models within the treatment group. Note that this analysis is based on data collected after the role model intervention was implemented, as we were unable to collect baseline student data. By necessity, we restrict the analysis of student characteristics to the schools for which we have First, Second and Third Endline survey data, respectively. Recall that the First Endline survey was conducted only on students enrolled in grades 5 to 8 in Fall 2018, while the Second and Third Endline surveys were conducted among students enrolled in grade 8 in Spring 2020 and Spring 2022, respectively. The implicit assumption is that the students surveyed in the First Endline were enrolled in the same schools and in baseline grades 4 to 7 when the intervention took place in Spring 2018. Similarly, we assume that the grade 8 students surveyed in 2020 and 2022 were enrolled in the same schools in grades 6 and 4 at the time of the intervention. We can partially verify the assumption regarding the First Endline data using information from our survey, which asked students about their school and grade in the previous academic year. Our data show that only 1% of students had changed schools the summer immediately following the intervention, which is reassuring. We are, however, unable to conduct the same checks for subsequent data collection waves.

From the First Endline, we have data for a total of 1,941 students with non-missing information for any of the variables in the sample,[26] of which 846 were in control schools and 1,095 in treated schools. From the Second Endline, we have data on 829 students (348 in control and 481 in treatment schools) enrolled in grade 8 at the time of the survey, hence likely in baseline grade 6 at the time of the intervention. From the Third Endline, we have data from 575 students (253

[26]We have missing age data for 251 students. We imputed their age using the average seen in the sample, hence increasing our sample size from 1690 to 1941 students.

in control and 322 in treatment schools), all enrolled in grade 8 in Spring 2022, hence likely in baseline grade 4 at the time of the role model visits.

The First Endline data show some imbalances in students' likelihood to have missed school due to poverty,[27] which is lower for students exposed to female role models than for students exposed to male role models or in control schools. We do not see statistically significant differences in the number of missed school days in the previous two weeks, the likelihood of living with at least one parent, the likelihood that a parent has completed secondary schooling, or the likelihood that a close relative (i.e., mother or sibling) will attend college. Specifically, about 16% of students had mothers who had completed secondary school and 24% had fathers who had completed secondary school, and only 6% had a close female relative who went to college, with no significant differences by treatment status.

Similarly, the comparison of student-level characteristics across treatment groups for the 2020 and 2022 graduating cohorts shows some small and not significant differences. Appendix Table 3.B.3 reports additional student-level comparisons across treatment groups, restricting the Second and Third Endline data to students with exam data. Again, it shows a few differences in student characteristics. Nevertheless, we control for students' characteristics in our main regression specifications.

3.4.2 Impacts on Gender Attitudes and College Aspirations

We start by discussing the differences in gender attitudes and college aspirations observed in our First, Second and Third Endline data. In the top panel of the appendix Figure 3.A.2, we report the simple (non-standardised) mean of the responses to the four questions that constitute the gender attitude index, described in Section 3.3.2, by treatment status and student gender. The index ranges from 1 to 4, with higher values indicating more gender equal attitudes. The bottom panel of the figure shows college aspirations, measured as the percentage of students in each treatment group and data collection wave who answered “college" when asked what the highest level of education they would like to complete “if nothing could stop them.” The figure suggests that the intervention, particularly visits by female role models, positively affected boys' and girls' gender attitudes in 2018 and 2020, with no significant difference between the treatment and control groups in 2022. We also note that both boys and girls become more gender-egalitarian over time.

[27]This is an indicator equal to 1 if the student stated that he/she missed school because he/she needed to work for pay or did not have money to pay for fees, books or other school costs.

The vast majority of students in the First and Second Endline (about 70% with no significant differences across genders) aspired to pursue a college education, as shown in the bottom panel of Figure 3.A.2. While this value may seem high, it is not unusual for children to express ambitious educational aspirations in surveys. Such responses may reflect a degree of naiveté or the challenges of accurately measuring aspirations, which can be perceived more as dreams than concrete goals. However, aspirations dropped quite dramatically, to about 40%, when we surveyed the 2022 graduating class in the Third Endline, possibly reflecting the sobering effects of the COVID-19 pandemic. This value, however, dropped quite dramatically, to about 40%, when we surveyed the 2022 graduating class in the Third Endline. Importantly, the raw data show no significant and consistent effect of the treatment on aspirations.[28]

Results from regression analysis, displayed in graphical form in Figure 3.1, as well as Table 3.3, largely confirm these findings. The estimates for gender attitudes are reported in standard deviations from the control mean. We find that classroom visits by female college students increased gender attitudes toward equality by about 0.36 standard deviations in the short term (First Endline in 2018), with no significant difference in the effects on male and female students. The gender gap observed in control schools, whereby female students hold attitudes that are about 0.28 standard deviations more toward equality as compared to male students, vanishes as a result of the female role model intervention (p-value = 0.251, Wald test for the sum of coefficients on Female and its interaction with Female RM). The results are robust to controlling for school and student characteristics, including school-pair fixed effects (exploiting the randomisation method we used to allocate schools to treatment and control groups), and to corrections for multiple hypothesis testing and for small numbers of clusters through wild bootstrapping. The First Endline data also show a marginally significant impact of the intervention on students in spillover classes, i.e., classes in treatment schools that did not receive role model visits.

In the middle and bottom panels of Figure 3.1, and in columns 3 and 5 of Table 3.3, we see the estimated role model impacts on the gender attitudes of the 2020 and the 2022 graduating students cohorts, who were surveyed in the Second and Third Endlines, respectively. We see that the estimated coefficient for the Female Role Model indicator remains positive but decreases over time and becomes statistically insignificant. [29]

[28]The Second Endline data for the 2020 graduating class suggest a positive effect of female role models on boys' college aspirations.

[29]The lack of impact on gender attitudes could be attributed either to the longer time-frame

As a robustness check, we conduct randomisation inference (Heß, 2017) and display the results in Table 3.B.4 (Columns 1 and 2) in the Appendix. This analysis confirms that the role model visits had a significant impact on students' gender attitudes, particularly those of boys [coefficient = 0.519; p-value (95% confidence interval) = 0.000, 0.009] six months post-intervention. The coefficients become substantially smaller and lose statistical significance when focusing on the 2020 and 2022 graduating classes (baseline grades 6 and 4), surveyed 2 and 4 years post-intervention, respectively.

Figure 3.1 and Table 3.3 also report estimates over time for college aspirations. We see no significant impacts of male and female role models on boys' and girls' aspirations in the First Endline. These null results are also observed for the 2020 graduating cohort (baseline grade 6), in the Second Endline data. However, the college aspirations of the 2022 graduating girls, who were surveyed in the Third Endline (4 years post-intervention), show evidence of a marginally significant impact of the female role model visits on their college aspirations. This finding suggests that aspirations may not shift immediately, but instead evolve through reflection and lived experience, shaped by exposure to opportunities and constraints, the accumulation of setbacks and achievements, and the encouragement or discouragement received from significant others.[30] The results are again confirmed using randomisation inference in Table 3.B.4 (Columns 2, 4 and 6) in the Appendix.

3.4.3 Impacts on Education Outcomes, Marital Status and Fertility

Our Second and Third Endline data collections enable us to examine the impact of the role model intervention on the educational outcomes of the surveyed students by linking survey responses with their performance on the standardised exam required for primary school graduation. In addition, through surveys of grade 8 teachers, we collected information on students' secondary school enrollment, marital status, and fertility outcomes. This analysis focuses on two grade 8 student cohorts: those surveyed in 2020 (baseline grade 6) and in 2022 (baseline grade 4). The sample is further restricted to schools that offered grade 8 and in which we were able to survey teachers who had taught the graduated students, as detailed in Section 3.3.2.3.

Following the empirical strategy outlined in Section 3.3.4, we pool the data

since the intervention, attrition, or the reduced power given the smaller sample size and only one grade surveyed.

[30] Recall that we are unable to analyze the effect of male role models in 2022 due to lack of data, as detailed in Section 3.3.3.

obtained for the 2020 and the 2022 graduating cohorts, and estimate the impact of role model visits on: i) the likelihood of sitting the grade 8 exam, ii) student performance in the exam (grade standardised around the control mean), iii) the likelihood of being enrolled in secondary school, and iv) the likelihood of being married and/or having children. For these outcomes, we primarily assess the impact of any role model visits, without distinguishing between male and female role models.

Approximately 80% of students took the mandatory exit exam, and the intervention did not impact this outcome. When examining the 984 students who took the exam across the two student cohorts, we observe evidence of a positive, albeit imprecise, impact of the role model visits on girls' exam performance (a 0.17 standard deviation increase), as shown in Table 3.4. The intervention also increased the likelihood of both boys' and girls' enrolling in secondary school by 6-10 percentage points, i.e., a 7-11% increase over the enrollment rate observed in the control schools (88%). We also see in column 4 a negative, though imprecise, impact of the intervention on students' likelihood of being married or having children. [31]. The point estimates on secondary school transition and the likelihood of being married or having children are identical for both boys and girls. However, positive effects on secondary school transition among girls are robust to multiple hypothesis testing and to a small number of clusters, but not to randomisation inference. In contrast, these effects among boys are only robust to randomisation inference.

3.4.4 Robustness: Sensitivity to Attrition and School-Level versus Grade-Level Treatment Effects

A key challenge in our empirical analysis is attrition at both the school and grade levels. As our initial design included schools that did not have grade 8, our mid- and long-term follow-up attempts and our analysis of education outcomes for the grade 8 graduating cohorts in 2020 and 2022 necessarily rely on a reduced sample. With our reduced sample, the number of intact randomisation pairs is too small to conduct pairwise comparisons. A further complication arises from an accidental imbalance in the allocation of male role models to baseline grade 4 classes in treatment schools, resulting in too few 2022 grade 8 students being

[31] We combine marriage and children outcomes, as the percentages of individuals who are married and those who have children are smaller than 10%, and those with children tend to be a subset of those who are married. When analysing marriage and children outcomes separately, we find negative impacts of the intervention on both, though only the effect on marital status is statistically significant. This could, however, be due to power issues, as the sample size for the children's outcome is reduced because not all surveyed teachers had this information on the graduated students. Results are available upon request.

exposed to them. As a consequence, our long-run analysis focuses on examining the impacts of any role model visit, without distinguishing between male and female role models.

We conduct a few additional analyses to further assess the validity and robustness of our primary findings. First, we evaluate whether our estimated short-term impacts on the survey outcomes (gender attitudes and aspirations) are robust to restricting the sample to the schools that are not subject to attrition in the Second and Third Endlines. Second, we estimate the impacts of the role model visits at the school level, i.e., without distinguishing between male and female role models, by estimating equation (2) of section 3.3.4. While this empirical strategy prevents us from comparing the impacts of male and female role models, it maximises statistical power by including all students in treatment and control schools for whom data are available. This is especially valuable when assessing the long-term effects of treatment on educational outcomes.

When restricting the analysis to schools revisited during the Second Endline data collection, we continue to observe a significant impact of female role models on both boys' and girls' gender attitudes, with point estimates virtually unchanged, as shown in Table 3.B.5 in the appendix. However, the impacts become statistically insignificant when the sample is further limited to schools that did not attrit from the Third Endline in 2022. For girls, the estimated effect of female role models remains comparable in magnitude to that obtained for the full sample (0.306 versus 0.361 standard deviations), but is less precisely estimated due to the smaller sample size. In the restricted sample, we also find evidence of a significant impact of male role models on boys' gender attitudes.

For consistency, we also replicate the analysis of short-term survey outcomes using a single school-level treatment indicator in place of the separate indicators for male and female role models. As expected, the estimated impacts on gender attitudes and college aspirations are similar in direction to those obtained for the female role model treatment, but smaller in magnitude and less statistically significant. This pattern suggests that the effects on these outcomes may be primarily driven by exposure to female role models.

To assure that the long-term outcomes are driven by the female role models, rather than male role models, we restrict our treatment sample to schools assigned to female role models. We then compare outcomes of students exposed to female role models versus those in control schools by substituting the school treatment dummy with a female role model dummy in Equation 3.2. This specification allows us to determine whether the long-term results are also driven by female role models, as with the short-term improvements in gender attitudes. Restricting the treatment schools to only those exposed to female role models, however, shrinks the sample size, leading to wider confidence intervals. The estimated impacts, shown in Table 3.4, across all the long-term outcomes remain largely similar in magnitude but with reduced precision. This pattern is expected, as the restriction reduces the size of the treated sample by 24%, thereby lowering statistical power. Nonetheless, the estimates from the pooled and restricted samples are statistically indistinguishable: the confidence intervals overlap, and the direction and approximate magnitude of the effect remain stable. This pattern is thus consistent with the short-run results, which point to female role models as the primary drivers of the impact.

3.4.5 Mechanisms

A natural concern is that teachers, rather than students, may be driving the results. Here, teachers may be more motivated after observing role models or just reporting more positive views of their pupils. We do not directly observe teacher effort, but we do measure outcomes that would shift if either of these channels were active. In control schools, Table 3.B.10 shows that teachers already hold near-universal expectations of success: 99% of pupils are expected to transition to secondary school, and 95% are expected to attend college, with no notable gender differences. These ceiling-level beliefs leave almost no room for treatment-driven updates. Likewise, if teachers were upwardly biased in reporting their students' outcomes, we would expect more gender-egalitarian attitudes among those in treated schools. Yet we find no treatment effects on teacher gender attitudes in Table 3.B.11. Together, this evidence makes teacher-driven mechanisms unlikely explanations.

By contrast, the pattern of results is consistent with student-level mechanisms rooted in student beliefs and aspirations. Our role models provided relatable examples of educational progression, reducing uncertainty about the returns on schooling and clarifying the steps required to succeed. Female role models were particularly effective in this context, where girls face stronger social constraints. By challenging restrictive norms, they expanded what girls believed was possible and increased support for their schooling. Our interpretation thus aligns closely with interventions that directly address internal belief updating (Nguyen, 2008; Golan and You, 2021) and gender attitude change (Beaman et al., 2012; Dhar, Jain, and Jayachandran, 2022).

3.5 Discussion

Our findings suggest that classroom visits by female role models had strong short-run effects on students' gender attitudes. Six months after the intervention, both boys and girls expressed more egalitarian views; however, these impacts diminished over time and became statistically insignificant in the longer term. This pattern may be partly explained by a general trend toward more gender-equal attitudes across all students, which may have diluted the intervention's relative impact, or by other factors affecting both treated and control students. In contrast, effects on aspirations emerged more slowly. We detect no changes in college aspirations in the short- or medium-run, but in the longer run, girls exposed to female role models reported higher aspirations to attend college. Baseline aspirations were already very high, which is consistent with other studies documenting over-optimism or naiveté among children when asked about educational ambitions; therefore, our ability to capture marginal changes is limited. Nevertheless, the delayed effect among girls suggests that role models may influence aspirations in ways that become salient only as students approach key educational transitions.

The longer-term impacts on educational outcomes cannot be attributed solely to shifts in college aspirations, since these effects are observed for both boys and girls. In contrast, aspirations increase only in the longer term for girls. A more plausible mechanism is that exposure to female role models challenged prevailing gender stereotypes and offered students concrete examples of resilience and persistence. By broadening students' sense of what is achievable in the face of adversity, the role models may have encouraged continued educational investment, even when explicit aspirations did not shift immediately. They may also have conveyed valuable information about the returns to education, whether through sharing their own job experiences or describing their career expectations. These match results from other role model interventions that link exposure to internal belief updating (Nguyen, 2008; Golan and You, 2021) and gender attitude change (Beaman et al., 2012; Dhar, Jain, and Jayachandran, 2022). Our mechanism analysis further shows no evidence that teachers changed their attitudes or their expectations for students, suggesting that teacher-driven channels or reporting bias are unlikely explanations.

The role model intervention was extremely cost-effective. The classroom visits we implemented were relatively short and inexpensive, costing approximately $53 per role model per school, with each role model visiting two classrooms. This included the cost of lunch and travel for each role model visiting a school, as well as the per-person cost of implementing a coaching day with all the role models. [32] Our estimated educational impacts are large yet comparable to those reported by Riley (2024) for an intervention most similar to ours in terms of content and target population. Specifically, Riley (2024) evaluated a program in which students in the final years of lower-secondary (grade 11) and upper-secondary (grade 13) school were taken to the cinema to watch *Queen of Katwe*, a Disney movie featuring a female role model. Riley (2024) finds a 0.20 SD increase in girls' grade 11 math standardised test scores, with stronger effects (0.33 SD) among low-achieving girls (below the median on a mock exam). She also documents an 11 percentage-point increase in girls' likelihood of remaining in school through the end of upper-level secondary school. These impacts are comparable to the standardised test gains we observe for girls exposed to female role models, as well as to the higher likelihood of secondary school enrollment that we find for both boys and girls (see Table 3.4). These effects are also comparable to those of other role model interventions in low and middle-income countries.[33]

We acknowledge the potential lack of external validity of the magnitude of our estimates. The impacts we document may reflect the disadvantaged backgrounds of the children in our study, many of whom were displaced or living in single-parent households. In this context, exposure to college students, particularly female college students, was highly salient: only about 5% of students reported having a female close relative who had attended college, and access to role models through television or social media was virtually absent. The intervention, therefore, provided a scarce yet powerful source of inspiration that may not be replicated in settings where role models are more visible in families or through the media. Nevertheless, the evidence from this fragile and understudied context is especially valuable, as it underscores both the severe constraints faced by children in such environments and the potential for low-cost interventions that address demand-side barriers to generate large impacts when supply-side constraints are also addressed.

[32] A more formal cost-effectiveness analysis, following the approach of Kremer, Brannen, and Glennerster (2013), generates striking values. With approximately 30 children per classroom, an investment of roughly 100 USD would enable a female role model to visit two schools, reach 40-60 girls, and yield estimated gains of 14–21 standard deviations per 100 USD spent. These large values reflect both the substantial impact on test scores—comparable to other role-model studies—and the exceptionally low cost of the intervention.

[33] Table 3.B.12 summarises education-focused role-model interventions we are aware of, including differences in content, delivery, and timing. Our intervention is comparable in terms of the intensity and magnitude of its effects.

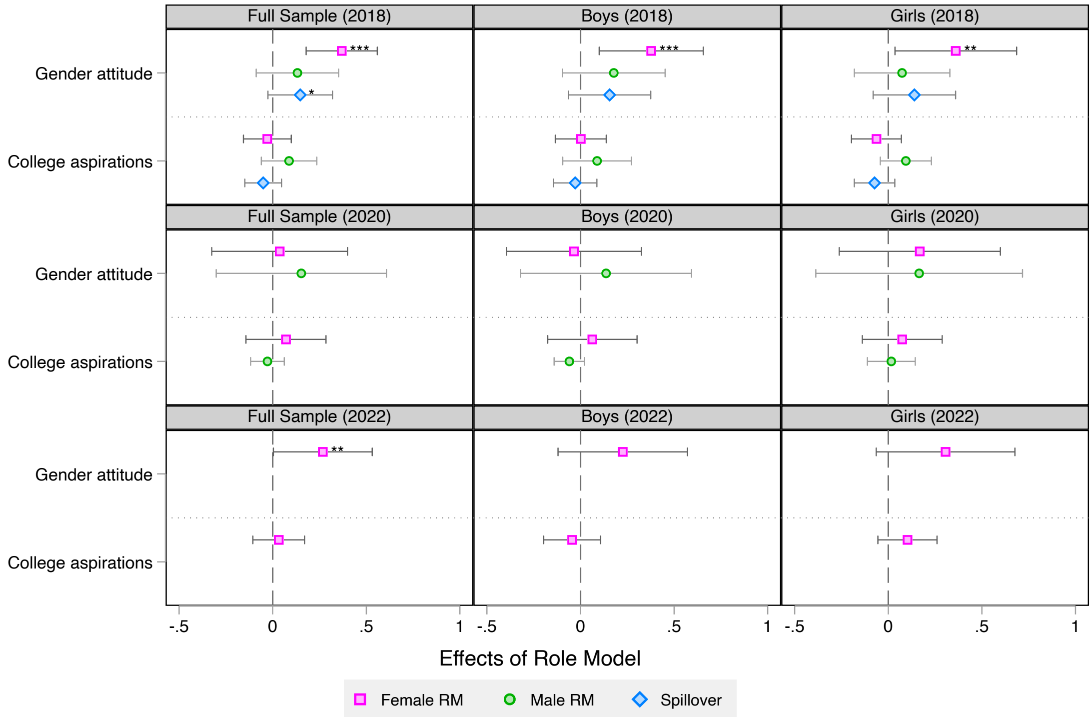
3.6 Conclusion

We evaluated a randomised role-model intervention comprising short classroom visits by male and female college students to primary schools in Somalia. Female college students visiting classrooms had a significant, albeit short-term, impact on the gender attitudes of both boys and girls. However, these effects gradually diminished over time as students' attitudes shifted toward greater equality. Longer-term results indicate a positive impact on girls' college aspirations, significant improvements in secondary school enrollment for both boys and girls, and a reduction in early marriage. Male role models did not significantly affect gender attitudes or aspirations for both boys and girls.

While the study's fragile context posed several challenges in following up with students over time and in recording their educational and post-graduation marital outcomes, the consistency of the findings across multiple outcomes and cohorts strengthens their credibility. Notably, the intervention was extremely low-cost yet produced large effects, comparable to those of other role model programs in education. Nevertheless, the findings highlight the importance of studying interventions in fragile contexts, where brief in-person interactions with relatable role models—who share similar disadvantaged backgrounds and openly recount their successes, setbacks, and strategies for overcoming challenges—can serve as a particularly powerful and cost-effective tool.

3.7 Figures

Figure 3.1: Role Model Impacts on Gender Attitudes and College Aspirations



Note: The figure displays the estimated coefficients and 95 per cent confidence intervals generated from the regression analyses reported in Table 3.3 in the Appendix. The gender attitude variable is standardised around the control mean and, therefore, expressed in standard deviations from that mean. The top panel of the figure shows estimates from the First Endline, which included students enrolled in grades 5-8 at the time of data collection in Fall 2018, and were likely in grades 4–7 at the time of the intervention in Spring 2018. The middle panel displays estimates obtained from the Second Endline for the 2020 graduating cohort. It therefore includes students who were in grade 8 in Spring 2020 and likely in grade 6 at the time of the intervention. The bottom panel reports estimates obtained from the Third Endline for the 2022 graduating cohort. It includes students who were in grade 8 in the Spring of 2022 and were therefore likely enrolled in grade 4 at the time of the intervention. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.8 Tables

Table 3.1: Baseline Balance

	Control	Treatment	Difference
	(1)	(2)	(3)
Charges school fees	0.348	0.348	0.000
Age of school	9.087	8.217	0.870
Has girl friendly space	0.174	0.174	0.000
Number of classrooms	6.522	8.913	-2.391
Number female teachers	1.652	1.826	-0.174
Total students enrolled	209.174	257.304	-48.130
Boys enrolled	106.130	139.870	-33.739
Girls enrolled	103.043	117.435	-14.391
Student-teacher ratio	30.500	34.493	-3.993
Number of grades	6.652	6.565	0.087
Observations	23	23	46

Note: Table shows school characteristics by treatment status using data originally obtained from Save the Children International before school randomization and implementation of the intervention. The data in this table refer to the entire universe of schools originally included in the study. For each characteristic listed in the rows, we report the averages observed in treatment and control schools, along with their differences. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.2: Student Characteristics by Endline Survey Wave (All)

	Female RM	Male RM	Spillover	P-value Female RM=Male RM	Control Mean	Observations
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Fall 2018						
Girl	-0.022 (0.033)	-0.093 (0.037)**	0.006 (0.031)	0.100	0.500	1,941
Student age	-0.083 (0.159)	0.017 (0.226)	0.105 (0.132)	0.685	14.636	1,912
Missed school	0.023 (0.061)	0.021 (0.046)	-0.018 (0.055)	0.981	0.741	1,941
Missed school due to poverty	-0.090 (0.041)**	0.001 (0.027)	-0.074 (0.036)**	0.051	0.139	1,941
Missed school due to chores	0.030 (0.024)	-0.038 (0.036)	-0.001 (0.023)	0.059	0.119	1,941
Missed school due to sickness	0.063 (0.062)	0.059 (0.042)	0.042 (0.052)	0.957	0.462	1,941
Days absent from school	-0.253 (0.472)	-0.362 (0.410)	-0.086 (0.379)	0.823	2.363	1,735
Lives with parent(s)	-0.003 (0.021)	0.032 (0.036)	0.005 (0.026)	0.321	0.882	1,941
Father or mother completed secondary	0.004 (0.040)	-0.057 (0.052)	0.058 (0.038)	0.270	0.242	1,700
Close female relative went to college	0.045 (0.030)	0.000 (0.022)	0.045 (0.019)**	0.167	0.057	1,941
Panel B: Spring 2020						
Girl	-0.134 (0.078)*	-0.083 (0.067)		0.498	0.471	829
Student age	-0.484 (0.448)	-0.292 (0.414)		0.737	15.321	829
Missed school	0.016 (0.095)	0.108 (0.064)		0.377	0.629	829
Missed school due to poverty	0.063 (0.106)	0.055 (0.120)		0.957	0.164	829
Missed school due to chores	0.038 (0.112)	0.079 (0.083)		0.740	0.402	829
Missed school due to sickness	-0.052 (0.068)	0.058 (0.075)		0.231	0.270	829
Days absent from school	-0.177 (0.484)	0.200 (0.275)		0.458	1.934	829
Close female relative went to college	-0.005 (0.083)	-0.002 (0.094)		0.978	0.253	829
Panel C: Spring 2022						
Girl	-0.051 (0.036)				0.514	575
Student age	-0.287 (0.272)				15.743	575
Missed school	0.051 (0.118)				0.605	575
Missed school due to poverty	0.071 (0.041)*				0.032	575
Missed school due to chores	0.043 (0.047)				0.130	575
Missed school due to sickness	-0.004 (0.088)				0.423	575
Days absent from school	0.492 (0.757)				2.170	575
Lives with parent(s)	0.033 (0.030)				0.874	575
Father or mother completed secondary	0.027 (0.089)				0.383	575
Close female relative went to college	-0.010 (0.055)				0.320	575

Note: Each row in each panel is a separate regression whereby the variable indicated in the row is the dependent variable, and treatment dummies are the key independent variables. For each regression, the omitted group is the control group. Regressions in Panel A also include baseline grade and randomisation strata fixed effects. Standard errors in Panel A are clustered at the school-grade level. Standard errors in Panels B and C are clustered at the school level because the analysis includes only a single grade per school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3: Short-term, medium-term and long-term impacts on gender attitudes and aspirations

	2018		2020		2022	
	Gender attitude	College aspirations	Gender attitude	College aspirations	Gender attitude	College aspirations
	(1)	(2)	(3)	(4)	(5)	(6)
Female RM	0.378	0.001	0.182	0.137	0.100	-0.032
	(0.140)***	(0.069)	(0.267)	(0.085)	(0.173)	(0.066)
	{0.016}	{0.994}	{0.578}	{0.242}	{0.586}	{0.702}
	[0.006]	[1.000]	[0.686]	[0.091]	[0.743]	[0.749]
Male RM	0.178	0.089	0.110	-0.026		
	(0.138)	(0.093)	(0.247)	(0.093)		
	{0.178}	{0.356}	{0.766}	{0.772}		
	[0.290]	[0.595]	[0.779]	[0.867]		
Spillover	0.156	-0.028				
	(0.111)	(0.059)				
	{0.206}	{0.646}				
	[0.278]	[0.918]				
Girl x Female RM	-0.017	-0.065	-0.043	-0.054	-0.024	0.153
	(0.236)	(0.046)	(0.245)	(0.103)	(0.224)	(0.061)**
	{0.958}	{0.134}	{0.894}	{0.624}	{0.970}	{0.020}
	[1.000]	[0.278]	[0.867]	[0.779]	[0.869]	[0.008]
Girl x Male RM	-0.104	0.005	-0.147	0.103		
	(0.152)	(0.073)	(0.219)	(0.067)		
	{0.538}	{0.890}	{0.538}	{0.158}		
	[0.814]	[1.000]	[0.687]	[0.104]		
Girl x Spillover	-0.016	-0.045				
	(0.138)	(0.055)				
	{0.950}	{0.382}				
	[1.000]	[0.728]				
Girl	0.271	0.027	0.163	0.011	0.330	-0.031
	(0.091)***	(0.030)	(0.154)	(0.045)	(0.150)**	(0.047)
Female RM + Girl × Female RM	0.361	-0.064	0.139	0.082	0.076	0.121
	(0.164)**	(0.067)	(0.210)	(0.128)	(0.208)	(0.066)*
	{0.034}	{0.366}	{0.552}	{0.556}	{0.738}	{0.110}
Male RM + Girl × Male RM	0.074	0.094	-0.037	0.077		
	(0.129)	(0.069)	(0.257)	(0.097)		
	{0.620}	{0.192}	{0.888}	{0.470}		
Spillover + Girl × Spillover	0.139	-0.074				
	(0.111)	(0.055)				
	{0.206}	{0.158}				
Grade fixed effects	Yes	Yes	No	No	No	No
Pair fixed effects	Yes	Yes	No	No	No	No
School controls	No	No	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.000	0.707	-0.000	0.693	0.000	0.431
Observations	1,941	1,941	829	829	575	575
Clusters	117	117	32	32	28	28

Note: The First Endline data generate estimates in columns 1 and 2 and include students who were enrolled in grades 5 to 8 at the time of the survey (Fall 2018), i.e., in grades 4-7 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated using the Second Endline data, which includes the 2020 graduating cohort, i.e., students who were likely enrolled in grade 6 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated using the Second Endline data, which includes the 2022 graduating cohort, i.e., students who were likely enrolled in grade 4 at the time of the intervention, in Spring 2018. The gender attitude index averages four survey questions and is standardised around the control mean. Therefore, the associated estimates are expressed in standard deviations from the control mean. College aspirations are measured by a dummy variable that equals 1 if the highest level of education a student would like to complete, if nothing could stop them, is college. Robust standard errors in parentheses are clustered at the grade-school level (Column 1- 2) and the school level (Column 3- 6). Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We pool the three types of treatment groups into one category in Table 3.B.6; we do not distinguish by the type of role model who visited them.

Table 3.4: Role Models' Impacts on Exit Exam, School Enrollment and Marriage (Treatment vs. Control)

	(1) Sat exam	(2) Exam score	(3) Enrolled secondary	(4) Married or has children
Treatment school	0.027	0.246	0.063	-0.026
	(0.036)	(0.209)	(0.039)	(0.017)
	{0.494}	{0.334}	{0.170}	{0.120}
	[0.558]	[0.211]	[0.093]	[0.105]
Girl × Treatment	-0.055	-0.075	0.036	-0.030
	(0.061)	(0.183)	(0.056)	(0.045)
	{0.398}	{0.686}	{0.604}	{0.594}
	[0.469]	[0.558]	[0.558]	[0.558]
Girl	0.002	0.074	-0.034	0.053
	(0.046)	(0.147)	(0.057)	(0.046)
Baseline grade==4	0.220	-0.003	0.091	-0.070
	(0.037)***	(0.175)	(0.035)**	(0.031)**
Treatment + Girl × Treatment	-0.027	0.170	0.099	-0.056
	(0.052)	(0.203)	(0.051)*	(0.047)
	{0.604}	{0.476}	{0.076}	{0.310}
Pair FE	No	No	No	No
School controls	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Control mean	0.787	0.006	0.877	0.077
Observations	1,262	984	984	984
Clusters	29	29	29	29

Note: The analysis includes the graduating cohorts of 2020 and 2022, who were likely enrolled in grades 6 and 4, respectively, at the time of the intervention in Spring 2018. For these cohorts, we have exit exam data and teacher-reported information on secondary school enrolment and marital status. *Sat exam* is an indicator equal to 1 if the student took the grade 8 standardised exit exam, and 0 otherwise. *Exam score* refers to the grade achieved on the exit exam, conditional on having taken it. Exam scores are standardised around the cohort's control mean and are expressed in standard deviations. *Enrolled secondary* is a 0-1 dummy variable equal to 1 if the student is enrolled in secondary school, as reported by the corresponding grade 8 class teacher. Similarly, *married or has children* is a 0-1 dummy equal to 1 if the student is married or has children, according to their grade 8 teacher. Estimates for the full 2020 graduating class are presented in the Appendix. Robust standard errors in parentheses are clustered at the school level. Wild cluster bootstrap p-values are shown in curly brackets. Romano-Wolf corrected p-values are in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 3.A Appendix Figures

Figure 3.A.1: Map Showing Study Schools

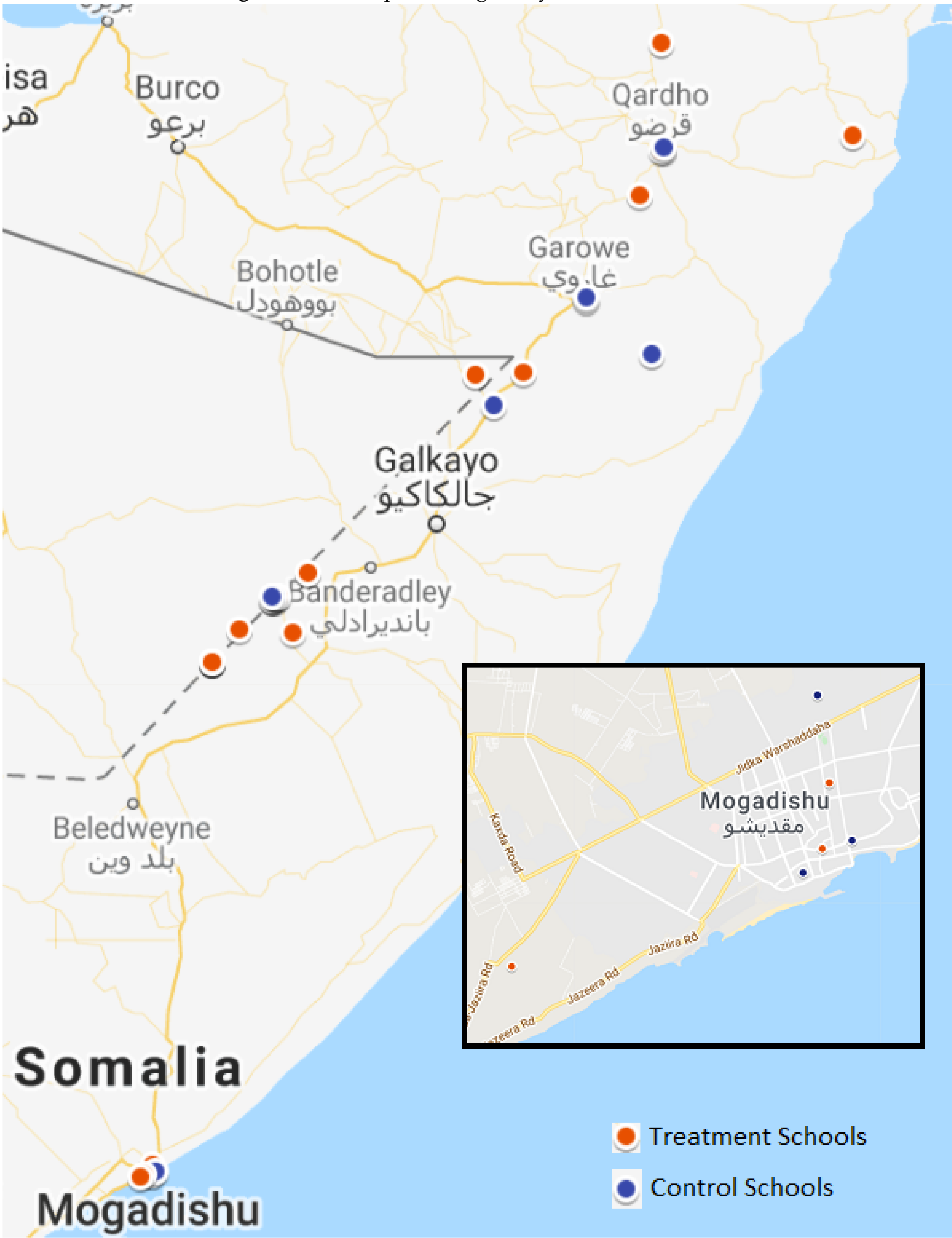
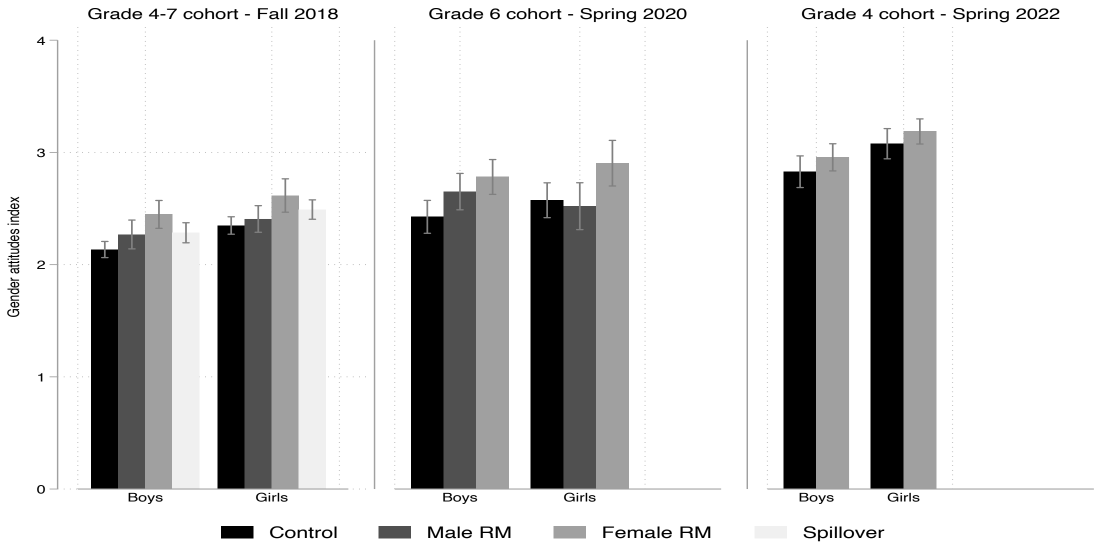
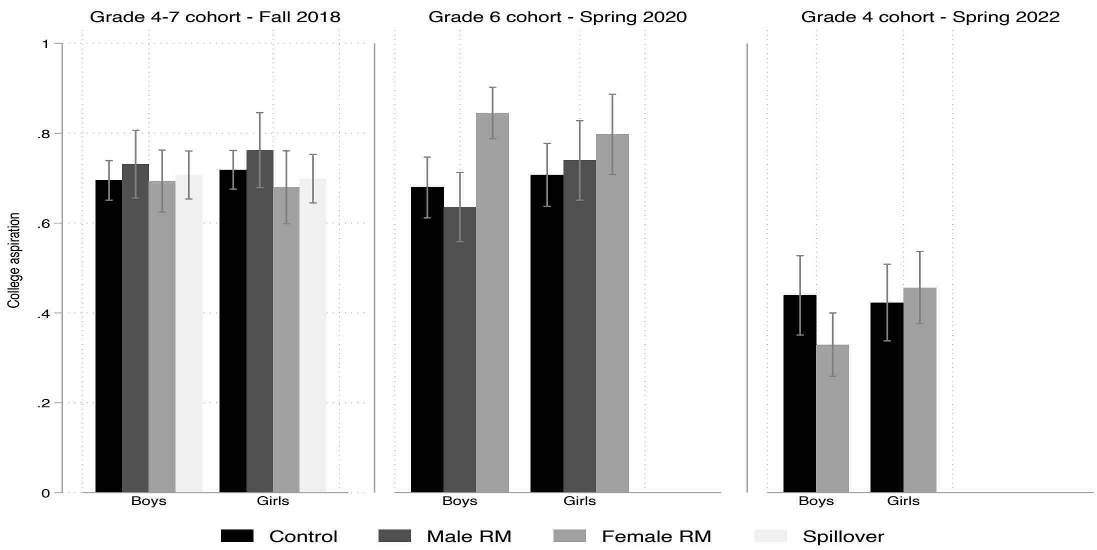


Figure 3.A.2: College Aspirations and Gender Attitudes by Treatment

(a) Gender Attitudes



(b) College Aspiration



Note: The figure reports the outcome variables, by student gender and treatment group, as measured in three separate survey waves. In the left panel, outcomes were measured in Fall 2018, 6 months after the intervention (Spring 2018), among students enrolled in grades 4–7 in Spring 2018. The outcomes measured in the Spring of 2020 among students who were enrolled in grade 6 in the Spring of 2018 are presented in the middle panel, while outcomes measured in the Spring of 2022 are shown in the right-hand panel. The gender attitude index in the top panel is an aggregate of four survey questions, ranging from 1 to 4, with higher numbers indicating a preference for greater gender equality. College Aspirations in the bottom panel refer to the percentage of students who stated that, if nothing were to stop them, their highest level of education would be college.

Appendix 3.B Appendix Tables

Table 3.B.1: School Characteristics in all Survey Rounds

	Fall 2018			Spring 2020			Spring 2022		
	Control	Treatment	Difference	Control	Treatment	Difference	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Charges school fees	0.368	0.364	0.005	0.438	0.500	-0.062	0.438	0.500	-0.062
Age of school	9.316	8.455	0.861	9.812	9.562	0.250	9.812	10.833	-1.021
Has girl friendly space	0.211	0.136	0.074	0.188	0.125	0.062	0.188	0.000	0.188
Number of classrooms	7.263	8.364	-1.100	7.125	9.125	-2.000	7.125	9.417	-2.292
Number female teachers	1.474	1.773	-0.299	1.438	1.312	0.125	1.438	1.417	0.021
Total students enrolled	209.158	263.773	-54.615	233.375	278.688	-45.312	233.375	297.333	-63.958
Boys enrolled	106.947	143.545	-36.598	118.125	152.125	-34.000	118.125	161.917	-43.792
Girls enrolled	102.211	120.227	-18.017	115.250	126.562	-11.312	115.250	135.417	-20.167
Student-teacher ratio	28.709	35.770	-7.061	31.977	33.568	-1.591	31.977	33.579	-1.602
Number of grades	6.947	6.682	0.266	7.625	7.500	0.125	7.625	7.500	0.125
Observations	19	22	41	16	16	32	16	12	28

Note: Table shows school characteristics by treatment status using data originally obtained from Save the Children International before school randomization and implementation of the intervention. In this panel, for each follow-up survey wave, we restrict the original baseline data to only schools covered in the corresponding follow-up survey. In the first three columns, the data is restricted to the schools included in the First Endline. Data in columns 4-6 are for the schools included in the Second Endline, while columns 7-9 are for the schools included in the Third Endline. For each data collection wave, we report the averages observed in treatment and control schools, along with the differences between them. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. For additional balance tests for students in classes assigned to male versus female role models within the treatment group, see Tables 3.2 and 3.B.3 in the Appendix.

Table 3.B.2: Attrition Tests

	Spring 2020			Spring 2022		
	Not Attrited	Attrited	Difference	Not Attrited	Attrited	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Student Survey						
Treatment school	0.500	0.500	0.000	0.429	0.611	-0.183
Charges school fees	0.469	0.071	0.397***	0.464	0.167	0.298**
Age of school	9.688	6.286	3.402*	10.250	6.167	4.083**
Has girl friendly space	0.156	0.214	-0.058	0.107	0.278	-0.171
Number of classrooms	8.125	6.786	1.339	8.107	7.111	0.996
Number female teachers	1.375	2.571	-1.196*	1.429	2.222	-0.794
Total students enrolled	256.031	181.143	74.888*	260.786	190.389	70.397
Boys enrolled	135.125	95.286	39.839	136.893	101.389	35.504
Girls enrolled	120.906	85.857	35.049*	123.893	89.000	34.893*
Student-teacher ratio	32.772	31.865	0.908	32.664	32.236	0.428
Number of grades	7.562	4.429	3.134***	7.571	5.111	2.460***
Observations	32	14	46	28	18	46
Panel B: Exit Exam and Teacher Survey						
Treatment school	0.444	0.579	-0.135	0.444	0.579	-0.135
Charges school fees	0.519	0.105	0.413***	0.481	0.158	0.324**
Age of school	10.519	6.000	4.519***	10.556	5.947	4.608***
Has girl friendly space	0.148	0.211	-0.062	0.074	0.316	-0.242**
Number of classrooms	7.519	8.000	-0.481	7.963	7.368	0.595
Number female teachers	1.333	2.316	-0.982	1.296	2.368	-1.072
Total students enrolled	257.481	198.789	58.692	260.593	194.368	66.224
Boys enrolled	136.148	104.316	31.832	137.407	102.526	34.881
Girls enrolled	121.333	94.474	26.860	123.185	91.842	31.343*
Student-teacher ratio	32.291	32.788	-0.498	33.170	31.539	1.631
Number of grades	7.556	5.263	2.292***	7.593	5.211	2.382***
Observations	27	19	46	27	19	46

Note: We compare school characteristics in the schools for which we have data in the Second and Third Endline (*Not Attrited* in the table) and those for which we have no follow-up data (*Attrited*). In Panel A, attrition is defined based on the existence of grade 8 in the school since the Second and Third Endline surveys target students graduating from primary schools. We collected data from all 32 schools that offered eight grades in 2020 and 2022. The Not Attrited sample fell to 28 in 2022, as we lost four schools from which we received only very few surveys back, as detailed in Section 3.3.2.3. In Panel B, attrition is defined based on the availability of exam data and teacher-generated outcomes for the graduating students surveyed. Here, the sample is further reduced due to the unavailability of the teachers who taught the graduating classes of 2020 or 2022. For more details, see the Section 3.3.2.3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.B.3: Student Characteristics: Sample with Exam Data

	Treatment school	Control Mean	Observations
	(1)	(2)	(3)
Panel B: Spring 2020			
Girl	-0.189 (0.076)**	0.460	473
Student age	-0.590 (0.417)	15.286	473
Missed school	0.058 (0.084)	0.591	473
Missed school due to poverty	0.086 (0.115)	0.186	473
Missed school due to chores	0.103 (0.084)	0.371	473
Missed school due to sickness	-0.075 (0.058)	0.253	473
Days absent from school	-0.162 (0.341)	1.814	473
Close female relative went to college	0.035 (0.087)	0.241	473
Panel C: Spring 2022			
Girl	-0.064 (0.039)	0.518	511
Student age	-0.224 (0.304)	15.728	511
Missed school	0.072 (0.123)	0.573	511
Missed school due to poverty	0.073 (0.040)*	0.023	511
Missed school due to chores	0.036 (0.050)	0.138	511
Missed school due to sickness	0.009 (0.089)	0.404	511
Days absent from school	0.515 (0.755)	2.028	511
Lives with parent(s)	0.033 (0.033)	0.872	511
Father or mother completed secondary	0.015 (0.093)	0.408	511
Close female relative went to college	-0.016 (0.063)	0.317	511

Note: The analysis is restricted to surveyed students from schools for which we have exit exam results and teacher survey responses. These students constitute our working sample for evaluating the role model intervention. Each row in each panel represents a separate regression, with the variable indicated in the row serving as the dependent variable and the treatment dummy as the key independent variable. For each regression, the omitted group is the control group. Standard errors for each regression are clustered at the school level because the analysis includes only a single grade per school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.4: Randomization Inference: Short-term, medium-term and long-term impacts on gender attitudes and aspirations

	2018		2020		2022	
	Gender attitude	College aspirations	Gender attitude	College aspirations	Gender attitude	College aspirations
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Boys						
Female RM	0.519	0.011	0.342	0.163	0.129	-0.105
	(0.119)	(0.077)	(0.373)	(0.101)	(0.190)	(0.079)
	[0.001, 0.009]	[0.872, 0.911]	[0.422, 0.484]	[0.199, 0.252]	[0.504, 0.566]	[0.199, 0.251]
Male RM	0.059	0.043	0.179	-0.049		
	(0.125)	(0.109)	(0.276)	(0.116)		
	[0.661, 0.720]	[0.698, 0.754]	[0.580, 0.641]	[0.716, 0.771]		
Panel B: Girls						
Female RM	0.282	-0.195	0.279	0.061	0.153	0.045
	(0.133)	(0.065)	(0.242)	(0.117)	(0.231)	(0.072)
	[0.145, 0.193]	[0.039, 0.068]	[0.326, 0.387]	[0.605, 0.666]	[0.490, 0.552]	[0.515, 0.577]
Male RM	0.055	0.173	-0.037	0.034		
	(0.148)	(0.072)	(0.273)	(0.111)		
	[0.690, 0.747]	[0.033, 0.060]	[0.860, 0.901]	[0.763, 0.815]		

Note: The table displays estimates obtained from randomisation inference, using the *ritest* command in Stata (Heß, 2017). The analysis is conducted separately for boys and girls. The estimates in columns 1 and 2 are generated from the First Endline, which targeted students who were enrolled in grades 5 to 8 at the time of the survey (Fall 2018), hence in grades 4–7 at the time of the intervention in Spring 2018. The estimates in columns 3 and 4 are generated by the Second Endline, which targeted the 2020 graduating cohort, i.e., students likely enrolled in grade 6 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated from the Third Endline, which targeted the 2022 graduating cohort, i.e., students likely enrolled in grade 4 at the time of the intervention in Spring 2018. All regressions include student-level controls. Standard errors in parentheses are clustered at the grade–school level in Columns 1 and 2, and at the school level in Columns 3 to 6, since the analysis in this case only includes a single grade per school. In square brackets, we report the 95% confidence interval of the p-value for the treatment effect coefficient based on randomisation inference.

Table 3.B.5: Attrition Sensitivity of Short-term Impacts

	All schools		Schools not attrited in 2020		Schools not attrited in 2022	
	Gender attitude	College aspirations	Gender attitude	College aspirations	Gender attitude	College aspirations
	(1)	(2)	(3)	(4)	(5)	(6)
Female RM	0.378	0.001	0.378	0.093	0.191	0.037
	(0.140)***	(0.069)	(0.148)**	(0.061)	(0.192)	(0.072)
	{0.016}	{0.994}	{0.030}	{0.156}	{0.448}	{0.614}
	[0.003]	[1.000]	[0.011]	[0.130]	[0.654]	[0.885]
Male RM	0.178	0.089	0.323	0.122	0.308	0.077
	(0.138)	(0.093)	(0.139)**	(0.105)	(0.145)**	(0.108)
	{0.178}	{0.356}	{0.022}	{0.268}	{0.034}	{0.566}
	[0.291]	[0.590]	[0.020]	[0.308]	[0.061]	[0.750]
Spillover	0.156	-0.028	0.197	0.002	0.229	-0.070
	(0.111)	(0.059)	(0.122)	(0.065)	(0.135)*	(0.075)
	{0.206}	{0.646}	{0.110}	{0.954}	{0.110}	{0.384}
	[0.285]	[0.913]	[0.130]	[0.977]	[0.182]	[0.667]
Girl x Female RM	-0.017	-0.065	-0.027	-0.080	0.116	-0.039
	(0.236)	(0.046)	(0.272)	(0.048)*	(0.389)	(0.048)
	{0.958}	{0.134}	{0.930}	{0.094}	{0.746}	{0.438}
	[1.000]	[0.285]	[0.977]	[0.130]	[0.885]	[0.750]
Girl x Male RM	-0.104	0.005	-0.203	-0.050	-0.217	-0.032
	(0.152)	(0.073)	(0.174)	(0.077)	(0.179)	(0.079)
	{0.538}	{0.890}	{0.308}	{0.560}	{0.276}	{0.764}
	[0.840]	[1.000]	[0.308]	[0.791]	[0.453]	[0.885]
Girl x Spillover	-0.016	-0.045	-0.092	-0.035	-0.219	0.006
	(0.138)	(0.055)	(0.153)	(0.061)	(0.142)	(0.065)
	{0.950}	{0.382}	{0.530}	{0.626}	{0.144}	{0.976}
	[1.000]	[0.735]	[0.791]	[0.791]	[0.249]	[0.898]
Girl	0.271	0.027	0.338	0.036	0.347	0.028
	(0.091)***	(0.030)	(0.105)***	(0.034)	(0.105)***	(0.033)
Female RM + Girl × Female RM	0.361	-0.064	0.351	0.012	0.306	-0.001
	(0.164)**	(0.067)	(0.195)*	(0.065)	(0.271)	(0.079)
	{0.034}	{0.366}	{0.084}	{0.814}	{0.324}	{0.994}
Male RM + Girl × Male RM	0.074	0.094	0.120	0.072	0.091	0.045
	(0.129)	(0.069)	(0.121)	(0.072)	(0.124)	(0.077)
	{0.620}	{0.192}	{0.356}	{0.364}	{0.502}	{0.582}
Spillover + Girl × Spillover	0.139	-0.074	0.105	-0.033	0.010	-0.064
	(0.111)	(0.055)	(0.126)	(0.062)	(0.128)	(0.069)
	{0.206}	{0.158}	{0.432}	{0.614}	{0.912}	{0.352}
Grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
School controls	No	No	No	No	No	No
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.000	0.707	0.074	0.708	0.074	0.708
Observations	1,941	1,941	1,643	1,643	1,450	1,450
Clusters	117	117	96	96	83	83

Note: All analyses in this table correspond to data collected during the First Endline Survey. Our models in columns 1 and 2 include all surveyed students who were enrolled in grades 5 to 8 at the time of the survey (Fall 2018), hence in grades 4-7 at the time of the intervention, in Spring 2018. In subsequent panels, we restrict the First Endline data to schools covered in subsequent endline surveys. Specifically, the data in columns 3 and 4 is restricted to a sub-sample of schools that were also covered in the Second Endline. On the other hand, the data in columns 5 and 6 is restricted to a sub-sample of schools that were also covered in the Third Endline. The gender attitude index averages four survey questions and is standardised around the control mean. Therefore, the associated estimates are expressed in standard deviations from the control mean. College aspirations are measured by a dummy variable that equals 1 if the highest level of education a student would like to complete, if nothing could stop them, is college. Robust standard errors in parentheses are clustered at the grade-school level. Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.6: Impacts on Gender Attitudes and Aspirations (Treatment vs. Control Schools)

	2018		2020		2022	
	Gender attitude	College aspirations	Gender attitude	College aspirations	Gender attitude	College aspirations
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment school	0.213	0.007	0.144	0.056	0.100	-0.032
	(0.088)**	(0.047)	(0.216)	(0.075)	(0.173)	(0.066)
	{0.018}	{0.860}	{0.512}	{0.542}	{0.586}	{0.702}
	[0.004]	[0.821]	[0.549]	[0.517]	[0.743]	[0.749]
Girl × Treatment	-0.042	-0.040	-0.100	0.026	-0.024	0.153
	(0.123)	(0.042)	(0.200)	(0.066)	(0.224)	(0.061)**
	{0.734}	{0.330}	{0.720}	{0.714}	{0.970}	{0.020}
	[0.821]	[0.321]	[0.701]	[0.701]	[0.869]	[0.008]
Girl	0.275	0.025	0.162	0.010	0.330	-0.031
	(0.091)***	(0.030)	(0.153)	(0.045)	(0.150)**	(0.047)
Treatment + Girl × Treatment	0.171	-0.033	0.044	0.083	0.076	0.121
	(0.093)*	(0.044)	(0.195)	(0.085)	(0.208)	(0.066)*
	{0.072}	{0.464}	{0.826}	{0.408}	{0.738}	{0.110}
Grade fixed effects	Yes	Yes	No	No	No	No
Pair fixed effects	Yes	Yes	No	No	No	No
School controls	No	No	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.000	0.707	-0.000	0.693	0.000	0.431
Observations	1,941	1,941	829	829	575	575
Clusters	117	117	32	32	28	28

Note: Unlike Table 3.3, we pool all students in the treatment school as one category, irrespective of the type of role model who visited them. The First Endline data generate estimates in columns 1 and 2 and include students who were enrolled in grades 5 to 8 at the time of the survey (Fall 2018), i.e., in grades 4–7 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated using the Second Endline data, which includes the 2020 graduating cohort, i.e., students who were likely enrolled in grade 6 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated using the Second Endline data, which includes the 2022 graduating cohort, i.e., students who were likely enrolled in grade 4 at the time of the intervention, in Spring 2018. The gender attitude index averages four survey questions and is standardised around the control mean. Therefore, the associated estimates are expressed in standard deviations from the control mean. College aspirations are measured by a dummy variable that equals 1 if the highest level of education a student would like to complete, if nothing could stop them, is college. Robust standard errors in parentheses are clustered at the grade–school level (Column 1–2) and at the school level (Column 3–6). Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.7: Female Role Models Impacts on Exit exam, School Enrollment and Marriage (Female Role Models vs. Control)

	Sat exam	Exam score	Enrolled secondary	Married or has children
	(1)	(2)	(3)	(4)
Female RM	0.032	0.235	0.046	-0.013
	(0.044)	(0.260)	(0.044)	(0.019)
	{0.572}	{0.518}	{0.396}	{0.482}
	[0.635]	[0.483]	[0.353]	[0.635]
Girl x Female RM	-0.051	-0.002	0.029	-0.022
	(0.062)	(0.193)	(0.058)	(0.047)
	{0.458}	{0.994}	{0.698}	{0.730}
	[0.572]	[0.979]	[0.702]	[0.702]
Girl	0.003	0.081	-0.035	0.054
	(0.048)	(0.149)	(0.057)	(0.047)
Baseline grade==4	0.223	0.042	0.117	-0.088
	(0.043)***	(0.184)	(0.039)***	(0.035)**
Female RM + Girl × Female RM	-0.019	0.233	0.076	-0.035
	(0.055)	(0.232)	(0.049)	(0.043)
	{0.766}	{0.394}	{0.172}	{0.512}
Pair fixed effects	No	No	No	No
School controls	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Control mean	0.787	0.006	0.877	0.077
Observations	1,071	859	859	859
Clusters	29	29	29	29

Note: The analysis includes the graduating cohorts of 2020 and 2022, who were likely enrolled in grades 6 and 4, respectively, at the time of the intervention in Spring 2018 and for whom we have exit exam data and teacher-generated data on secondary school enrolment and marital status. *Sat exam* is an indicator equal to 1 if the student took the grade 8 standardized exit exam, and 0 otherwise. *Exam score* refers to the grade obtained in the exit exam, conditional on having taken it. Exam scores are standardised around the cohort's control mean and therefore expressed in standard deviations. *Enrolled secondary* is a 0-1 dummy variable equal to 1 if the student is enrolled to secondary school as reported by the corresponding grade 8 class teacher. Similarly, *married or has children* is a 0-1 dummy equal to 1 if the student is married or has had children, according to his/her grade 8 teacher. Since we cannot test for male role model effects for the 2022 graduating cohort (baseline grade 4) due to insufficient data, this analysis excludes the 2020 graduating students who were assigned to male role models at the time of the intervention. Robust standard errors in parentheses are clustered at the school level. Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.8: Randomisation Inference: Exit Exam, School Enrollment and Marriage (Treatment vs. Control)

	Sat exam	Exam score	Enrolled secondary	Married or has children
	(1)	(2)	(3)	(4)
Panel A: Boys				
Treatment school	0.023	0.282	0.076	-0.039
	(0.039)	(0.219)	(0.041)	(0.020)
	[0.521, 0.583]	[0.164, 0.214]	[0.064, 0.099]	[0.120, 0.164]
Panel B: Girls				
Treatment school	-0.027	0.119	0.070	-0.027
	(0.051)	(0.201)	(0.034)	(0.033)
	[0.653, 0.712]	[0.584, 0.645]	[0.476, 0.538]	[0.874, 0.913]

Note: The table shows estimates obtained from randomisation inference, using the *ritest* command in Stata (Heß, 2017). The analysis is conducted separately for boys and girls. The analysis utilises data generated from the Second and Third Endlines, which followed the 2020 and 2022 graduating cohorts of students (baseline grades 6 and 4, respectively). All regressions include grade fixed effects, school-level and student-level controls. Standard errors in parentheses are clustered at the school level. In square brackets, we report a 95% confidence interval of the p-value for the treatment effect coefficient based on randomisation inference.

Table 3.B.9: Role Models' Impacts on Exit Exam, School Enrollment and Marriage (Treatment vs. Control with Pair Fixed Effects)

	(1) Sat exam	(2) Exam score	(3) Enrolled secondary	(4) Married or has children
Treatment school	0.044	0.474	0.135	-0.072
	(0.036)	(0.222)**	(0.045)***	(0.036)*
	{0.244}	{0.010}	{0.008}	{0.018}
	[0.283]	[0.018]	[0.003]	[0.024]
Girl × Treatment	-0.034	-0.210	-0.003	-0.011
	(0.068)	(0.176)	(0.055)	(0.041)
	{0.660}	{0.288}	{0.964}	{0.774}
	[0.781]	[0.283]	[0.917]	[0.844]
Girl	0.003	0.087	-0.003	0.036
	(0.056)	(0.179)	(0.053)	(0.037)
Baseline grade==4	0.211	0.028	0.106	-0.071
	(0.042)***	(0.196)	(0.035)***	(0.029)**
Treatment + Girl × Treatment	0.010	0.263	0.132	-0.083
	(0.047)	(0.123)**	(0.065)*	(0.063)
	{0.808}	{0.052}	{0.052}	{0.254}
Pair FE	Yes	Yes	Yes	Yes
School controls	No	No	No	No
Student controls	Yes	Yes	Yes	Yes
Control mean	0.758	-0.058	0.850	0.092
Observations	1,044	793	793	793
Clusters	24	24	24	24

Note: The analysis includes a subsample of cohorts graduating in 2020 and 2022, who were likely enrolled in grades 6 and 4, respectively, at the time of the intervention in Spring 2018. For these cohorts, we restrict the analysis to students whose tertiary outcomes (exit exam, secondary school enrolment, and marital status) are available for both schools in a randomisation pair. *Sat exam* is an indicator equal to 1 if the student took the grade 8 standardised exit exam, and 0 otherwise. *Exam score* refers to the grade achieved on the exit exam, conditional on having taken it. Exam scores are standardised around the cohort's control mean and are expressed in standard deviations. *Enrolled secondary* is a 0-1 dummy variable equal to 1 if the student is enrolled in secondary school, as reported by the corresponding grade 8 class teacher. Similarly, *married or has children* is a 0-1 dummy equal to 1 if the student is married or has children, according to their grade 8 teacher. Regressions in this table include randomisation strata fixed effects. Robust standard errors in parentheses are clustered at the school level. Wild cluster bootstrap p-values are shown in curly brackets. Romano-Wolf corrected p-values are in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.10: Effects of Role Models on Projected Education Trajectories

	Will attend secondary school	Will attend college or university
	(1)	(2)
Female RM	0.000	-0.009
	(0.001)	(0.053)
	{0.876}	{0.916}
	[0.921]	[0.921]
Male RM	-0.010	0.024
	(0.010)	(0.043)
	{0.556}	{0.614}
	[0.545]	[0.715]
Girl x Female RM	0.006	-0.069
	(0.006)	(0.040)*
	{0.436}	{0.104}
	[0.545]	[0.479]
Girl x Male RM	0.017	-0.059
	(0.012)	(0.049)
	{0.210}	{0.330}
	[0.499]	[0.545]
Girl	-0.006	0.044
	(0.006)	(0.036)
Female RM + Girl × Female RM	0.007	-0.079
	(0.007)	(0.052)
	{0.580}	{0.188}
Male RM + Girl × Male RM	0.007	-0.036
	(0.005)	(0.047)
	{0.298}	{0.656}
Grade fixed effects	No	No
Pair fixed effects	No	No
School controls	Yes	Yes
Student controls	No	No
Control mean	0.997	0.953
Observations	659	659
Clusters	32	32

Note: The estimates in this table are derived from grade 8 teacher interviews conducted during the Second Endline data collection. For each teacher surveyed, we asked them to estimate whether each student in their class would transition to a) secondary school and b) college or university. Consequently, Column 1 is a dummy variable equal to 1 if the teacher anticipates a specific student will move on to secondary school. Similarly, Column 2 is a dummy indicating whether the teacher expects the student to progress to college or university. We differentiate between students exposed to male or female role models and examine gender-specific effects in teachers' projections. However, we do not control for students' characteristics beyond gender, as these data were not collected in the teacher surveys nor reliably obtained from student surveys. Robust standard errors in parentheses are clustered at the school level, with wild cluster bootstrap p-values in curly brackets and Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.B.11: Teacher's Gender Attitudes

	Control	Treatment	Difference
	(1)	(2)	(3)
Sons should be encouraged than daughters to go to college (Reversed)	3.562	3.125	0.438
It is more important for boys than girls to do well in school (Reversed)	3.375	3.062	0.312
Boys are better leaders than girls (Reversed)	3.125	3.188	-0.062
Girls should be concerned with becoming good wives rather than a career (Reverse	3.500	3.312	0.188
Teacher gender attitudes	3.391	3.172	0.219
Observations	16	16	32

Note: The table shows teachers' gender attitudes by treatment status, based on data from the teacher survey conducted as part of the Second Endline survey. The data here are responses from Grade 8 teachers, one from each school, who helped link students from the graduating class with their exit exam grades from school records. Each statement, listed in a row, uses a reversed Likert scale from 1 to 4, with higher numbers indicating a preference for greater gender equality. The gender attitude index averages these four reversed statements, with higher scores indicating a stronger preference for gender equality. For each statement in the rows, we report the average teacher scores in treatment and control schools, along with the difference between them. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.B.12: Review of Selected Role Model Interventions

Country (Study)	Intervention	Study Design	Findings
Somalia – This paper	One-off 40–60 min classroom talk by male/female college student; narrative + Q&A.	School–grade RCT, Grades 2–8.	Female role models raise gender attitudes, improve girls' secondary transition; boys and male role models show less robust gains
Uganda – Riley (2024)	2-hour screening of a relatable female role-model film, *Queen of Katwe*, vs placebo.	Individual RCT in lower- and upper-secondary.	Math failure drops 27%→15%; girls gain 0.20 SD (0.33 SD low-ability); upper-secondary persistence increases by 11 p.p.
Madagascar – Nguyen (2008)	45–60 min information session; some arms include ex-student role models of poor/rich background.	School-level RCT in rural primaries.	Info + poor-background role model increases test scores (0.2 SD; up to 0.37 SD for pessimistic students) and modestly raises attendance.
Peru – Agurto et al. (2023)	20-minute presentation by female engineering students + brief Q&A.	Cluster RCT in Grade 11.	Small overall effects; high-ability girls increase engineering preference by 9 p.p. (45%) and show confidence gains.
China (urban) – Di (2024)	30–40 min talk with varying messages (study strategies, aspirations, or both).	Multi-arm class-level RCT in middle schools.	A combined message improves girls' scores (0.07–0.18 SD) but increases stress; weaker effects in boys.
China (rural) – Golan and You (2021)	One 40–60 min visit by a successful returnee/migrant from a similar background.	Cluster RCT in rural primary & junior-secondary.	Raises aspirations (+0.12–0.20 SD) and test scores (+0.06–0.10 SD); largest gains for low-aspiring students.

Notes: SD = Standard deviation. Q&A = Question and answer. p.p. = percentage points

Chapter 4

Returns to Socio-emotional Skills Training in the Labour Market: Evidence from Youth in Tanzania[1]

with Rachel Cassidy, Smita Das, Clara Delavallade and Munshi Sulaiman

4.1 Introduction

There is growing recognition of the importance of socio-emotional skills (SES) – also known as soft skills, life skills, or non-cognitive skills. These skills complement technical skills and are transferable across contexts, yielding substantial returns in the labour market (Heckman, Stixrud, and Urzua, 2006; Heckman and Kautz, 2012; Deming, 2017). As a result, many upskilling programs are increasingly incorporating SES, and policymakers now recognise them as foundational skills, alongside literacy and numeracy (Cunningham and Villaseñor, 2016; Morsy and Mukasa, 2020). This growing demand raises key policy questions: can SES be effectively taught, and if so, which skills should be prioritised for training? Existing empirical studies have typically focused on a single skill or narrow set of skills such as negotiation (Ashraf et al., 2020) or personal initiative (Campos et al., 2017; Ubfal et al., 2022) – or embedded more holistic SES training within labour market interventions with other components, making it difficult to estimate the contributions of trainings on whole sets of SES (Groh et al., 2016; Prada, Rucci, and Urzúa, 2019; Acevedo et al., 2020; Chioda et al., 2021; Barrera-Osorio, Kugler, and Silliman, 2023; Brudevold-Newman and Ubfal, 2024; Dammert and Nansamba,

[1] This work was funded by the International Development Research Centre (IDRC), the Wellspring Philanthropic Fund and the World Bank Umbrella Fund for Gender Equality (UFGE). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organisations, or those of the Executive Directors of the World Bank or the governments they represent. AEA RCT Registry ID: AEARCTR-0005548. The Tanzania Commission for Science and Technology approved this research project (Research Permit Number - 2019-381-NA-2019-270.)

2023). Given the high cost of training and the wide array of potentially relevant skills, it is essential to identify those skill sets that are both malleable and likely to yield labour market improvements.

Theoretically, SES can influence economic outcomes through multiple pathways. First, an individual's *awareness skills* — namely, their ability to reflect on their own strengths, weaknesses, and goals, as well as those of others — inform how they pursue opportunities. Setbacks can also foster productive introspection, helping individuals adjust their strategies for better outcomes (Tamir and Ford, 2009; Kreemers, Hooft, and Vianen, 2022). Second, in contexts with limited job prospects, job seekers also require intrapersonal and interpersonal *management skills*. They must not only plan and persevere, but also manage the emotional toll of repeated rejections (Berger et al., 2022). Grit is widely acknowledged as crucial for long-term success (e.g., Alan, Boneva, and Ertac, 2019). A proactive mindset enables individuals to recognise and seize opportunities as they arise (see Campos et al., 2017). Many goals are pursued in interpersonal contexts, making communication and collaboration vital to success (Borghans et al., 2008). When institutional constraints and resource scarcity prevail, individuals with stronger interpersonal skills, such as negotiation and persuasion, may be better equipped to navigate or challenge such barriers (Ashraf et al., 2020).

To unpack SES, we employ the ESTEEM framework, which spans 14 skills across two domains: four awareness-related and 10 management-related.[2] We test the effects of training on awareness skills, management skills, or both, on SES, perceived skill levels, labour market outcomes, and well-being of young men and women in three-month and one-year post-intervention. For this purpose, we designed two novel SES curricula and implemented a randomised controlled trial involving 4,728 NEET (not in education, employment, or training) youth across three peri-urban regions of Tanzania. Each curriculum targeted a specific skill domain and comprised 14 sessions, each lasting 1–2 hours, for a total of 27 hours. Participants were randomly assigned to receive awareness training, management training, or both. Training was delivered over one week in community venues by trained, gender-matched facilitators with tertiary education. Uptake was high, with over three-quarters of the participants attending at least 80% of the sessions. SES and perceived skills were assessed via behavioural tasks and self-reported

[2]The Effective Socio-emotional Skills To Gain Economic Empowerment (ESTEEM) framework identifies 14 SES considered malleable and likely to influence labour market outcomes, available here: https://poverty-action.org/socio-emotional-skills. These were selected based on existing SES frameworks, consultations with psychologists, focus groups, empirical literature on SES returns, and gender-based theoretical considerations (Marsh et al., 2025; Delavallade, Das, et al., 2025; Delavallade, Rouanet, and Das, 2020).

scales at baseline, three months, and one year post-intervention.

We find that socio-emotional skills (SES) training leads to significant short-term improvements in SES, particularly when measured via self-reports. Specifically, most of the 14 self-reported skills improved in the three months post-intervention, whereas only three interpersonal skills showed behavioural gains. Despite short-term SES gains, these effects are not sustained 1 year after training, regardless of the SES measure. Additionally, combining awareness and management modules — which doubles training time — does not yield stronger effects than offering either alone. This suggests that SES domains may be interconnected, with spillovers from targeted skills into non-targeted areas, and that focusing on a few key skills might be sufficient.

Overall, SES training alone appears insufficient to improve labour outcomes for the average youth. Nevertheless, one year post-intervention, SES training modestly increases income-generating activity among individuals who reported being job seekers at baseline. Specifically, among male participants actively seeking jobs at baseline, SES training increases participation in income-generating activities by six percentage points. However, we do not observe such labour market gains among the average NEET youth or women.

Our findings suggest that SES training may induce upwardly biased self-assessments of one's SES, which appear to generate labour market returns for men but not for women — consistent with evidence on gendered rewards to overconfidence (see Exley and Kessler, 2022; Ozen et al., 2020). Importantly, our intervention aimed to enhance SES in general rather than in specific contexts (e.g., entrepreneurship). Nor did it address external constraints such as limited job networks, transportation challenges, or employer bias—each of which may limit the effectiveness of SES training.[3]

This study contributes to the growing body of literature on the development of soft skills among youth. These skills are thought to complement technical skills and thus potentially yield substantial returns in the labour market (Heckman, Stixrud, and Urzua, 2006; Heckman and Kautz, 2012; Deming, 2017). While prior studies have shown that SES are malleable during adolescence and early adulthood (e.g., Groh et al., 2016; Campos et al., 2017; Prada, Rucci, and Urzúa, 2019; Ashraf et al., 2020; Chioda et al., 2021; Ubfal et al., 2022; Adhvaryu, Kala, and Nyshadham, 2023; Barrera-Osorio, Kugler, and Silliman, 2023; Brudevold-Newman and Ubfal, 2024), most focus on a single skill or narrow skill set, or

[3]Youth-targeted training programs often combine SES with technical skills or other labour market interventions (Groh et al., 2016; Chioda et al., 2021; Acevedo et al., 2020; Barrera-Osorio, Kugler, and Silliman, 2023; Brudevold-Newman and Ubfal, 2024; Dammert and Nansamba, 2023).

embed SES in broader interventions. We advance this literature by experimentally varying the type of SES training. Like Brudevold-Newman and Ubfal (2024), our combined arm addresses the same skill set, but we also test awareness and management curricula separately. Our management arm overlaps the skills studied in Ubfal et al. (2022). In line with these studies, SES levels improve in the three months after the intervention, but the effects are short-lived, fading away within one year. Our findings further suggest that SES components are interrelated, indicating that training either domain yields similar benefits.

Our study also expands the literature on SES measurement, which remains a challenge in evaluating the impact of SES training. Researchers often rely on self-reported scales, such as the Big Five Inventory, which are seldom validated in low-income settings. This has led to observations of a male advantage in self-reported SES (Ajayi et al., 2022; Hossain and Jukes, 2024), although recent data from behavioural tasks find no gender gap (Cassidy et al., 2024). Our dual-measurement approach —utilising both self-reports and behavioural tasks — enables us to assess whether training impacts skills as measured via different methods and whether these changes translate into labour market gains. We find that self-reports are more responsive to training, though their effects are transitory and gender-neutral in the short term.

Our study focuses on a policy-relevant segment of the population — NEET (not in education, employment, or training) youth. High youth unemployment, particularly among those with higher education, is a common challenge across Sub-Saharan Africa. In Tanzania, the context of our study, 23% of young women and 14% of young men aged 15-24 are neither in education nor in employment (Tanzania National Bureau of Statistics, 2021). Yet, employers often report difficulties in finding workers with the appropriate skills, especially soft skills (Cunningham and Villaseñor, 2016; Munishi, 2016; Tan, Bashir, and Tanaka, 2016; Morsy and Mukasa, 2020; Bassi and Nansamba, 2021). Information frictions may partially explain this skill mismatch—firms struggle to observe the soft skill levels of youth, often holding miscalibrated beliefs about their potential. Previous studies have documented this phenomenon, where such misperceptions hinder job formation (Bassi and Nansamba, 2021; Abebe, Caria, Fafchamps, Falco, Franklin, Quinn, and Shilpi, 2023).

Holding inflated self-assessments, although inefficient, has been associated with labour market returns. First, individuals with inflated self-assessments are more successful at convincing others of their abilities, particularly when their competence is uncertain (Burks et al., 2013; Schwardmann and Weele, 2019; Soldà et al., 2021; Demiral and Mollerstrom, 2024). Secondly, it can serve as intrinsic motivation, driving persistence and goal pursuit (Reuben, Wiswall, and Zafar, 2017; Hoffman and Burks, 2020; Barron and Gravert, 2022). Our findings generally align with these two mechanisms. First, SES training enhances participants' self-perception of their skills; however, with minimal improvements in actual skill levels, this creates a self-assessment gap. Second, this increase in the self-assessment gap is particularly pronounced among individuals with higher initial gaps in their self-assessment. Third, consistent with the literature (e.g. Reuben, Wiswall, and Zafar, 2017; Bordalo et al., 2019; Adamecz-Völgyi and Shure, 2022; Exley and Kessler, 2022), we observe that the self-assessment gap is disproportionately prevalent among men. Finally, the larger self-assurance improves labour market outcomes among male job seekers who exhibited larger self-assessment at the onset. However, such labour market returns are not observed for women, who likely face additional barriers or constraints that limit the efficacy of the SES training for them[4]. Thus, while SES training may help male job seekers signal their skills more effectively, it may not sufficiently mitigate the gendered disadvantages women face in the labour market.

4.2 Conceptual Framework

Socio-emotional skills (SES) may enhance labour productivity by shaping how individuals convert effort and aspirations into economic activity. These skills may influence the behavioural technology of labour supply—affecting entry into income-generating work, persistence in job search, and coordination in informal production. SES may complement cognitive and technical skills by reducing behavioural frictions, improving reliability, and strengthening cooperation, especially where supervision is limited, and contracts are incomplete (Borghans et al., 2008; Heckman and Kautz, 2012). Subsequently, they may raise productivity on both the extensive and the intensive margin. Yet SES are broad and multidimensional, spanning behavioural, emotional, and social competencies. Existing classifications—such as intrapersonal versus interpersonal—describe where skills operate rather than how they shape decisions. To identify the mechanisms through which SES affect productivity, this study adopts a functional framing that distinguishes SES into Awareness and Management as summarised in Figure

[4]Returns to overconfidence are unlikely to be a gender-neutral phenomenon. For example, literature documents that men benefit more from higher skill signals (see extant literature on men's hubris Burks et al., 2013; Hoffman and Burks, 2020; Adamecz-Völgyi and Shure, 2022), while women could even face penalties for signalling skills that are perceived as counterstereotypical (Ozen et al., 2020; Exley and Nielsen, 2024).

4.A.1. [5] These two complementary capacities may jointly influence activation, execution, and coordination in the labour market.

Awareness—encompassing self- and social-awareness— may enhance choice quality and activation by aligning aspirations with realistic opportunities (Ray, 2006; Lybbert and Wydick, 2018) and by building the relational capital necessary for productive engagement. Management—covering self- and relationship-management— on the other hand, may improve execution and coordination by sustaining effort, punctuality, and teamwork (see Alan, Boneva, and Ertac, 2019; Ashraf et al., 2020; Campos et al., 2017; Borghans et al., 2008; Heckman and Kautz, 2012). These two competencies may be complementary since activation without execution risks aspirational drift, while execution without activation risks disciplined misallocation (Genicot and Ray, 2020). Their complementarity implies that improved orientation (from Awareness) translates into stronger and more reliable enactment (from Management). In the labour market, Awareness may primarily affect entry margins. At the same time, Management may deepen hours and income. Thus, their joint exposure is likely to produce larger gains across both margins as summarised in Figure 4.A.2. This motivates our 2×2 factorial design, which enables us to test for synergistic effects, as predicted by dynamic-complementarity models (Cunha and Heckman, 2007).

Beyond improving actual behaviour, SES training can also shape how individuals perceive their capabilities. Exposure to training often prompts participants to revise self-assessments upward, creating a gap between perceived and measured competence. This overestimation arises broadly from increased self-reflection, new behavioural vocabulary, and motivational framing. Such effects may operate differently across domains: Awareness may elevate perceived orientation and self-efficacy, while Management may boost perceived reliability and agency. From a behavioural standpoint, this mechanism aligns with Bénabou and Tirole (2002) theory of self-confidence, which argues that optimistically biased self-beliefs sustain motivation and persistence. Within the logic of signalling theory (Spence, 1973), overestimation can be privately beneficial: stronger self-beliefs encourage assertive job search, bolder negotiation, and improved self-presentation. These expressive behaviours may alter employer perceptions, especially in low-information markets where confidence serves as a credible proxy for ability.

However, the returns to overestimation could differ by gender. While both men

[5] Several established frameworks in Social and Emotional Learning (SEL) and Emotional Intelligence (EI) make a similar functional distinction. These include the Collaborative for Academic, Social, and Emotional Learning (CASEL) five-competency model, Mayer and Salovey's four-branch Emotional Intelligence framework, and the OECD/Heckman–Kautz socio-emotional skill taxonomy.

and women may experience similar perceived gains, the labour market rewards assertiveness asymmetrically. Men are typically rewarded for confidence and ambition, which align with prevailing masculine norms, whereas women often face penalties for comparable self-assertion (Bassi and Nansamba, 2021; Exley and Kessler, 2022). Consequently, the same confidence that propels men's job search and income gains may yield muted or even adverse effects for women, whose assertiveness can be viewed as norm deviation rather than competence. Thus, while SES training enhances perceived empowerment for all, it may reproduce gendered inequities in how that empowerment is recognised and rewarded in the labour market. Therefore, in addition to gender, we test whether labour market effects vary by baseline job-seeking behaviours.

4.3 Context, Intervention and Study Design

4.3.1 Tanzanian Context

Tanzania, like most African countries, continues to witness a youth bulge. Tanzania's youth aged 15–24 account for nearly one-fifth of the population and about one-third of the working-age population. Yet their transition into stable employment remains low, with many entering the informal economy[6], as they enter the labour market with low academic qualifications.[7] According to the 2020/21 Integrated Labour Force Survey (ILFS), youth unemployment among youth (15–24) stands at around 15% (women 18%, men 12%), while 19% (women 23%, men 14%) of youth are not in employment, education or training (NEET)[8]. The slack in labour demand, alongside low educational levels, implies that these youths face a protracted school-to-work transition, taking nearly 20 months. These youths mostly search for jobs through informal networks, reflecting both limited access to formal job-matching systems and low returns to online or institutional search. The COVID-19 pandemic further constrained labour demand, pushing many youth into informal work.

Despite the recent expansion of secondary and tertiary education, many young Tanzanians, especially those with higher levels of education, struggle to find jobs that match their qualifications (ILO, 2018). Employers frequently report difficulty finding suitably skilled candidates—especially soft skills such

[6]Agriculture is the dominant economic activity in Tanzania, employing about 60% of the total workforce. Besides agriculture, about 90% of employed persons are in informal employment.

[7]A significant proportion of labour market entrants drop out of the formal education system at the primary (58%) or O-level secondary education (38%) levels.

[8]The NEET rate increased from 14% likely due to the economic downturn in 2020 caused by the Covid-19 pandemic

as work ethic, teamwork, problem-solving, and communication (Munishi, 2016; Tan, Bashir, and Tanaka, 2016). For example, in a recent nationwide survey of technicians' employers, more than 60% of the employers rated almost all soft skills as inadequate among technician graduates (NACTE, 2020). Subsequently, employers are forced to retrain employees due to weak soft skills at entry (Juma, 2025). In the absence of these re-training efforts, better-educated youth are forced to settle for low-skill jobs, even as firms face skill shortages.

The Tanzanian government recognises this mismatch and has instituted various initiatives to upskill its workforce. First, the National Skills Development Strategy (NSDS 2016–2026) emphasises aligning training with labour-market demand. Secondly, the Five-Year Development Plan III (2021/22–2025/26) prioritises youth employment, private-sector job creation, and modernisation of the technical and vocational education and training (TVET) institutions. Third is the establishment of the Tanzania Employment Services Agency (TaESA), which provides placement, counselling, and internship facilitation, while youth enterprise funds support self-employment. Yet these efforts primarily emphasise technical and entrepreneurial skills. Systematic investment in the development of socio-emotional skills remains limited, leaving a key lever of youth employability underdeveloped.

4.3.2 Intervention

The intervention took place from September to December 2021 and was implemented by BRAC Tanzania. The research team supervised the implementation. Depending on their treatment assignment, study participants were invited to participate in one of three training programs: a five-day training on "awareness" skills, a five-day training on "management" skills, or a five-day training on awareness skills followed by a five-day training on management skills (i.e., ten days of training in total). The curricula for the interventions were developed by Alkimia Consulting in collaboration with the research team, building on recent advances in psychology and behavioural economics as well as the curriculum of the BRAC's Empowerment and Livelihood for Adolescents (ELA)[9].

The skills included in the curriculum were derived from the ESTEEM frame-

[9] Empowerment and Livelihoods in Action (ELA) is a BRAC flagship program, which aims to enhance young women's economic, health & social outcomes. BRAC initiated efforts to revamp the ELA curriculum; hence, the curriculum evaluated in this study is part of a broader effort to adjust the life skills curriculum to suit changing economic and social contexts. Although substantial portions of the curriculum under this study were incorporated into the new ELA curriculum being implemented in multiple countries, the content covered only a portion of that curriculum, which extends beyond SES. The curricula developed are available here

work of 14 SES (Delavallade, Rouanet, and Das, 2020; Delavallade, Das, et al., 2025; Marsh et al., 2025). The list of skills was designed to be as exhaustive and mutually exclusive as possible and includes four sub-categories: self-awareness skills (emotional awareness, self-awareness); social awareness skills (listening, empathy); self-management skills (emotional regulation, self-control, perseverance, personal initiative, problem-solving, and decision making); and relationship management skills (expressiveness, interpersonal relatedness, influence, negotiation, and collaboration). The Awareness training focused on the four awareness skills, and the Management training focused on the 10 management skills as shown in Figure 4.A.1.

Both training sets consisted of 14 one- to two-hour sessions, totalling 27 hours of content. The training sets were based on an experiential learning approach, involving self-reflection and continuous engagement through open-ended questions, case studies, role-plays, and powerful visuals. For example, to develop self-awareness, participants are asked to use a tree as a metaphor for their life to understand their history, values, and goals. During the last three Awareness training sessions, individuals practised channelling their knowledge of strengths and desires into creating small businesses and into job searches. During the management training, exercises were more closely linked to employment and entrepreneurship. For example, to learn about problem-solving, groups were given business scenarios and practised creating creative solutions, listing the advantages and disadvantages of each solution, and developing an action plan.

A team of 18 facilitators led the two curriculum sets, with an equal number assigned to each, split by gender. The competitively recruited facilitators were a mix of tertiary college and university graduates. Facilitators first received the training themselves, which also served as a training pilot. They were then trained for five days in facilitation skills, including instruction on using the materials, managing groups, challenging participants, and conducting teach-back and feedback sessions. The extended training of facilitators also served as a platform for piloting curricula. Subsequently, the curriculum development consultant made relevant context adjustments during this facilitator training. Due to COVID-related travel restrictions, the facilitators' training was conducted in a hybrid format, with some trainers of trainers facilitating remotely. At the same time, all facilitators attended in person at a single venue. The participants' training sessions were held in person, divided by gender, and facilitated by facilitators of their gender.

4.3.3 Sample and randomisation

Our study targeted young men and women from 40 communities across three urban areas in Tanzania: Dodoma, Dar es Salaam, and Iringa. We conducted a census in January and February 2021 across these three areas. Communities were included in our sampling frame if the listing exercise identified them as having at least 120 eligible young men and 120 eligible young women within a one-mile radius, community leaders who confirmed interest in the program, and an existing venue suitable for hosting a training. We randomly selected 40 communities from the list of communities meeting these criteria. Within selected communities, the same listing exercise provided an individual sampling frame of all eligible young men and women aged 16-26 who were not in full-time education, training, or employment. We drew a random sample of 60 young women and 60 young men from each community in this sampling frame for the baseline survey conducted in May and June 2021.

Our study sample consisted of 2,358 young men and 2,370 young women who were neither in education, training, nor in full-time employment[10]. Only half of these participants were actively seeking jobs during the baseline survey. Despite similar ages (average = 21 years) and educational levels (average = 9 years) between men and women, we observe gender differences in labour outcomes at baseline. For example, compared to young women, we find young men were 10 percentage points (pp) more likely to be engaged in part-time income-generating activities and 10 pp more likely to be actively searching for a job at the time of the survey. Another marker of gender inequality was the disproportionate allocation of household chores to women, where they spent an average of two extra hours on these chores per day relative to young men. Both men and women were equally drawn from households with low socioeconomic status. For example, an average participant had guardians who had barely completed primary education, and approximately one-third of the participants came from female-headed households, which is typically correlated with poverty in this context.

We then randomised the respondents to one of four groups. Randomisation was conducted at the individual level, stratified by community, gender, and age (classified into below median age or not). Individuals within each stratum were randomly allocated with equal probability to one of three treatment arms – 27 hours Awareness (A), 27 hours Management (M), 54 hours A+M – or to the control arm (no training). As shown in Table 4.B.1, participants allocated to

[10]Nearly a third of them were engaged in part-time jobs, mostly informal jobs, and were own-account workers.

the three treatment arms were, on average, similar to those in the control arm. The training, which took place from September to December 2021, had a high level of participation. For example, 89% of participants allocated to receive the training attended at least one of their sessions, and training completion was also high: 70% (A), 71% (M), and 61% (A+M). Due to the shorter duration, single-arm participants were more likely to miss any of their allocated sessions (Column 1 of Table 4.B.2). As expected, training completion was slightly lower among those in the combined arm (Column 3 of Table 4.B.2). However, after accounting for the difference in training length, the results in Column 4 of Table 4.B.2 show no significant difference in the proportion of sessions attended. We also find no significant difference in treatment uptake by gender. Further exploratory analysis does not show any other systemic differences between those who take up and complete allocated treatment and those who do not take up but do not complete it.

We conducted the first follow-up survey between February and March 2022 (N=4,683, attrition 1%). This was approximately three months after the last batch of participants had completed their assigned training. We also conducted a second follow-up survey in October/November 2022 (N=4,645, attrition rate 2%), approximately 1 year after training completion. We used similar survey instruments across our three survey rounds. We had very low attrition: only 3% of our study participants were not observed across all three survey rounds.

4.3.4 Measuring SES

For each of the 14 skills, we included one self-report scale and one behavioural measure in each survey round. Among the self-report scales, five skills are measured using original items and based in theory, and nine are adapted from existing scales with selected original items[11]. For each skill, the self-report scale included 6 to 12 items and utilised a five-point Likert scale. To guard against zeroed-out scores, we aggregated skills within each category via a modified geometric mean. Final scores for individuals' skills and aggregated skills were standardised relative to the control group's mean and standard deviation.

These SES measures were developed and validated in three Sub-Saharan African countries (Marsh et al., 2025; Delavallade, Das, et al., 2025). The behavioural measures were either a situational judgment test (SJT) for nine skills or a task-based measure for five skills. Three out of the five task-based measures

[11] The sources for these items can be found in Marsh et al. (2025) and Delavallade, Das, et al. (2025), but include scales from seminal papers such as Schutte et al. (1998), Schwarzer and Jerusalem (1995), Duckworth and Quinn (2009), and Frese et al. (1997).

are original items, while two were adapted from existing measures, including negotiation (Selman et al., 1986) and perseverance (Alan, Boneva, and Ertac, 2019). The other three task-based measures included a simulated SMS conversation with responses encoded by enumerators to measure collaboration and an enumerator post-survey assessment to measure self-control. The SJTs, on the other hand, involved presenting respondents with two to three scenarios and asking about their likelihood of taking a specified action. Actions mirrored the self-reported scales and included a list of "good" actions associated with successful skill use and "poor" actions associated with poor skill use. The final scores for SJTs were a net count of "good" actions. The scores for task-based measures were based on simple averages, except for perseverance, which used a weighted average of the difficulty of the puzzle attempted by the respondent. We adopted a similar approach to self-reported scores in aggregating and standardising the final scores from behavioural measures.

4.3.5 Estimation Strategy

We estimate ITT effects for each follow-up survey, using the following ANCOVA specification for our main estimation:

$$\begin{aligned} Y_{is,t=1,2} = \alpha + \beta_1 A_{is} + \beta_2 M_{is} + \beta_3 (A+M)_{is} + \delta_1 Y_{is,t=0} \\ + \delta_2 Z_{is,t=0} + \delta_3 X'_{is,t=0} + \mu_s + \epsilon_{ist} \end{aligned} \tag{4.1}$$

$Y_{is,t}$ is the outcome of individual i in stratum s at time t and $Y_{is,t=0}$ its baseline. A_{is}, M_{is}, and $(A+M)_{is}$ is the treatment assignment of individual i in stratum s into either awareness only, management only and both training, respectively. $Z_{is,t=0}$ whether baseline value was missing. $X'_{is,t=0}$ vector of unbalanced baseline covariates. β_1, β_2, and β_3 coefficients of interest.

Our heterogeneity analysis examines the differences in treatment effects by the participants' gender. We interact each treatment dummy with the covariates in Equation 4.1 and fit the regression separately for our men and women sub-samples. In addition to the participant's gender, our pre-specified heterogeneity covariates, as measured at baseline, include socioeconomic status (SES), educational level, cognitive level, and age. We expand this list of heterogeneity covariates to include baseline labour market attachment, that is, participants' job-seeking status. Our study involves drawing inferences from multiple hypotheses; thus, we correct for the false discovery rate and report adjusted q-values, following Benjamini and Hochberg (1995). We report the pre-specified outcomes and follow the sequence outlined in our pre-analysis plan when discussing our findings.

4.4 Results

4.4.1 Impacts on Socio-Emotional Skills

SES training has a positive impact on men's and women's self-reported SES three months after completion. Column 1 in Panel A of Table 4.1 shows that training in awareness-related SES, management-related SES or their combinations increase aggregate SES by approximately 0.10 standard deviations ($p < 0.05$). We find similar positive gains in the SES sub-aggregates (Table 4.1, Columns 2 and 3, Panel A). Turning to comparability between self-reported SES versus behaviourally-elicited SES, our results indicate that SES improvements are more pronounced when SES is elicited via self-reported measures relative to behavioural measures. For example, the pairwise comparison between Columns 1–3 versus 4–6 in Table 4.1 shows that while the effects by self-reported measures are about 0.10 standard deviations, those noted based on behavioural measures are generally smaller, less than 0.05 standard deviations. Further, when we examine SES at more disaggregated levels, we find significant improvements ($p < 0.05$) in self-reported SES in 13 out of the 14 disaggregated SES, compared to only three (active listening, interpersonal relatedness and influence) that show significant improvements ($p < 0.05$) in behaviourally elicited SES (Figure 4.A.5). Among these disaggregated skills, those that showed improved behavioural measures yielded effect sizes comparable to those elicited by self-reported measures.

The short-term improvements in SES attributable to the SES training fade away within one year after training completion. Our estimates in Table 4.1 (Columns 1–6, Panel B) indicate that SES training has no significant effects at the endline, regardless of the measures adopted or the type of SES training received. The fade-out is pronounced when SES is elicited via self-report measures. Specifically, we can rule out equivalence ($p < 0.10$) between three-month and one-year SES training effects on self-reported SES (Table 4.1, Columns 1–3, Panel B). While we cannot rule out equivalence between the three-month and one-year effects on behaviourally elicited SES ($p > 0.10$), the sign of the differences in point estimates nonetheless suggests a fade-out effect (Table 4.1, Columns 4–6, Panel B), albeit a smaller starting point.

Both types of SES training—awareness-related and management-related—lead to comparable overall improvements in SES. For instance, Column 3 of Table 4.1 shows that participants trained in awareness-related SES improve their awareness-specific and overall SES. Similarly, Column 2 of Table 4.1 shows that those trained in management-related SES experience gains beyond the targeted domain. Over time, the fade-out of SES effects is comparable across participants trained in either curriculum or their combination. Moreover, Table 4.B.4 indicates no statistically significant difference in the effects on men's and women's SES in both three months (Columns 1-3) and one year post-intervention (Columns 4-6). That said, the awareness-related SES training yields marginally significant results among men. In contrast, when combined with management-related SES training, the effects observed in women are nearly half those observed in men. These findings suggest that management-related SES training consistently improves self-reported SES for both men and women, particularly within the three-month post-intervention.

The divergence of effect sizes in SES between self-reported and behavioural measures of SES potentially increases the skill self-assessment gap[12]. Following Anderson et al., 2012; Adamecz-Völgyi and Shure, 2022, we define the skill self-assessment gap[13] as the differences in the percentile rank between self-assessed and behaviourally-measured skills. First, we express each skill measure as a within-sample percentile rank. We then obtain the self-assessment gap as the residual from regressing self-assessed percentiles on behavioural measure percentiles. Consistent with the literature (e.g. Reuben, Wiswall, and Zafar, 2017; Bordalo et al., 2019; Adamecz-Völgyi and Shure, 2022; Exley and Kessler, 2022), the results in Column 1 of Table 4.B.3 indicate that men exhibit a larger skills self-assessment gap than women. Additionally, consistent with the Dunning–Kruger effect (Feld, Sauermann, and Grip, 2017), the gap is larger among individuals with lower skills when assessed via behavioural measures. Further exploration also shows a strong positive correlation ($\rho = 0.68$) in the skills self-assessment gap in various domains of SES. Turning to the impact estimates, we find that SES training initially increased the self-assessment gap among both men and women; however, this effect fades over time (Figure 4.1). Examining the conditional average treatment effects also indicates that the transient increases in self-assessment gap induced by the SES training are higher among men who already had higher gaps (Figure 4.A.4(a)). The fade-out of the self-assessment gap within one year plausibly indicates individuals' tendency to recalibrate their self-beliefs in the absence of reinforcement or further skill acquisition.

One key concern is whether the gains in self-reported SES and the attendant

[12]While we did not pre-specify this outcome, we build on the findings from our descriptive baseline paper (Cassidy et al., 2024). In that paper, we argued that the gap between self-reported skills and behaviourally-measured skills among men partly reflects men's overestimation of their skills

[13]The skills self-assessment gap in our study can alternatively be framed as skills overestimation, one of the three forms of overconfidence. Our study cannot adduce the other two forms of overconfidence, that is, the overplacement of one's skills compared to others and overprecision beliefs (or overcertainty) about one's skill levels.

self-assessment gap are artefacts of experimenter demand. As a sub-genre of the experimenter demand effects, we would expect participants with a higher propensity for social desirability bias to not only be upward biased in their judgments of their skills but also be motivated to respond with choices they deem to be in line with the researcher's hypothesis (Quidt, Haushofer, and Roth, 2018). In the latter case, the ITT estimates will be upwardly biased, as participants in treatment groups may believe the researcher's goal is to enhance their skills through SES training. We adopted two approaches to address this concern. First, the intervention delivery team was separated from the interviewing team to reduce the researcher's salience. Secondly, we elicit participants' propensity to give socially desirable answers as part of our baseline survey[14], and explore whether participants with high propensity to give socially desirable answers have different treatment effects. Reassuringly, we do not find consistent evidence of heterogeneity of treatment effects by participants' propensity to provide socially desirable answers. Specifically, the results in Table 4.B.9 (Panel A) show that men with high social desirability biases experienced minimal gains in SES following the training, whereas women showed the opposite trend. These findings suggest that the upward revision in participants' self-evaluation is a real change in self-assessment induced by the SES training rather than an artefact of experimenter demand.

We additionally explore whether awareness-related SES is foundational to the development of management-related SES. This hypothesis, grounded in psychological theory (e.g. Bandura, 1997; Dweck, 2006), posits that self-awareness and social awareness form the basis for the development of self-regulation and interpersonal management skills. If true, we would expect (i) stronger effects in the combined training arm and (ii) greater management skill gains among participants with higher baseline awareness SES. Our results do not support either of these predictions. First, we find no significant difference in SES gains across the three treatment arms (Table 4.1), suggesting that combining awareness and management training does not yield additional benefits.

Next, we turn to the second prediction: baseline awareness of SES significantly moderates the development of SES, especially management-related SES. Here, we examine the heterogeneity of SES effects by baseline SES levels, separately for awareness-related and management-related SES. We split our baseline sample into two groups based on whether their baseline SES was above or below the median of their respective community. We then examined whether treatment effects

[14]While analysing this baseline data Cassidy et al. (2024) found that self-reported measures of SES are positively correlated with social desirability bias, especially among men.

varied by baseline SES. Focusing on the conditional average treatment effects of the management SES training (Column 3 of Table 4.B.5, Panel A), we find that SES effects are equally null even among those with higher baseline awareness-related SES. And thus, we cannot rule out the equivalence of management SES training alone among those endowed with higher baseline awareness SES relative to those with lower baseline awareness SES who also received the single management SES training. Further exploratory analysis, however, reveals larger improvements in behaviorally elicited SES among participants with higher levels of education and cognitive abilities (Columns 1 - 3 of Table 4.B.7). These findings suggest that SES training could enhance skills at the top of the distribution, rather than awareness-related SES being foundational for the development of management-related SES.

4.4.2 Impacts on Economic Participation

On average, no meaningful changes in labour market outcomes are attributable to the SES training. Results in Table 4.2 show that SES training does not change the likelihood of labour force participation at three months and one year post-intervention (Column 2). Characteristically, SES training effect sizes on extensive margin labour force participation are tightly estimated around zero. Despite the null effects at the extensive margin, Column 3 (Panel A) results indicate that SES training increases labour force participation at the intensive margin, particularly in management-related areas. Relative to those in the control group, men and women trained in management-related SES increase the hours they dedicate to income-generating activities by an average of 0.10 standard deviations (Figure 4.A.3). These average intensive margin effects fade within one year, especially for women. The significant increase in hours spent on income-generating activities is not accompanied by increases in earnings (Column 4 of Table 4.2, Panel A), three months and one year post-intervention. Similarly, while point estimates indicate increased monthly earnings, the changes in earnings are not sufficiently large to achieve statistical significance at conventional levels.

An alternative explanation for the limited effects on labour force participation is that respondents devoted more time to job searching or pursuing further education. The results, however, do not, on average, support these arguments. The results in Table 4.2 show that SES training does not increase the likelihood that men or women seek wage employment. Similarly, SES training does not alter men's and women's aspirations to start a business. We next turn to educational outcomes. We note from the onset that education pursuit[15], and future education

[15]Youths in full-time education were screened out of our baseline sample. We only included

aspirations are low. The proportion of participants in our study sample who were pursuing or aspiring to pursue education dropped further in subsequent survey rounds. For example, in the control group, the number of youths who aspired to further education decreased by 58% between the two follow-up rounds. Examining causal estimates in Column 6 indicates that SES training does change the likelihood of pursuing further education and training. However, we observe a temporary decline in aspirations to pursue further education due to SES training. Specifically, the results in Table 4.2 (Column 7) show that SES training reduces the aspiration to pursue further education by up to three percentage points three months post-intervention (Panel A). Still, this decline is precisely null one year later (Panel B). Our results thus suggest SES may accelerate the revision of education pursuit aspirations among young men and women, at least in the three-month post-intervention period.

4.4.3 Returns to Self-Assessment Gap in the Labour Market

An upwardly biased self-evaluation can shape labour market outcomes via two mechanisms. On the one hand, under asymmetric information, it can serve as a persuasive signal that convinces others of one's competence, thereby improving one's employment prospects—even if their actual skills remain unchanged (Burks et al., 2013; Schwardmann and Weele, 2019; Soldà et al., 2021; Demiral and Mollerstrom, 2024). Overconfident individuals may project competence more convincingly during interviews or networking interactions. On the other hand, overconfidence may also serve as an intrinsic motivator, increasing job search intensity and persistence in the face of setbacks (Reuben, Wiswall, and Zafar, 2017; Hoffman and Burks, 2020). However, the returns to inflated self-evaluations are unlikely to be gender-neutral. For example, studies (such as Exley and Kessler, 2022; Ozen et al., 2020) show that labour markets often reward men—but penalise women—for assertive self-promotion. Subsequently, labour market returns to upwardly biased self-evaluations could be mediated by gender and baseline labour market engagement–yielding positive returns for men but lower effects for women. We thus explore the gender-differentiated response to SES training among active job seekers and those not actively seeking jobs.

We find significant gender-differentiated effects of SES training on labour market outcomes depending on whether they sought a job at baseline. One year post-intervention, we observe increased participation in income-generating activities due to SES training at both extensive and intensive margins, but only

youths who were either not pursuing education or, at most, pursuing education on a part-time basis.

among men seeking jobs at baseline (Table 4.4, Panel B). Specifically, the results in Column 2 indicate that SES training increased the probability of engaging in income-generating activities by six percentage points among men seeking jobs at baseline. Furthermore, we observed an increase in hours worked of approximately half an hour (or 6 to 10% relative to the control group) due to SES training among men seeking jobs at baseline. Although not significant at conventional levels, the results also suggest that these participants registered income increases of 6–15%. These effects were driven by men seeking to enter the labour market. We, however, do not find a similar heterogeneity in the impact of SES training in the women's sample.

Suggestive evidence supports the positive correlation between upwardly biased self-evaluations and labour market outcomes. First, at baseline, we find significant positive correlations between skill self-assessment gap and the likelihood of seeking employment, particularly among men (Column 3 of Table 4.B.3). The positive correlation between men's skill self-assessment gap and job-seeking likelihood persisted even after adjusting for the actual skill levels. Secondly, since conditional average treatment effects indicate that the increases in the self-assessment gap induced by the SES training are larger among men who already had higher gaps, it follows that the impact of the SES training on the skill self-assessment gap is concentrated among those who were already seeking jobs at baseline. Finally, this is the cadre of participants - men who (i) were active job-seekers at baseline and (ii) had larger increases in self-assessment gap after training - who registered improvements in labour market outcomes in the one-year post-intervention follow-up (Table 4.4, Panel B).

4.4.4 Impacts on Broader Well-being

Our final family of prespecified outcomes concerned psychological well-being. We examine participants' subjective well-being on a Cantril ladder, levels of depression and anxiety. Estimates in Column 1–3 of Table 4.3 show that the effects of SES training on participants' psychological well-being are mixed. On the one hand, we find that SES training marginally improves an individual's subjective well-being; on the other hand, it increases the likelihood of reporting mild depressive and mild anxiety symptoms in the three months after the intervention. All these effects on participants' psychological well-being also fade away within one year after the intervention. Our extensive heterogeneity analysis of the SES training effects on the psychological well-being of men and women yields no meaningful patterns.

Regarding downstream outcomes, we also prespecified three indicators of women's empowerment: disagreement with gender-regressive attitudes, participation in decision-making, and the ability to make one's own decisions. We restrict our analysis in Columns 4–6 of Table 4.3 to the women participants only. Results in Table 4.3 indicate that SES training does not shift women's views towards gender-egalitarian ideals, both in the three-month post-intervention and after one year. The SES training does not change women's decision-making power in either study horizon. We also examine the effects of SES training on women's time use and find no significant changes in their time-use patterns. This latter result could indicate a greater demand on women's time, as they, at least in the three-month post-intervention period, dedicate more time to income-generating activities as a result of the SES training; yet there was no corresponding change in the time they devote to household chores.

4.5 Discussion and Conclusion

We evaluate the impact of socio-emotional skills (SES) training on 4,728 young Tanzanian men and women who were neither in full-time employment, education, nor training (NEET), using a randomised controlled trial. We find that SES training leads to short-term improvements in self-reported SES across both awareness-related and management-related skill domains, with some behavioural gains as well. These gains, although not sustained over time, demonstrate the potential for SES training to shift short-term perceptions of skill and agency.

Notably, SES training yields modest but significant labour market improvements among a subgroup of men actively seeking employment at baseline, who experience increased participation in income-generating activities at both extensive and intensive margins. These effects appear to be driven by upward-biased self-evaluations induced by the training. This effect may act as a persuasive signal of competence in the absence of formal qualifications. Although women exhibit similar increases in self-assessment, these do not translate into improved labour outcomes, suggesting the presence of deeper structural barriers, such as restrictive gender norms or employer discrimination.

While combining awareness and management training doubles training time, it does not yield additional benefits beyond offering either module alone, suggesting strong interconnections between SES domains and potential spillovers across skill types. Exploratory analyses suggest slightly higher returns to management-related skills, though both domains contribute meaningfully.

Our study contributes to the growing literature on non-cognitive skills and their relationship with labour market outcomes. First, we isolate the relative returns to different SES components by linking measured skills to a specific curriculum. Second, by targeting a particularly vulnerable population—NEET youth—we shed light on two plausible SES-related drivers of their labour market exclusion. On the one hand, these youth may lack the SES employers demand, which justifies the policy's emphasis on upskilling programs. On the other hand, informational frictions or employer biases may prevent accurate observation of these skills,[16] even when youth possess relevant capabilities leading to suboptimal hiring decisions (Bassi and Nansamba, 2021; Abebe, Caria, Fafchamps, Falco, Franklin, Quinn, and Shilpi, 2023).

Our results highlight a complementary mechanism: even modest changes in self-perceived SES—absent measurable improvements in actual skill levels—can positively affect job search outcomes, particularly for young men. This aligns with evidence on the labour market returns to male overconfidence (Burks et al., 2013; Ozen et al., 2020; Hoffman and Burks, 2020; Adamecz-Völgyi and Shure, 2022), though it contrasts with studies that find negative effects of job-seeker overoptimism (Kiss et al., 2023; Abebe, Caria, Fafchamps, Falco, Franklin, Quinn, and Shilpi, 2023). One explanation lies in sample selection: unlike prior studies targeting positively selected populations, our sample includes early-stage labour market entrants who dropped out midway through secondary school. In such contexts, higher self-belief and expressive signalling may serve as valuable substitutes for formal certification, particularly for men.

The improvements in self-reported skills and subsequent inflated self-assessments could partly reflect social desirability bias. Several patterns, however, suggest the observed effects are not purely artefactual. First, we note improvements in a few behaviourally elicited SES. Second, the notable improvements in labour market outcomes, albeit on a sub-sample, long after program completion, suggest partial internalisation of these competencies and perceptions rather than transient response bias. Thirdly, consistent with evidence from other randomised control trials, soft-skills training frequently elevates self-perceptions—optimism and self-efficacy—even when employment effects are muted or delayed (such as Acevedo et al., 2020; Groh et al., 2016).

The intervention's cost-effectiveness and delivery model compare favourably with other SES training interventions. Our 27-hour community-based training de-

[16] Recent evidence from certification programs in South Africa (Abel, Burger, and Piraino, 2020; Carranza et al., 2022), Ethiopia (Abebe, Caria, Fafchamps, Falco, Franklin, and Quinn, 2020), and Uganda (Bassi and Nansamba, 2021) confirms the potential of such interventions to mitigate informational frictions and improve employment outcomes.

livered modest but cost-effective psychosocial improvements relative to lengthier, more comprehensive initiatives in Rwanda (Brudevold-Newman and Ubfal, 2024) and Jordan (Groh et al., 2016), among others (See Appendix Table 4.B.11) that achieved larger employment gains. Our study presents evidence suggesting that internal constraints—e.g., confidence—may limit the effectiveness of upskilling programs. Nonetheless, internal transformation alone cannot overcome weak labour demand. Sustained youth empowerment requires addressing broader constraints to improve the labour market outcomes of NEET youth.

4.6 Tables

Table 4.1: Effect of SES Training on Aggregate SES

	Self-reported SES			Behavioural SES		
	(1)	(2)	(3)	(4)	(5)	(6)
	All SES	Awareness	Management	All SES	Awareness	Management
Panel A: Midline: Pooled Sample						
Awareness (A)	0.097 (0.037)***	0.088 (0.037)**	0.095 (0.037)**	0.022 (0.028)	0.029 (0.028)	0.011 (0.029)
	[0.027]	[0.041]	[0.027]	[0.488]	[0.424]	[0.692]
Management (M)	0.135 (0.037)***	0.127 (0.037)***	0.131 (0.038)***	0.028 (0.029)	0.024 (0.028)	0.017 (0.029)
	[0.004]	[0.004]	[0.004]	[0.424]	[0.488]	[0.602]
A + M	0.097 (0.037)***	0.088 (0.039)**	0.095 (0.037)**	0.042 (0.028)	0.048 (0.029)*	0.028 (0.029)
	[0.027]	[0.045]	[0.027]	[0.228]	[0.167]	[0.424]
P-value (A=M=A+M)	0.510	0.497	0.556	0.764	0.661	0.831
Control mean	-0.001	-0.001	-0.001	-0.005	-0.009	-0.001
Observations	4,607	4,607	4,607	4,323	4,590	4,331
Panel B: Endline: Pooled Sample						
Awareness (A)	0.023 (0.032)	0.012 (0.033)	0.028 (0.032)	-0.005 (0.028)	0.020 (0.026)	-0.019 (0.028)
	[0.873]	[0.873]	[0.873]	[0.900]	[0.873]	[0.873]
Management (M)	0.042 (0.031)	0.021 (0.033)	0.052 (0.032)	-0.009 (0.027)	0.008 (0.026)	-0.015 (0.027)
	[0.873]	[0.873]	[0.873]	[0.873]	[0.873]	[0.873]
A + M	0.016 (0.032)	0.009 (0.033)	0.018 (0.033)	0.013 (0.028)	-0.003 (0.027)	0.021 (0.028)
	[0.873]	[0.873]	[0.873]	[0.873]	[0.900]	[0.873]
P-value (A=M=A+M)	0.679	0.930	0.531	0.725	0.679	0.316
Control mean	0.000	-0.000	0.001	-0.000	-0.001	-0.000
Observations	4,607	4,607	4,607	4,350	4,594	4,352
Awareness (A) (Endline - Midline)	-0.073 (0.048)	-0.076 (0.049)	-0.067 (0.049)	-0.031 (0.039)	-0.009 (0.038)	-0.039 (0.040)
Management (M) (Endline - Midline)	-0.093 (0.049) *	-0.106 (0.050) **	-0.080 (0.049)	-0.042 (0.039)	-0.016 (0.038)	-0.040 (0.040)
A + M (Endline - Midline)	-0.081 (0.049) *	-0.079 (0.051)	-0.077 (0.049)	-0.034 (0.040)	-0.052 (0.039)	-0.016 (0.040)

Dependent variables in each column are standardised values of the corresponding SES (sub)aggregates. Each SES score is standardised relative to the values of the control group, separately for each survey wave. For each survey wave, we test the equality of treatment effects reported under our three treatment arms and report the p-value associated with this Wald test. Differences (Endline - Midline) in Panel B test differences between endline and midline coefficients. These coefficient differences are extracted from a pooled regression that includes a full interaction between treatment and survey wave dummies, while also adjusting for a full interaction between baseline value and the survey wave. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.2: Effect of SES Training on Labour Market Outcomes

	(1) Engaged any IGA	(2) Hours IGA	(3) Monthly income	(4) Searched for job	(5) Planning IGA	(6) Attending school	(7) Aspire education
			Panel A: Midline: Pooled Sample				
Awareness (A)	0.005 (0.019)	0.106 (0.154)	4.311 (7.553)	0.004 (0.018)	-0.003 (0.014)	0.005 (0.011)	-0.033 (0.015)**
	[0.896]	[0.825]	[0.825]	[0.896]	[0.896]	[0.825]	[0.241]
Management (M)	0.009 (0.018)	0.428 (0.155)***	9.372 (7.665)	0.013 (0.018)	0.009 (0.014)	0.003 (0.011)	-0.023 (0.015)
	[0.825]	[0.078]	[0.825]	[0.825]	[0.825]	[0.896]	[0.825]
A + M	-0.008 (0.018)	0.129 (0.155)	-0.001 (7.431)	0.014 (0.018)	0.012 (0.014)	0.006 (0.011)	-0.014 (0.015)
	[0.873]	[0.825]	[1.000]	[0.825]	[0.825]	[0.825]	[0.825]
P-value (A=M=A+M)	0.632	0.064	0.467	0.827	0.496	0.969	0.407
Control mean	0.507	4.735	143.795	0.671	0.747	0.074	0.186
Observations	4,609	4,609	4,609	4,609	4,606	4,609	4,609
			Panel B: Endline: Pooled Sample				
Awareness (A)	-0.002 (0.018)	-0.008 (0.142)	-4.765 (6.115)	0.007 (0.016)	0.005 (0.013)	-0.004 (0.008)	-0.001 (0.012)
	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]
Management (M)	0.019 (0.018)	0.019 (0.144)	2.064 (6.205)	-0.004 (0.016)	0.002 (0.013)	-0.003 (0.008)	-0.001 (0.013)
	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]
A + M	0.006 (0.018)	-0.010 (0.144)	-0.623 (6.101)	-0.017 (0.016)	0.004 (0.013)	-0.002 (0.009)	-0.006 (0.012)
	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]	[0.962]
P-value (A=M=A+M)	0.485	0.975	0.522	0.286	0.970	0.977	0.869
Control mean	0.480	4.836	136.269	0.761	0.722	0.049	0.108
Observations	4,607	4,607	4,605	4,607	4,604	4,607	4,607
Awareness (A) (Endline - Midline)	-0.007 (0.026)	-0.114 (0.209)	-9.075 (9.718)	0.003 (0.024)	0.008 (0.019)	-0.009 (0.014)	0.032 (0.019) *
Management (M) (Endline - Midline)	0.009 (0.025)	-0.409 (0.211) *	-7.308 (9.862)	-0.017 (0.024)	-0.007 (0.019)	-0.006 (0.014)	0.022 (0.020)
A + M (Endline - Midline)	0.014 (0.026)	-0.139 (0.212)	-0.622 (9.614)	-0.031 (0.024)	-0.008 (0.019)	-0.008 (0.014)	0.007 (0.020)

Dependent variables in each column are dummy variables except Columns 2 and 3. Column 1 indicates whether a respondent was engaged in any income-generating activity at the time of the survey, which was derived from the seven-day recall preceding the survey. Column 2 indicates participation in income-generating activities at the intensive margin and records the number of hours respondents spent on these activities on a typical day in the week preceding the survey. Column 3 represents the average monthly income earned by the respondent, expressed in constant 2017 USD, adjusted for inflation and purchasing power parity at the time of the survey. Column 4 reports whether the respondent searched for a job in the three months preceding the survey and was derived from an aggregate of various job search activities, including submitting applications, inquiring about potential vacancies, and networking with others. Column 5 indicates whether a respondent was planning to start an income-generating activity at the time of the survey. The last two columns report whether the respondent is enrolled in school or training and whether the respondent aspires to pursue further education, respectively. For each survey wave, we test the equality of treatment effects reported under our three treatment arms and report the p-value associated with this Wald test. Differences (Endline - Midline) in Panel B test differences between endline and midline coefficients. These coefficient differences are extracted from a pooled regression that includes a full interaction between treatment dummies and survey wave dummies, while adjusting for a full interaction between baseline values and the survey wave. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.3: Effect of SES Training on Gender Outcomes and Psychological Wellbeing

	Psychological Wellbeing			Gender Outcomes		
	(1) Life ladder	(2) Depression score	(3) Anxiety score	(4) Gender equitable index	(5) Influences some decisions	(6) Can make own decisions
Panel A: Midline						
Awareness (A)	0.090 (0.047)*	0.095 (0.121)	0.058 (0.116)	0.009 (0.017)	0.003 (0.009)	-0.005 (0.009)
	[0.358]	[0.650]	[0.790]	[0.674]	[0.744]	[0.674]
Management (M)	0.039 (0.047)	0.214 (0.122)*	0.104 (0.116)	0.028 (0.017)*	0.009 (0.008)	0.007 (0.008)
	[0.650]	[0.358]	[0.650]	[0.674]	[0.674]	[0.674]
A + M	0.060 (0.048)	0.041 (0.125)	-0.003 (0.114)	0.009 (0.016)	0.006 (0.008)	0.005 (0.009)
	[0.634]	[0.834]	[0.981]	[0.674]	[0.674]	[0.674]
Control mean	3.617	3.797	3.256	0.312	0.767	0.773
Observations	4,609	4,609	4,608	2,323	2,312	2,235
Panel B: Endline						
Awareness (A)	0.054 (0.041)	0.069 (0.122)	0.020 (0.117)	-0.019 (0.017)	0.000 (0.008)	0.006 (0.009)
	[0.551]	[0.850]	[0.865]	[0.959]	[0.959]	[0.959]
Management (M)	-0.018 (0.041)	0.062 (0.122)	-0.035 (0.117)	-0.005 (0.017)	-0.015 (0.009)*	0.002 (0.010)
	[0.850]	[0.850]	[0.858]	[0.959]	[0.692]	[0.959]
A + M	0.058 (0.042)	-0.137 (0.119)	-0.156 (0.113)	-0.002 (0.016)	-0.001 (0.008)	0.006 (0.009)
	[0.551]	[0.563]	[0.551]	[0.959]	[0.959]	[0.959]
Control mean	3.820	4.618	4.142	0.284	0.770	0.762
Observations	4,607	4,607	4,607	2,322	2,311	2,222
Awareness (A) (Endline - Midline)	-0.036 (0.062)	-0.025 (0.172)	-0.038 (0.165)	-0.028 (0.024)	-0.003 (0.012)	0.007 (0.013)
Management (M) (Endline - Midline)	-0.057 (0.063)	-0.152 (0.173)	-0.140 (0.165)	-0.033 (0.024)	-0.025 (0.012) **	-0.009 (0.013)
A + M (Endline - Midline)	-0.002 (0.064)	-0.178 (0.172)	-0.153 (0.161)	-0.011 (0.023)	-0.007 (0.012)	-0.003 (0.013)

Dependent variables in this table are continuous variables. The first column is based on the respondents' subjective evaluation of their overall life at the time of the survey, using a 10-point Cantril ladder, where 0 represents the worst possible life and 10 represents the best possible life. Column 2 is an aggregate of how frequently a respondent experienced symptoms of depression in the two weeks preceding the survey; questions are based on the Patient Health Questionnaire. Similarly, Column 3 is an aggregate from the Generalised Anxiety Disorder 7-item (GAD-7) scale in the last two weeks. Column 4 is also an aggregate of how often the respondent disagrees with four gender-regressive statements. Column 5 averages how frequently the respondent's opinions are considered in a series of 7 decisions concerning her. The last column represents the proportion of 7 decisions that respondents can make independently. The estimation of the effects of SES training on psychological outcomes in this table is based on the full sample; however, the estimation of effects on gender outcomes is restricted to the women's sub-sample. Differences (Endline - Midline) in Panel B test differences between endline and midline coefficients. These coefficient differences are extracted from a pooled regression that includes a full interaction between treatment and survey wave dummies, while also adjusting for a full interaction between baseline value and the survey wave. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

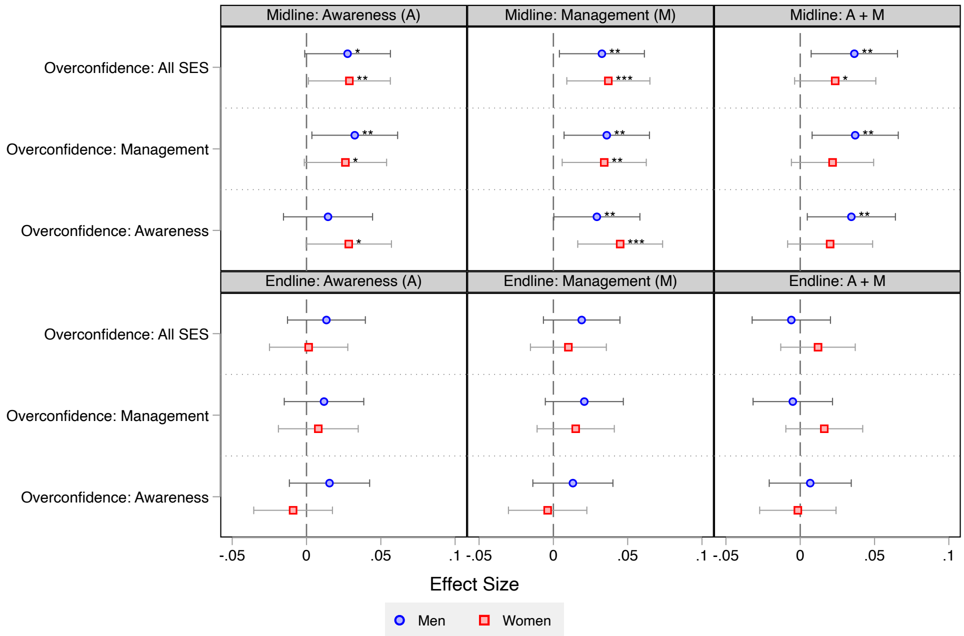
Table 4.4: Labour Market Outcomes Heterogeneity by Searching for a job

	Men			Women		
	(1) Engaged any IGA	(2) Hours IGA	(3) Monthly income	(4) Engaged any IGA	(5) Hours IGA	(6) Monthly income
Panel A: Midline						
SES treat	0.001 (0.033)	0.017 (0.067)	0.039 (0.075)	0.009 (0.029)	0.058 (0.052)	0.089 (0.049)*
	[0.964]	[0.964]	[0.964]	[0.750]	[0.399]	[0.217]
Searching for a job × SES treat	0.021 (0.044)	0.046 (0.089)	-0.044 (0.102)	-0.031 (0.043)	0.004 (0.076)	-0.112 (0.076)
	[0.665]	[0.665]	[0.665]	[0.701]	[0.962]	[0.415]
Searching for a job	0.009 (0.039)	0.003 (0.079)	0.106 (0.090)	0.009 (0.037)	-0.001 (0.067)	0.081 (0.068)
SES treat + Searching for a job × SES treat	0.022 (0.028)	0.062 (0.057)	-0.005 (0.070)	-0.022 (0.031)	0.062 (0.054)	-0.023 (0.056)
	[0.653]	[0.653]	[0.946]	[0.678]	[0.678]	[0.678]
Control mean: Not searching for a job	0.580	0.278	0.201	0.450	-0.269	-0.257
Panel B: Endline						
SES treat	-0.037 (0.030)	-0.102 (0.065)	-0.097 (0.072)	-0.047 (0.028)	-0.095 (0.053)*	-0.029 (0.047)
	[0.230]	[0.230]	[0.230]	[0.151]	[0.151]	[0.531]
Searching for a job × SES treat	0.096 (0.041)**	0.233 (0.085)***	0.180 (0.093)*	0.074 (0.042)*	0.094 (0.079)	0.005 (0.077)
	[0.030]	[0.019]	[0.053]	[0.233]	[0.352]	[0.951]
Searching for a job	-0.036 (0.036)	-0.089 (0.074)	-0.074 (0.082)	-0.050 (0.037)	-0.038 (0.070)	0.012 (0.069)
SES treat + Searching for a job × SES treat	0.059 (0.027)**	0.130 (0.054)**	0.083 (0.062)	0.027 (0.031)	-0.000 (0.059)	-0.025 (0.062)
	[0.043]	[0.043]	[0.180]	[0.997]	[0.997]	[0.997]
Control mean: Not searching for a job	0.535	0.261	0.183	0.480	-0.183	-0.140

Columns 1 and 4 indicate whether a respondent was engaged in any income-generating activity at the time of the survey, and were derived from the seven-day recall preceding the survey. Columns 2 and 5 indicate participation in income-generating activities at an intensive margin and record the number of hours respondents spent on these activities on a typical day in the week preceding the survey. Columns 3 and 6 represent the average monthly income earned by the respondent, expressed in constant 2017 USD, adjusted for inflation and purchasing power parity at the time of the survey. Income and hours spent on income-generating activities are standardised relative to the control values in a survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.7 Figures

Figure 4.1: Effects of SES Training on Overconfidence



Note: Each row, in the top and bottom panels, is an extract from a single regression as specified in Equation 4.1. Within each survey wave, we interact the treatment dummy with the gender dummy and extract appropriate linear combinations for each specified outcome in a row. Each dependent variable, overconfidence, is a continuous variable indicating the residuals obtained by regressing percentile ranks from self-reported measures against percentile ranks from behavioural measures for each SES (sub)aggregate, following Anderson et al., 2012; Adamecz-Völgyi and Shure, 2022. Regressions also include fixed effects for the enumerator and randomisation strata. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 4.A Appendix Figures

Figure 4.A.1: Skills definitions with levels of skills aggregation

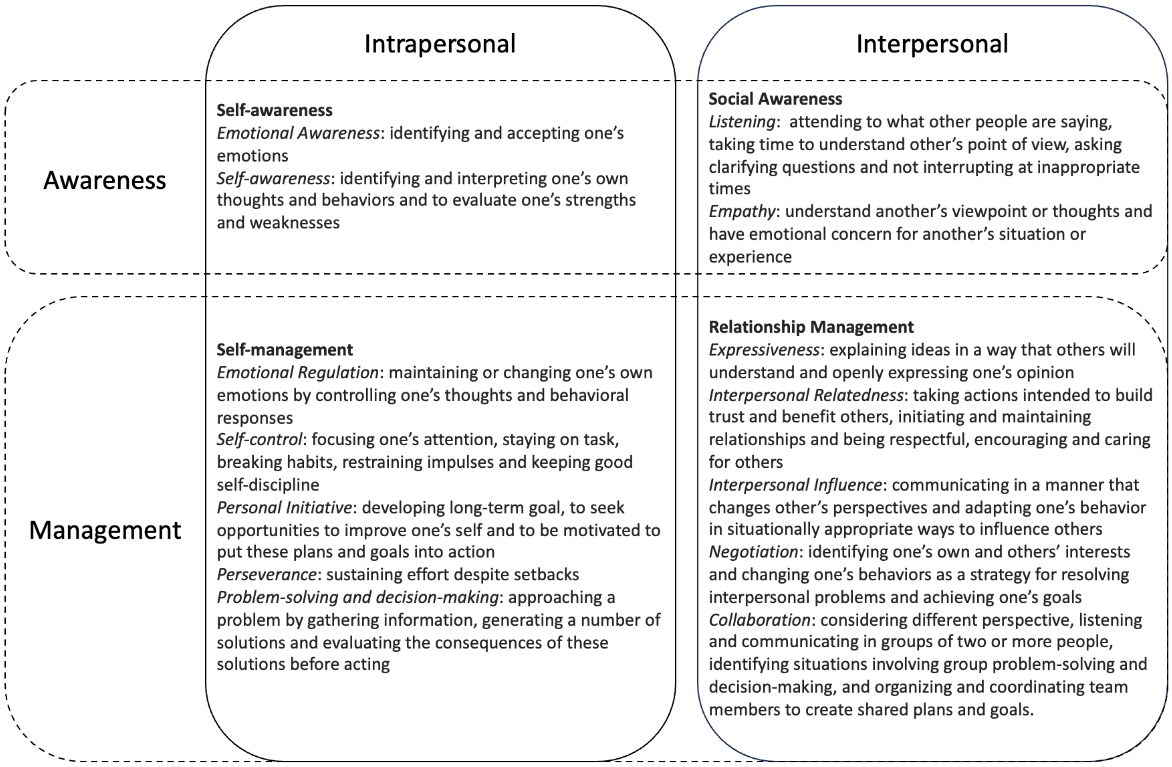


Figure 4.A.2: Summary of the Conceptual Framework

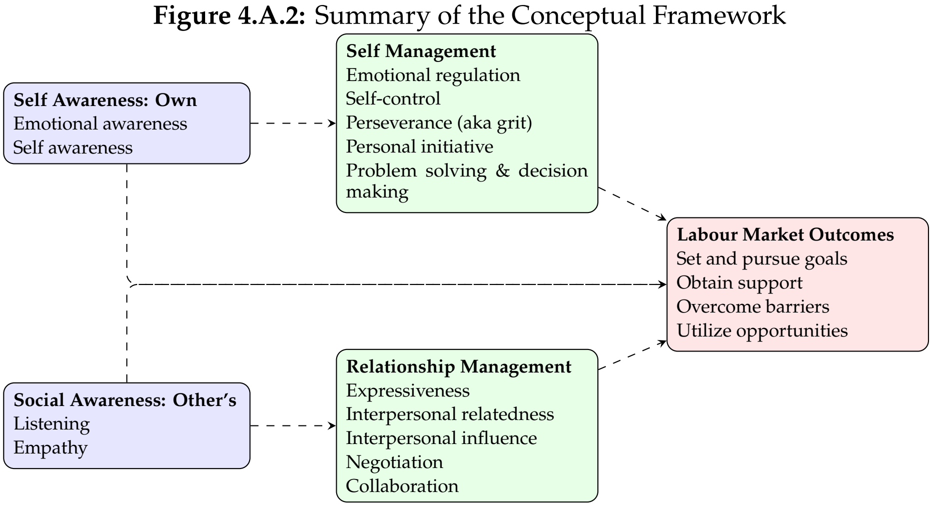
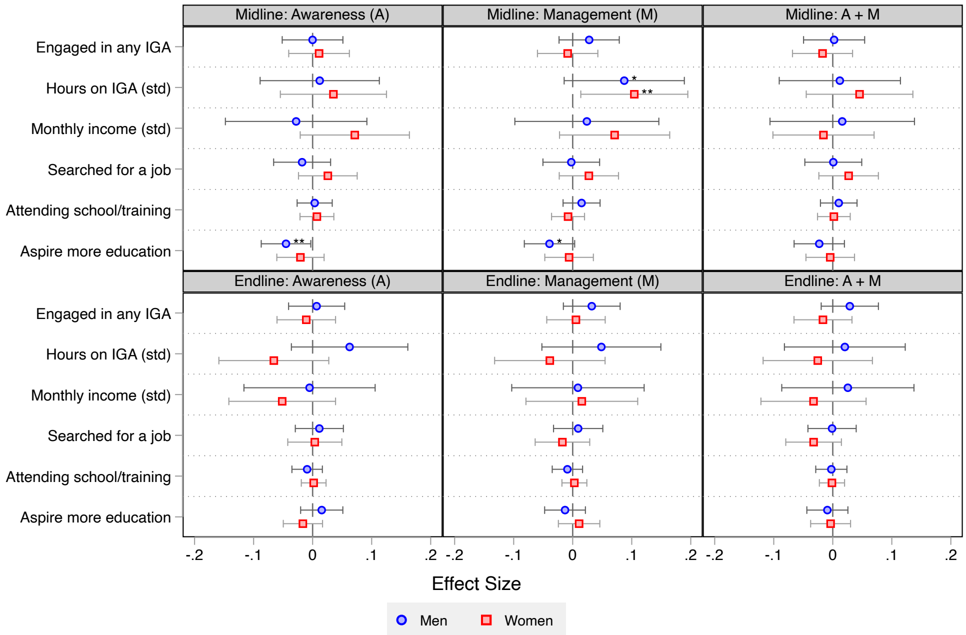


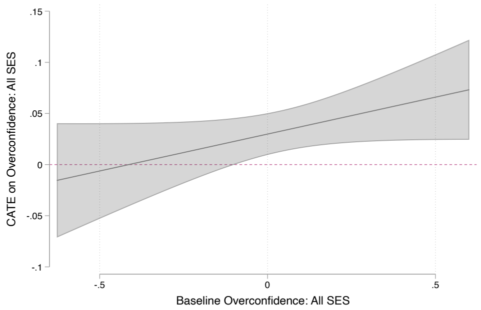
Figure 4.A.3: Effect of SES Training on Labour Market Outcomes by Sex



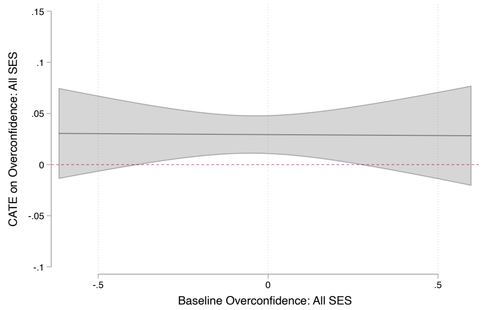
Note: Each row, in the top and bottom panels, is an extract from a single regression as specified in Equation 4.1. Within each survey wave, we interact the treatment dummy with the gender dummy and extract appropriate linear combinations for each specified outcome in a row. Engaged in IGA indicates whether a respondent was engaged in any income-generating activity at the time of the survey, as reported through a seven-day recall preceding the survey. IGA hours records the number of hours respondents spent on these activities on a typical day in the week preceding the survey, while income is the average monthly income earned by the respondent. Both hours and income are standardised relative to the control group for each survey wave. Job search is a dummy variable indicating whether the respondent searched for a job in the three months preceding the survey, derived from an aggregate of various job search activities, including submitting applications, inquiring about potential vacancies, and networking with others. IGA plan indicates whether a respondent was planning to start an income-generating activity at the time of the survey. The last two rows report whether the respondent is enrolled in school or training and whether the respondent aspires to pursue further education, respectively. Each regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4.A.4: Conditional Average Treatment Effects (CATE) on Overconfidence

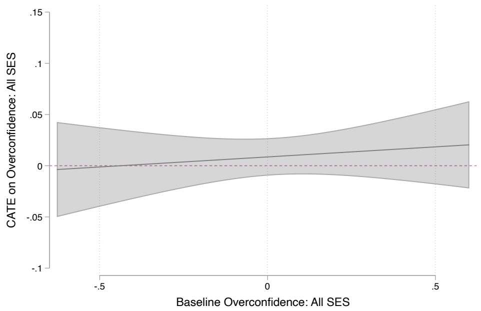
(a) CATE at Midline (Men)



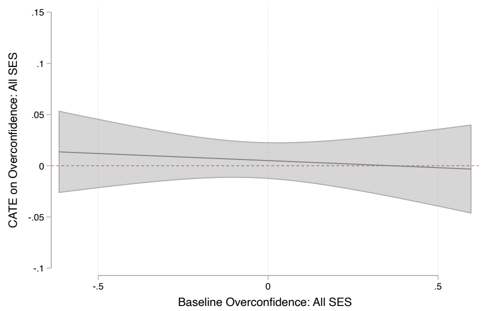
(b) CATE at Midline (Women)



(c) CATE at Endline (Men)



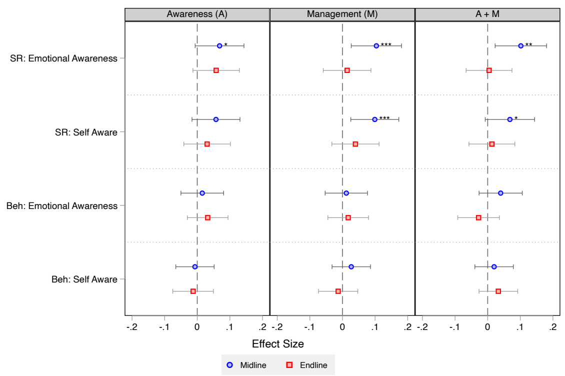
(d) CATE at Endline (Women)



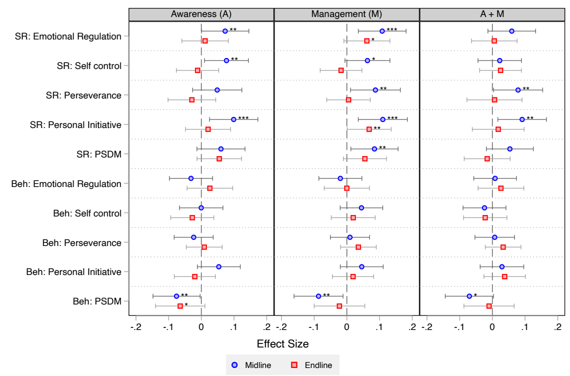
Note: Each panel shows the conditional average treatment effects (CATE) of SES training on levels of overconfidence at various levels of baseline overconfidence. Each panel correspond to a follow-up survey wave, separately for men and women. For each survey wave, we fully interact a participant's baseline overconfidence levels with the treatment dummy. We do not distinguish between our three treatment arms. Instead, we pool our treatments into a single dummy variable to increase power. For each panel, the dependent variable, overconfidence, is a continuous variable indicating the residuals obtained by regressing percentile ranks from self-reported measures against percentile ranks from behavioural measures, following Anderson et al., 2012; Adamecz-Völgyi and Shure, 2022. Regressions also include fixed effects for the enumerator and randomisation strata. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. The shaded region represents the 95% confidence interval for the CATE.

Figure 4.A.5: Effects on SES Training on Disaggregated SES

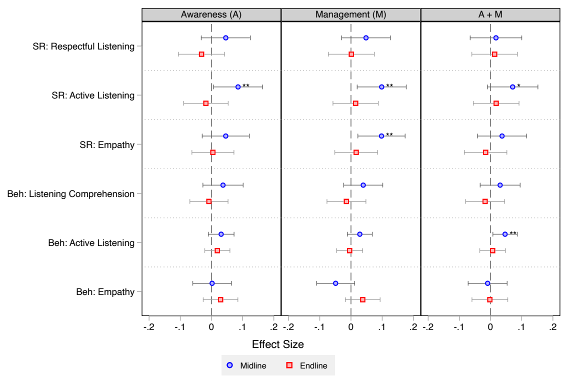
(a) Self Awareness



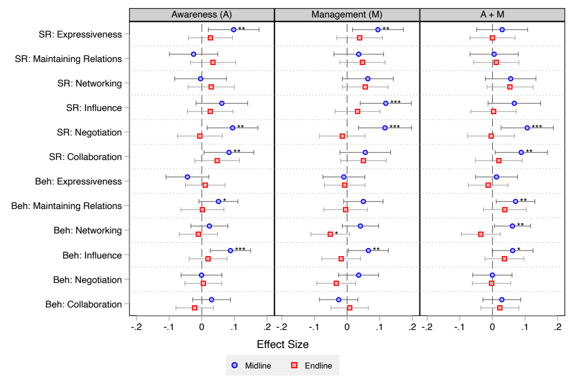
(b) Self Management



(c) Social Awareness



(d) Relationship Management



Note: Each row in the top and bottom panels is an extract from a single regression as specified in Equation 4.1. Within each survey wave, we interact the treatment dummy with the gender dummy and extract appropriate linear combinations for each specified outcome in a row. Each dependent variable is a continuous variable indicating the scores described in Section 4.3.4. Regressions also include fixed effects for the enumerator and randomisation strata. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 4.B Appendix Tables

Table 4.B.1: Baseline Characteristics Balance

	Awareness (A)	Management (M)	A + M	Joint P-value (A=M=A+M=0)	Men Control	Women Control
Age in years	-0.016(0.065)	0.073(0.064)	-0.026(0.064)	0.397	21.086	21.122
Married/Cohabitating	-0.025(0.012)**	-0.006(0.012)	-0.006(0.012)	0.154	0.059	0.168
Years of education	-0.227(0.127)*	-0.083(0.129)	-0.110(0.130)	0.354	9.168	8.959
Cognitive ability index	-0.019(0.010)**	-0.006(0.009)	0.016(0.009)*	0.002	0.709	0.685
Respondent is household head	0.001(0.013)	0.014(0.013)	-0.002(0.013)	0.607	0.194	0.062
Has a bank account	-0.002(0.010)	0.020(0.011)*	0.007(0.010)	0.183	0.088	0.046
Total savings (USD 2017/18 PPP)	0.186(5.242)	4.227(5.219)	3.948(5.175)	0.764	47.203	37.259
Female-headed household	0.003(0.019)	0.003(0.019)	0.014(0.019)	0.893	0.241	0.430
Household size	0.140(0.100)	-0.034(0.096)	0.103(0.098)	0.252	4.884	5.358
Household owns smart phone	-0.024(0.020)	0.015(0.020)	-0.007(0.020)	0.264	0.569	0.554
Major construction of walls is cement	0.004(0.011)	-0.008(0.011)	0.005(0.011)	0.640	0.888	0.889
Type of toilet is flush toilet	0.011(0.015)	-0.004(0.015)	0.009(0.015)	0.713	0.747	0.765
Mother's years of education	-0.174(0.099)*	-0.075(0.097)	-0.190(0.100)*	0.182	7.667	7.478
Father's years of education	-0.160(0.104)	-0.035(0.104)	-0.088(0.106)	0.444	8.087	7.616
SR: All SES	-0.044(0.039)	0.005(0.039)	-0.054(0.039)	0.323	0.111	-0.107
SR: Awareness	-0.032(0.038)	0.027(0.039)	-0.048(0.038)	0.214	0.088	-0.094
SR: Management	-0.047(0.039)	-0.005(0.040)	-0.053(0.039)	0.406	0.115	-0.106
Beh: All SES	-0.044(0.033)	-0.014(0.032)	0.006(0.031)	0.427	0.029	-0.017
Beh: Awareness	-0.006(0.030)	-0.030(0.030)	0.005(0.029)	0.636	-0.009	0.019
Beh: Management	-0.052(0.033)	-0.007(0.033)	0.002(0.033)	0.337	0.041	-0.035
Engaged in any IGA	0.010(0.018)	0.017(0.018)	0.016(0.018)	0.772	0.376	0.263
Time spent on IGA (Hours)	-0.231(0.126)*	0.063(0.129)	-0.069(0.129)	0.110	3.738	1.760
Hours on IGA (std)	-0.065(0.035)*	0.018(0.036)	-0.019(0.036)	0.110	0.283	-0.272
Monthly income (USD 2017/18 PPP)	-2.934(5.458)	0.877(5.581)	-3.565(5.478)	0.798	91.008	48.736
Monthly income (std)	-0.020(0.037)	0.006(0.038)	-0.024(0.038)	0.798	0.150	-0.140
Searched for a job	0.007(0.019)	-0.020(0.019)	0.003(0.019)	0.484	0.565	0.467
Planning new IGA	-0.001(0.016)	-0.005(0.016)	0.014(0.016)	0.622	0.594	0.614
Attending school/training	-0.003(0.007)	0.004(0.008)	0.007(0.008)	0.638	0.048	0.025
Aspire more education	0.010(0.016)	-0.001(0.016)	0.013(0.016)	0.779	0.212	0.228
Life ladder now	-0.125(0.054)**	-0.042(0.054)	-0.088(0.053)*	0.099	3.471	3.448
Depression scale (PHQ-8)	0.088(0.129)	0.115(0.127)	0.065(0.128)	0.829	3.882	4.340
GAD-7 Anxiety	-0.000(0.123)	0.058(0.125)	-0.056(0.125)	0.826	3.307	3.536
Gender equitable index	-0.018(0.012)	-0.016(0.012)	-0.004(0.012)	0.367	0.248	0.364
Influences at least some decisions	-0.006(0.007)	0.005(0.007)	0.000(0.007)	0.513	0.757	0.746
Can moderately make own decisions	-0.009(0.008)	-0.000(0.008)	-0.012(0.008)	0.275	0.781	0.756

Each row represents a separate regression, where the variable listed is regressed against treatment dummies. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. The first three columns display the coefficients associated with these treatment dummies, illustrating the difference between a specific treatment and the control arm. We additionally test that these coefficients are jointly zero, or in other words, the three treatment arms are jointly similar to the control group. The last two columns are mean values for the control group, separately for men and women. Heteroskedasticity robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.2: SES Training Uptake

	(1) Attended any session	(2) Attended atleast 80% of sessions	(3) Attended all sessions	(4) Proportion of sessions attended
Awareness (A)	-0.052 (0.018)***	0.061 (0.022)***	0.110 (0.027)***	-0.004 (0.018)
Women × Awareness (A)	0.006 (0.024)	-0.018 (0.030)	-0.038 (0.037)	0.009 (0.025)
Management (M)	-0.034 (0.017)*	0.074 (0.023)***	0.129 (0.027)***	0.011 (0.018)
Women × Management (M)	-0.014 (0.024)	-0.015 (0.030)	-0.068 (0.037)*	-0.004 (0.025)
A + Women × A	-0.046 (0.016)***	0.043 (0.021)**	0.073 (0.025)***	0.005 (0.017)
M + Women × M	-0.048 (0.016)***	0.058 (0.020)***	0.062 (0.025)**	0.007 (0.017)
Men: P-value (A=M)	0.327	0.564	0.470	0.408
Women: P-value (A=M)	0.904	0.438	0.653	0.888
Women A+M Mean	0.929	0.802	0.669	0.839
Men A+M Mean	0.905	0.748	0.553	0.803
Observations	3,549	3,549	3,549	3,549

Data in this table are derived from training attendance administrative records. Dependent variables in this table are dummies, except Column 4. Column 1 indicates whether a participant attended any allocated sessions. Column 2 indicates whether a participant attended at least four days out of the five anticipated days in each single arm or at least eight out of the ten anticipated days in the combined arm. Column 3 indicates whether a participant attended all five days in the single arm or all ten days in the combined arm. In Column 4, we express the number of sessions attended by each participant as a fraction of the training duration, that is, five days for each of the single arms and ten days for the combined arm. Each dependent variable listed in a corresponding column is regressed against treatment dummies. Since there was no uptake among the control group, we restricted the analysis to participants who were randomised to receive SES training. Therefore, the omitted category is those receiving the combined treatment. The coefficients therefore show the difference in training uptake among those randomised to receive the single arms relative to those who were randomised to receive the combined arm. We additionally interact the regressors with the gender dummy to explore the differential uptake between men and women. Each regression also includes enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.3: Overconfidence and Job Seeking Behaviour at Baseline

	(1) Overconfidence All SES	(2) Overconfidence All SES	(3) Searched for job	(4) Searched for job
Overconfidence: All SES			0.092 (0.024)***	0.102 (0.038)***
Beh: All SES	-0.024 (0.005)***	-0.020 (0.007)***	0.014 (0.010)	0.020 (0.009)**
Cognitive ability index	0.107 (0.018)***	0.065 (0.025)**	-0.012 (0.032)	0.000 (0.041)
Years of education	0.009 (0.002)***	0.007 (0.002)***	0.004 (0.002)*	0.000 (0.004)
Social desirability	0.125 (0.011)***	0.138 (0.013)***	-0.029 (0.019)	-0.007 (0.029)
Age	0.009 (0.001)***	0.012 (0.002)***	0.016 (0.002)***	0.017 (0.004)***
Women	-0.047 (0.008)***	0.018 (0.087)	-0.091 (0.016)***	0.039 (0.154)
Women × Overconfidence: All SES				-0.024 (0.055)
Women × Beh: All SES		-0.010 (0.009)		-0.012 (0.013)
Women × Cognitive ability index		0.086 (0.033)**		-0.025 (0.056)
Women × Social desirability		-0.023 (0.017)		-0.045 (0.033)
Women × Years of education		0.004 (0.003)		0.007 (0.005)
Women × Age		-0.004 (0.003)		-0.001 (0.005)
Observations	4,431	4,431	4,431	4,431

Estimates in this table correspond to correlations at baseline. Each column is an extract from a single regression, where the variable listed is regressed against the independent variables listed in rows. Overconfidence is a continuous variable indicating the residuals obtained by regressing percentile ranks from self-reported measures against percentile ranks from behavioural measures following Anderson et al., 2012; Adamecz-Völgyi and Shure, 2022. Job search, on the other hand, is a dummy variable indicating whether the respondent searched for a job in the three months preceding the survey, derived from an aggregate of various job search activities, including submitting applications, inquiring about potential vacancies, and networking with others. Cognitive ability is the proportion of correct responses from the Raven's Progressive Matrices test, while social desirability is an index aggregating the eight-item Balanced Inventory of Desirable Responding scale. In odd-numbered columns, we pool the correlations for the entire sample, while in even-numbered columns, we allow the correlates to have differential slopes by respondent gender. Each regression additionally controls for enumerator fixed effects. Standard errors shown in parentheses are clustered at the street level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.4: Effect of SES Training on Aggregate SES by Sex

	Midline			Endline		
	(1)	(2)	(3)	(4)	(5)	(6)
	All SES	Awareness	Management	All SES	Awareness	Management
Awareness (A)	0.083 (0.052)	0.058 (0.053)	0.091 (0.052)*	0.046 (0.046)	0.034 (0.047)	0.050 (0.046)
	[0.265]	[0.501]	[0.233]	[0.887]	[0.924]	[0.887]
Women × Awareness (A)	0.027 (0.074)	0.060 (0.075)	0.009 (0.075)	-0.045 (0.063)	-0.044 (0.065)	-0.042 (0.064)
Management (M)	0.117 (0.052)**	0.102 (0.052)*	0.117 (0.052)**	0.058 (0.045)	0.050 (0.047)	0.060 (0.045)
	[0.100]	[0.168]	[0.100]	[0.887]	[0.887]	[0.887]
Women × Management (M)	0.037 (0.075)	0.051 (0.074)	0.028 (0.075)	-0.031 (0.063)	-0.057 (0.066)	-0.016 (0.063)
A + M	0.134 (0.053)**	0.127 (0.056)**	0.129 (0.053)**	0.002 (0.047)	0.009 (0.049)	-0.002 (0.047)
	[0.100]	[0.100]	[0.100]	[0.978]	[0.978]	[0.978]
Women × A + M	-0.074 (0.074)	-0.077 (0.077)	-0.067 (0.074)	0.026 (0.064)	0.001 (0.066)	0.039 (0.065)
A + Women × A	0.110 (0.052)**	0.118 (0.052)**	0.099 (0.053)*	0.001 (0.044)	-0.010 (0.045)	0.007 (0.044)
	[0.126]	[0.100]	[0.181]	[0.978]	[0.978]	[0.978]
M + Women × M	0.153 (0.054)***	0.152 (0.053)***	0.145 (0.054)***	0.027 (0.043)	-0.007 (0.045)	0.044 (0.044)
	[0.076]	[0.076]	[0.090]	[0.924]	[0.978]	[0.887]
A + M + Women × A + M	0.060 (0.052)	0.050 (0.053)	0.062 (0.052)	0.029 (0.044)	0.009 (0.044)	0.037 (0.045)
	[0.461]	[0.567]	[0.461]	[0.924]	[0.978]	[0.887]
Men: P-value (A=M=A+M)	0.628	0.453	0.763	0.443	0.692	0.350
Women: P-value (A=M=A+M)	0.232	0.176	0.311	0.785	0.896	0.669
Women Control Mean	-0.024	-0.018	-0.025	-0.077	-0.066	-0.079
Men Control Mean	0.022	0.016	0.023	0.079	0.067	0.081
Observations	4,607	4,607	4,607	4,607	4,607	4,607

Dependent variables in each column are standardised values of corresponding SES (sub) aggregates derived from self-reported measures at a corresponding survey wave. Each self-reported SES score is standardised relative to the values of the control group, separately for each survey wave. Columns 1-3 refer to midline estimates while 4–6 refer to estimates at endline. For each survey wave, we interact with the treatment dummy with the gender dummy. This enables us to assess the equality of treatment effects reported across our three treatment arms for both men and women. Subsequently, we report the p-value associated with this Wald test. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.5: Aggregate Behavioural SES Skills Heterogeneity by Awareness SES (Beh) above median

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	Beh: All SES	Beh: Awareness	Beh: Management	Beh: All SES	Beh: Awareness	Beh: Management
Panel A: Midline						
Awareness (A)	-0.027 (0.057)	0.037 (0.055)	-0.049 (0.059)	0.039 (0.058)	0.068 (0.059)	-0.001 (0.058)
	[0.713]	[0.642]	[0.609]	[0.886]	[0.886]	[0.981]
Awareness SES (Beh) above median × A	0.090 (0.079)	0.023 (0.081)	0.110 (0.080)	-0.024 (0.083)	-0.126 (0.082)	0.045 (0.084)
	[0.912]	[0.408]	[0.452]	[0.939]	[0.939]	[0.939]
Management (M)	0.007 (0.059)	0.060 (0.058)	-0.031 (0.060)	0.006 (0.058)	-0.002 (0.058)	-0.016 (0.058)
	[0.529]	[0.529]	[0.529]	[0.898]	[0.886]	[0.886]
Awareness SES (Beh) above median × M	0.053 (0.083)	-0.022 (0.081)	0.094 (0.085)	0.024 (0.080)	-0.009 (0.081)	0.061 (0.081)
	[0.071]	[0.166]	[0.071]	[0.949]	[0.939]	[0.939]
A + M	-0.006 (0.057)	0.064 (0.059)	-0.055 (0.059)	0.051 (0.061)	0.066 (0.059)	0.027 (0.062)
	[0.529]	[0.779]	[0.529]	[0.898]	[0.886]	[0.886]
Awareness SES (Beh) above median × A + M	0.141 (0.081)*	0.029 (0.082)	0.188 (0.083)**	-0.040 (0.082)	-0.086 (0.084)	-0.004 (0.084)
	[0.166]	[0.810]	[0.071]	[0.939]	[0.939]	[0.965]
Awareness SES (Beh) above median	-0.066 (0.059)	-0.000 (0.067)	-0.097 (0.059)*	0.038 (0.062)	0.097 (0.070)	0.011 (0.062)
A + Awareness SES (Beh) above median × A	0.063 (0.054)	0.060 (0.057)	0.061 (0.054)	0.015 (0.059)	-0.058 (0.057)	0.044 (0.060)
	[0.911]	[0.683]	[0.780]	[0.977]	[0.977]	[0.977]
M + Awareness SES (Beh) above median × M	0.060 (0.058)	0.038 (0.056)	0.063 (0.061)	0.030 (0.056)	-0.011 (0.055)	0.045 (0.057)
	[0.683]	[0.780]	[0.683]	[0.977]	[0.977]	[0.977]
A + M + Awareness SES (Beh) above median × A + M	0.135 (0.055)**	0.094 (0.056)*	0.133 (0.057)**	0.011 (0.054)	-0.020 (0.058)	0.024 (0.055)
	[0.780]	[0.881]	[0.683]	[0.977]	[0.977]	[0.977]
Control mean:Awareness SES (Beh) below median	-0.046	-0.136	0.026	-0.014	-0.001	-0.019
Panel B: Endline						
Awareness (A)	-0.023 (0.053)	0.035 (0.053)	-0.052 (0.053)	-0.079 (0.053)	0.034 (0.049)	-0.126 (0.055)**
	[0.830]	[0.773]	[0.698]	[0.420]	[0.694]	[0.187]
Awareness SES (Beh) above median × A	0.072 (0.078)	-0.016 (0.077)	0.103 (0.079)	0.096 (0.079)	-0.059 (0.070)	0.160 (0.081)**
	[0.970]	[0.970]	[0.970]	[0.965]	[0.965]	[0.965]
Management (M)	-0.034 (0.054)	0.050 (0.052)	-0.071 (0.056)	-0.035 (0.054)	-0.004 (0.051)	-0.046 (0.055)
	[0.698]	[0.830]	[0.698]	[0.770]	[0.694]	[0.694]
Awareness SES (Beh) above median × M	0.032 (0.077)	-0.031 (0.075)	0.058 (0.079)	0.077 (0.076)	-0.024 (0.072)	0.123 (0.078)
	[0.970]	[0.970]	[0.970]	[0.965]	[0.965]	[0.965]
A + M	0.028 (0.054)	0.030 (0.055)	0.015 (0.055)	-0.015 (0.059)	-0.008 (0.053)	-0.006 (0.059)
	[0.698]	[0.831]	[0.698]	[0.503]	[0.694]	[0.222]
Awareness SES (Beh) above median × A + M	0.002 (0.078)	-0.014 (0.078)	0.016 (0.079)	0.012 (0.083)	-0.045 (0.077)	0.035 (0.083)
	[0.981]	[0.970]	[0.970]	[0.965]	[0.965]	[0.965]
Awareness SES (Beh) above median	0.038 (0.056)	0.037 (0.065)	0.021 (0.054)	-0.058 (0.057)	-0.008 (0.062)	-0.078 (0.055)
A + Awareness SES (Beh) above median × A	0.049 (0.057)	0.018 (0.055)	0.051 (0.057)	0.017 (0.057)	-0.025 (0.050)	0.034 (0.060)
	[0.916]	[0.916]	[0.916]	[0.731]	[0.934]	[0.731]
M + Awareness SES (Beh) above median × M	-0.002 (0.054)	0.019 (0.053)	-0.013 (0.055)	0.042 (0.054)	-0.029 (0.050)	0.077 (0.055)
	[0.969]	[0.916]	[0.916]	[0.731]	[0.731]	[0.728]
A + M + Awareness SES (Beh) above median × A + M	0.030 (0.055)	0.016 (0.054)	0.031 (0.056)	-0.003 (0.058)	-0.053 (0.053)	0.030 (0.058)
	[0.916]	[0.916]	[0.916]	[0.731]	[0.826]	[0.728]
Control mean:Awareness SES (Beh) below median	-0.029	-0.051	-0.011	0.001	0.012	-0.000

Dependent variables in each column are standardised values of corresponding SES (sub)aggregates derived from behavioural measures at a corresponding survey wave. Each behavioural SES score is standardised relative to the values of the control group, separately for each survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.6: Aggregate Behavioural SES Skills Heterogeneity by Management SES (Beh) above median

	Men			Women		
	(1) Beh: All SES	(2) Beh: Awareness	(3) Beh: Management	(4) Beh: All SES	(5) Beh: Awareness	(6) Beh: Management
Panel A: Midline						
Awareness (A)	-0.009 (0.054)	0.039 (0.056)	-0.016 (0.056)	0.014 (0.056)	0.025 (0.058)	-0.004 (0.058)
	[0.861]	[0.810]	[0.861]	[0.906]	[0.906]	[0.945]
Management SES (Beh) above median × A	0.058 (0.077)	0.020 (0.080)	0.051 (0.079)	0.024 (0.085)	-0.044 (0.083)	0.049 (0.086)
	[0.634]	[0.368]	[0.822]	[0.644]	[0.699]	[0.699]
Management (M)	0.039 (0.060)	0.059 (0.056)	0.031 (0.061)	-0.030 (0.059)	-0.067 (0.058)	-0.016 (0.061)
	[0.810]	[0.810]	[0.810]	[0.906]	[0.906]	[0.906]
Management SES (Beh) above median × M	-0.005 (0.083)	-0.018 (0.081)	-0.020 (0.086)	0.091 (0.082)	0.112 (0.081)	0.060 (0.084)
	[0.634]	[0.634]	[0.634]	[0.431]	[0.431]	[0.431]
A + M	0.058 (0.057)	0.113 (0.055)**	0.019 (0.060)	-0.039 (0.057)	-0.023 (0.059)	-0.025 (0.058)
	[0.810]	[0.861]	[0.810]	[0.906]	[0.906]	[0.906]
Management SES (Beh) above median × A + M	0.018 (0.078)	-0.065 (0.081)	0.050 (0.081)	0.134 (0.083)	0.089 (0.084)	0.099 (0.084)
	[0.822]	[0.634]	[0.695]	[0.431]	[0.431]	[0.431]
Management SES (Beh) above median	-0.080 (0.062)	-0.004 (0.059)	-0.080 (0.068)	-0.042 (0.068)	-0.026 (0.058)	-0.014 (0.073)
A + Management SES (Beh) above median × A	0.049 (0.055)	0.059 (0.056)	0.035 (0.056)	0.038 (0.062)	-0.019 (0.059)	0.045 (0.062)
	[0.948]	[0.948]	[0.948]	[0.688]	[0.611]	[0.790]
M + Management SES (Beh) above median × M	0.034 (0.057)	0.041 (0.057)	0.011 (0.060)	0.061 (0.056)	0.045 (0.056)	0.044 (0.056)
	[0.948]	[0.948]	[0.948]	[0.611]	[0.611]	[0.611]
A + M + Management SES (Beh) above median × A + M	0.076 (0.053)	0.048 (0.058)	0.068 (0.055)	0.095 (0.058)	0.066 (0.058)	0.073 (0.060)
	[0.948]	[0.948]	[0.948]	[0.611]	[0.611]	[0.611]
Control mean:Management SES (Beh) below median	-0.007	-0.053	0.024	-0.038	0.024	-0.062
Panel B: Endline						
Awareness (A)	-0.000 (0.055)	0.007 (0.054)	-0.005 (0.053)	-0.095 (0.053)*	-0.024 (0.048)	-0.111 (0.054)**
	[0.995]	[0.995]	[0.995]	[0.218]	[0.696]	[0.218]
Management SES (Beh) above median × A	0.025 (0.078)	0.033 (0.078)	0.013 (0.079)	0.132 (0.078)*	0.062 (0.071)	0.134 (0.081)*
	[0.981]	[0.981]	[0.981]	[0.910]	[0.910]	[0.910]
Management (M)	-0.036 (0.055)	0.012 (0.054)	-0.057 (0.056)	-0.017 (0.055)	-0.036 (0.052)	0.007 (0.056)
	[0.995]	[0.995]	[0.995]	[0.667]	[0.667]	[0.697]
Management SES (Beh) above median × M	0.034 (0.076)	0.040 (0.075)	0.027 (0.078)	0.044 (0.077)	0.038 (0.073)	0.022 (0.079)
	[0.981]	[0.981]	[0.981]	[0.910]	[0.910]	[0.910]
A + M	0.038 (0.056)	0.043 (0.058)	0.021 (0.056)	-0.023 (0.057)	-0.068 (0.053)	0.022 (0.058)
	[0.995]	[0.995]	[0.995]	[0.218]	[0.667]	[0.218]
Management SES (Beh) above median × A + M	-0.021 (0.078)	-0.042 (0.078)	0.004 (0.080)	0.031 (0.083)	0.074 (0.075)	-0.015 (0.085)
	[0.981]	[0.981]	[0.981]	[0.910]	[0.910]	[0.910]
Management SES (Beh) above median	-0.075 (0.062)	-0.051 (0.056)	-0.018 (0.068)	-0.051 (0.062)	-0.021 (0.050)	-0.024 (0.067)
A + Management SES (Beh) above median × A	0.025 (0.056)	0.040 (0.056)	0.008 (0.058)	0.037 (0.057)	0.038 (0.050)	0.023 (0.060)
	[0.928]	[0.928]	[0.928]	[0.959]	[0.959]	[0.959]
M + Management SES (Beh) above median × M	-0.002 (0.052)	0.052 (0.051)	-0.029 (0.054)	0.026 (0.054)	0.003 (0.050)	0.029 (0.055)
	[0.974]	[0.928]	[0.928]	[0.959]	[0.959]	[0.959]
A + M + Management SES (Beh) above median × A + M	0.017 (0.054)	0.001 (0.051)	0.024 (0.055)	0.009 (0.059)	0.006 (0.052)	0.007 (0.060)
	[0.928]	[0.928]	[0.928]	[0.959]	[0.959]	[0.959]
Control mean:Management SES (Beh) below median	-0.025	-0.070	0.005	0.031	0.074	0.004

Dependent variables in each column are standardised values of corresponding SES (sub)aggregates derived from behavioural measures at a corresponding survey wave. Each behavioural SES score is standardised relative to the values of the control group, separately for each survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.7: Aggregate Behavioural SES Skills Heterogeneity by Cognition above median

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	Beh: All SES	Beh: Awareness	Beh: Management	Beh: All SES	Beh: Awareness	Beh: Management
Panel A: Midline						
Awareness (A)	-0.022 (0.065)	0.089 (0.067)	-0.057 (0.065)	-0.026 (0.065)	-0.022 (0.061)	-0.046 (0.067)
	[0.738]	[0.582]	[0.582]	[0.719]	[0.719]	[0.719]
Cognition above median × A	0.067 (0.082)	-0.064 (0.085)	0.105 (0.083)	0.090 (0.085)	0.046 (0.083)	0.113 (0.086)
	[0.288]	[0.833]	[0.281]	[0.586]	[0.641]	[0.523]
Management (M)	0.009 (0.068)	0.026 (0.065)	0.010 (0.067)	-0.077 (0.066)	-0.107 (0.063)*	-0.064 (0.067)
	[0.582]	[0.693]	[0.582]	[0.651]	[0.719]	[0.651]
Cognition above median × M	0.042 (0.085)	0.038 (0.083)	0.016 (0.085)	0.153 (0.086)*	0.162 (0.083)*	0.128 (0.087)
	[0.031]	[0.037]	[0.066]	[0.365]	[0.470]	[0.365]
A + M	-0.085 (0.075)	-0.015 (0.073)	-0.091 (0.074)	-0.042 (0.066)	-0.029 (0.061)	-0.057 (0.069)
	[0.582]	[0.582]	[0.582]	[0.651]	[0.719]	[0.651]
Cognition above median × A + M	0.219 (0.088)**	0.137 (0.088)	0.193 (0.088)**	0.120 (0.086)	0.084 (0.083)	0.136 (0.088)
	[0.037]	[0.183]	[0.065]	[0.365]	[0.470]	[0.365]
Cognition above median	-0.042 (0.060)	-0.028 (0.062)	-0.026 (0.060)	-0.061 (0.063)	-0.031 (0.058)	-0.060 (0.065)
A + Cognition above median × A	0.045 (0.049)	0.025 (0.050)	0.048 (0.051)	0.064 (0.053)	0.024 (0.056)	0.067 (0.053)
	[0.891]	[0.891]	[0.891]	[0.308]	[0.269]	[0.339]
M + Cognition above median × M	0.051 (0.051)	0.064 (0.050)	0.026 (0.054)	0.075 (0.053)	0.055 (0.053)	0.063 (0.053)
	[0.891]	[0.891]	[0.891]	[0.274]	[0.337]	[0.308]
A + M + Cognition above median × A + M	0.134 (0.046)***	0.121 (0.048)**	0.102 (0.049)**	0.077 (0.052)	0.056 (0.055)	0.079 (0.053)
	[0.891]	[0.891]	[0.891]	[0.269]	[0.269]	[0.274]
Control mean:Cognition below median	-0.054	-0.075	-0.026	0.089	0.106	0.050
Panel B: Endline						
Awareness (A)	-0.055 (0.062)	-0.069 (0.065)	-0.049 (0.062)	-0.052 (0.059)	-0.011 (0.053)	-0.059 (0.059)
	[0.478]	[0.466]	[0.480]	[0.842]	[0.842]	[0.842]
Cognition above median × A	0.111 (0.080)	0.152 (0.081)*	0.082 (0.080)	0.038 (0.078)	0.029 (0.070)	0.022 (0.080)
	[0.996]	[0.996]	[0.996]	[0.613]	[0.686]	[0.530]
Management (M)	0.086 (0.063)	0.063 (0.062)	0.073 (0.063)	0.046 (0.060)	-0.052 (0.053)	0.076 (0.063)
	[0.466]	[0.355]	[0.523]	[0.842]	[0.842]	[0.842]
Cognition above median × M	-0.163 (0.080)**	-0.046 (0.078)	-0.182 (0.081)**	-0.070 (0.078)	0.059 (0.072)	-0.099 (0.081)
	[0.996]	[0.996]	[0.996]	[0.613]	[0.642]	[0.642]
A + M	0.063 (0.066)	-0.000 (0.069)	0.075 (0.066)	0.059 (0.062)	-0.030 (0.060)	0.090 (0.061)
	[0.466]	[0.355]	[0.466]	[0.842]	[0.842]	[0.842]
Cognition above median × A + M	-0.054 (0.082)	0.027 (0.084)	-0.078 (0.083)	-0.111 (0.082)	-0.001 (0.076)	-0.128 (0.082)
	[0.996]	[0.996]	[0.996]	[0.530]	[0.985]	[0.530]
Cognition above median	0.066 (0.055)	0.049 (0.059)	0.055 (0.056)	0.066 (0.054)	0.016 (0.050)	0.078 (0.055)
A + Cognition above median × A	0.056 (0.050)	0.083 (0.047)*	0.033 (0.051)	-0.014 (0.051)	0.018 (0.046)	-0.036 (0.054)
	[0.312]	[0.404]	[0.369]	[0.670]	[0.670]	[0.670]
M + Cognition above median × M	-0.078 (0.048)	0.016 (0.046)	-0.108 (0.050)**	-0.024 (0.050)	0.007 (0.048)	-0.023 (0.051)
	[0.242]	[0.723]	[0.124]	[0.737]	[0.885]	[0.737]
A + M + Cognition above median × A + M	0.008 (0.048)	0.027 (0.046)	-0.003 (0.049)	-0.052 (0.054)	-0.032 (0.047)	-0.037 (0.055)
	[0.124]	[0.622]	[0.124]	[0.670]	[0.670]	[0.670]
Control mean:Cognition below median	-0.030	-0.093	0.013	-0.003	0.058	-0.032

Dependent variables in each column are standardised values of corresponding SES (sub)aggregates derived from behavioural measures at a corresponding survey wave. Each behavioural SES score is standardised relative to the values of the control group, separately for each survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.8: Aggregate Behavioural SES Skills Heterogeneity by Searching for a job

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	Beh: All SES	Beh: Awareness	Beh: Management	Beh: All SES	Beh: Awareness	Beh: Management
Panel A: Midline						
Awareness (A)	0.048 (0.058)	0.103 (0.058)*	0.019 (0.058)	0.069 (0.057)	0.126 (0.057)**	0.011 (0.057)
	[0.965]	[0.687]	[0.965]	[0.504]	[0.087]	[0.848]
Searching for a job × A	-0.046 (0.077)	-0.098 (0.080)	-0.013 (0.080)	-0.089 (0.083)	-0.251 (0.081)***	0.018 (0.085)
	[0.604]	[0.532]	[0.821]	[0.945]	[0.945]	[0.945]
Management (M)	0.066 (0.059)	0.042 (0.058)	0.066 (0.062)	0.046 (0.057)	0.030 (0.055)	0.040 (0.058)
	[0.965]	[0.965]	[0.965]	[0.848]	[0.087]	[0.848]
Searching for a job × M	-0.051 (0.080)	0.016 (0.081)	-0.081 (0.082)	-0.065 (0.082)	-0.083 (0.081)	-0.057 (0.083)
	[0.532]	[0.532]	[0.604]	[0.945]	[0.945]	[0.945]
A + M	0.063 (0.061)	0.084 (0.062)	0.038 (0.064)	0.053 (0.058)	0.031 (0.058)	0.036 (0.060)
	[0.965]	[0.965]	[0.965]	[0.511]	[0.018]	[0.848]
Searching for a job × A + M	0.011 (0.080)	-0.008 (0.084)	0.013 (0.083)	-0.049 (0.082)	-0.021 (0.084)	-0.024 (0.083)
	[0.921]	[0.921]	[0.921]	[0.945]	[0.945]	[0.945]
Searching for a job	0.056 (0.057)	0.056 (0.060)	0.046 (0.058)	0.038 (0.061)	0.084 (0.058)	-0.006 (0.062)
A + Searching for a job × A	0.002 (0.052)	0.006 (0.053)	0.006 (0.054)	-0.020 (0.060)	-0.125 (0.057)**	0.029 (0.062)
	[0.735]	[0.789]	[0.735]	[0.745]	[0.751]	[0.745]
M + Searching for a job × M	0.015 (0.056)	0.058 (0.055)	-0.014 (0.057)	-0.018 (0.058)	-0.053 (0.059)	-0.017 (0.058)
	[0.848]	[0.735]	[0.848]	[0.767]	[0.745]	[0.767]
A + M + Searching for a job × A + M	0.074 (0.051)	0.075 (0.054)	0.051 (0.053)	0.004 (0.057)	0.010 (0.059)	0.013 (0.057)
	[0.789]	[0.848]	[0.735]	[0.745]	[0.745]	[0.745]
Control mean: Not searching for a job	-0.139	-0.145	-0.087	0.007	0.051	-0.025
Panel B: Endline						
Awareness (A)	-0.009 (0.058)	-0.031 (0.057)	0.010 (0.059)	-0.030 (0.052)	-0.001 (0.048)	-0.041 (0.054)
	[0.991]	[0.991]	[0.991]	[0.991]	[0.991]	[0.991]
Searching for a job × A	0.044 (0.080)	0.103 (0.079)	-0.011 (0.081)	-0.002 (0.079)	0.014 (0.071)	-0.012 (0.081)
	[0.867]	[0.867]	[0.867]	[0.965]	[0.965]	[0.965]
Management (M)	0.050 (0.060)	0.024 (0.056)	0.058 (0.061)	-0.050 (0.053)	-0.058 (0.049)	-0.034 (0.055)
	[0.991]	[0.867]	[0.991]	[0.991]	[0.991]	[0.991]
Searching for a job × M	-0.120 (0.078)	0.017 (0.076)	-0.177 (0.080)**	0.116 (0.077)	0.090 (0.073)	0.109 (0.079)
	[0.867]	[0.867]	[0.988]	[0.965]	[0.965]	[0.965]
A + M	0.025 (0.059)	-0.033 (0.060)	0.053 (0.058)	-0.023 (0.058)	-0.055 (0.053)	0.012 (0.058)
	[0.991]	[0.867]	[0.991]	[0.991]	[0.991]	[0.991]
Searching for a job × A + M	0.007 (0.080)	0.097 (0.081)	-0.052 (0.080)	0.033 (0.080)	0.052 (0.074)	0.004 (0.081)
	[0.988]	[0.867]	[0.867]	[0.965]	[0.965]	[0.965]
Searching for a job	0.035 (0.055)	-0.027 (0.057)	0.069 (0.056)	-0.035 (0.055)	-0.046 (0.051)	-0.018 (0.056)
A + Searching for a job × A	0.036 (0.054)	0.071 (0.053)	-0.001 (0.054)	-0.032 (0.058)	0.013 (0.051)	-0.054 (0.060)
	[0.531]	[0.759]	[0.531]	[0.440]	[0.366]	[0.556]
M + Searching for a job × M	-0.070 (0.049)	0.041 (0.050)	-0.118 (0.051)**	0.065 (0.056)	0.031 (0.054)	0.075 (0.056)
	[0.347]	[0.531]	[0.120]	[0.366]	[0.556]	[0.366]
A + M + Searching for a job × A + M	0.032 (0.052)	0.064 (0.051)	0.001 (0.053)	0.010 (0.056)	-0.003 (0.051)	0.016 (0.057)
	[0.347]	[0.819]	[0.120]	[0.366]	[0.366]	[0.366]
Control mean: Not searching for a job	-0.029	-0.042	-0.021	0.004	0.056	-0.026

Dependent variables in each column are standardised values of corresponding SES (sub)aggregates derived from behavioural measures at a corresponding survey wave. Each behavioural SES score is standardised relative to the values of the control group, separately for each survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.9: Aggregate Self-reported SES Skills Heterogeneity by Social desirability above median

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	SR: All SES	SR: Awareness	SR: Management	SR: All SES	SR: Awareness	SR: Management
Panel A: Midline						
Awareness (A)	0.187 (0.077)**	0.134 (0.079)*	0.203 (0.078)***	-0.026 (0.080)	0.023 (0.082)	-0.049 (0.080)
	[0.070]	[0.162]	[0.070]	[0.781]	[0.781]	[0.697]
Social desirability above median × A	-0.186 (0.105)*	-0.146 (0.108)	-0.195 (0.106)*	0.204 (0.108)*	0.135 (0.108)	0.227 (0.109)**
	[0.059]	[0.059]	[0.059]	[0.875]	[0.875]	[0.875]
Management (M)	0.145 (0.081)*	0.133 (0.079)*	0.142 (0.081)*	0.099 (0.085)	0.087 (0.085)	0.099 (0.084)
	[0.980]	[0.980]	[0.980]	[0.057]	[0.064]	[0.057]
Social desirability above median × M	-0.046 (0.110)	-0.069 (0.110)	-0.032 (0.111)	0.087 (0.111)	0.101 (0.112)	0.076 (0.111)
	[0.361]	[0.381]	[0.361]	[0.805]	[0.805]	[0.805]
A + M	0.203 (0.082)**	0.197 (0.084)**	0.193 (0.081)**	0.012 (0.077)	-0.022 (0.080)	0.028 (0.076)
	[0.162]	[0.266]	[0.162]	[0.105]	[0.317]	[0.083]
Social desirability above median × A + M	-0.111 (0.110)	-0.126 (0.115)	-0.096 (0.110)	0.062 (0.107)	0.103 (0.111)	0.038 (0.106)
	[0.381]	[0.381]	[0.381]	[0.875]	[0.805]	[0.875]
Social desirability above median	0.082 (0.076)	0.075 (0.078)	0.084 (0.076)	-0.068 (0.073)	-0.093 (0.074)	-0.050 (0.074)
A + Social desirability above median × A	0.002 (0.071)	-0.012 (0.074)	0.008 (0.071)	0.178 (0.070)**	0.158 (0.069)**	0.178 (0.071)**
	[0.278]	[0.278]	[0.278]	[0.435]	[0.461]	[0.435]
M + Social desirability above median × M	0.099 (0.071)	0.064 (0.074)	0.110 (0.071)	0.186 (0.071)***	0.188 (0.070)***	0.175 (0.072)**
	[0.298]	[0.578]	[0.281]	[0.037]	[0.037]	[0.044]
A + M + Social desirability above median × A + M	0.093 (0.072)	0.071 (0.076)	0.097 (0.072)	0.074 (0.071)	0.081 (0.073)	0.066 (0.072)
	[0.760]	[0.681]	[0.773]	[0.487]	[0.471]	[0.495]
Control mean:Social desirability below median	-0.052	-0.053	-0.049	-0.009	0.021	-0.023
Panel B: Endline						
Awareness (A)	0.061 (0.074)	0.039 (0.077)	0.070 (0.074)	-0.045 (0.067)	-0.059 (0.068)	-0.037 (0.068)
	[0.906]	[0.906]	[0.906]	[0.589]	[0.589]	[0.589]
Social desirability above median × A	-0.024 (0.096)	-0.008 (0.100)	-0.032 (0.096)	0.084 (0.089)	0.092 (0.091)	0.078 (0.090)
	[0.914]	[0.914]	[0.914]	[0.640]	[0.659]	[0.640]
Management (M)	0.095 (0.072)	0.087 (0.075)	0.095 (0.071)	-0.017 (0.067)	-0.054 (0.072)	0.003 (0.067)
	[0.906]	[0.906]	[0.906]	[0.589]	[0.589]	[0.589]
Social desirability above median × M	-0.080 (0.095)	-0.083 (0.100)	-0.074 (0.095)	0.070 (0.089)	0.074 (0.094)	0.066 (0.089)
	[0.914]	[0.930]	[0.914]	[0.365]	[0.659]	[0.365]
A + M	0.023 (0.075)	0.019 (0.078)	0.023 (0.074)	-0.047 (0.068)	-0.029 (0.067)	-0.056 (0.071)
	[0.906]	[0.933]	[0.906]	[0.589]	[0.589]	[0.589]
Social desirability above median × A + M	-0.044 (0.096)	-0.024 (0.101)	-0.052 (0.096)	0.129 (0.090)	0.061 (0.091)	0.160 (0.093)*
	[0.914]	[0.914]	[0.914]	[0.365]	[0.640]	[0.365]
Social desirability above median	0.088 (0.070)	0.080 (0.073)	0.091 (0.069)	0.003 (0.062)	0.009 (0.062)	0.001 (0.065)
A + Social desirability above median × A	0.037 (0.059)	0.030 (0.061)	0.038 (0.059)	0.039 (0.058)	0.033 (0.061)	0.041 (0.059)
	[0.647]	[0.647]	[0.647]	[0.902]	[0.693]	[0.960]
M + Social desirability above median × M	0.015 (0.061)	0.004 (0.064)	0.020 (0.061)	0.053 (0.058)	0.020 (0.060)	0.069 (0.059)
	[0.903]	[0.949]	[0.903]	[0.693]	[0.902]	[0.693]
A + M + Social desirability above median × A + M	-0.021 (0.060)	-0.005 (0.062)	-0.029 (0.060)	0.082 (0.058)	0.032 (0.060)	0.104 (0.059)*
	[0.647]	[0.647]	[0.647]	[0.693]	[0.693]	[0.693]
Control mean:Social desirability below median	0.034	0.041	0.029	-0.084	-0.073	-0.087

Dependent variables in each column are standardised values of corresponding SES (sub)aggregates derived from self-reported measures at a corresponding survey wave. Each self-reported SES score is standardised relative to the values of the control group, separately for each survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.10: Aggregate Self-reported SES Skills Heterogeneity by Searching for a job

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	SR: All SES	SR: Awareness	SR: Management	SR: All SES	SR: Awareness	SR: Management
Panel A: Midline						
Awareness (A)	-0.078 (0.079)	-0.055 (0.080)	-0.086 (0.080)	0.090 (0.074)	0.088 (0.076)	0.084 (0.074)
	[0.362]	[0.493]	[0.360]	[0.453]	[0.453]	[0.453]
Searching for a job × A	0.293 (0.105)***	0.194 (0.108)*	0.325 (0.106)***	0.002 (0.104)	0.021 (0.103)	-0.007 (0.106)
	[0.649]	[0.350]	[0.746]	[0.981]	[0.981]	[0.981]
Management (M)	-0.000 (0.075)	0.037 (0.078)	-0.019 (0.074)	0.197 (0.073)***	0.157 (0.074)**	0.205 (0.073)***
	[0.006]	[0.098]	[0.005]	[0.453]	[0.453]	[0.453]
Searching for a job × M	0.216 (0.107)**	0.102 (0.109)	0.260 (0.107)**	-0.103 (0.108)	-0.028 (0.107)	-0.134 (0.110)
	[0.016]	[0.134]	[0.014]	[0.981]	[0.981]	[0.981]
A + M	0.054 (0.081)	0.103 (0.086)	0.026 (0.079)	0.047 (0.073)	0.043 (0.075)	0.046 (0.073)
	[0.012]	[0.107]	[0.006]	[0.984]	[0.984]	[0.984]
Searching for a job × A + M	0.158 (0.110)	0.045 (0.114)	0.205 (0.109)*	-0.002 (0.106)	-0.017 (0.109)	0.005 (0.106)
	[0.272]	[0.746]	[0.134]	[0.981]	[0.981]	[0.981]
Searching for a job	-0.187 (0.076)**	-0.153 (0.078)**	-0.192 (0.076)**	0.053 (0.072)	0.046 (0.071)	0.053 (0.074)
A + Searching for a job × A	0.215 (0.069)***	0.139 (0.072)*	0.240 (0.070)***	0.092 (0.073)	0.110 (0.070)	0.077 (0.075)
	[0.996]	[0.814]	[0.895]	[0.033]	[0.101]	[0.033]
M + Searching for a job × M	0.216 (0.074)***	0.139 (0.074)*	0.241 (0.075)***	0.094 (0.080)	0.129 (0.077)*	0.071 (0.081)
	[0.016]	[0.107]	[0.012]	[0.357]	[0.213]	[0.430]
A + M + Searching for a job × A + M	0.212 (0.073)***	0.148 (0.074)**	0.230 (0.073)***	0.045 (0.075)	0.026 (0.077)	0.051 (0.074)
	[0.096]	[0.523]	[0.045]	[0.430]	[0.793]	[0.357]
Control mean: Not searching for a job	0.075	0.061	0.078	-0.015	-0.008	-0.017
Panel B: Endline						
Awareness (A)	0.107 (0.069)	0.126 (0.072)*	0.090 (0.068)	0.022 (0.062)	0.011 (0.064)	0.026 (0.064)
	[0.365]	[0.365]	[0.419]	[0.858]	[0.858]	[0.858]
Searching for a job × A	-0.102 (0.096)	-0.162 (0.100)	-0.065 (0.096)	-0.039 (0.089)	-0.036 (0.091)	-0.038 (0.090)
	[0.803]	[0.803]	[0.803]	[0.959]	[0.967]	[0.959]
Management (M)	0.058 (0.070)	0.060 (0.075)	0.054 (0.069)	0.020 (0.062)	-0.012 (0.062)	0.037 (0.063)
	[0.941]	[0.757]	[0.776]	[0.858]	[0.858]	[0.858]
Searching for a job × M	-0.006 (0.093)	-0.028 (0.098)	0.005 (0.092)	0.007 (0.090)	0.000 (0.095)	0.010 (0.091)
	[0.803]	[0.803]	[0.803]	[0.959]	[0.959]	[0.959]
A + M	0.030 (0.074)	0.049 (0.078)	0.018 (0.073)	0.013 (0.064)	-0.003 (0.063)	0.021 (0.066)
	[0.516]	[0.365]	[0.746]	[0.858]	[0.858]	[0.858]
Searching for a job × A + M	-0.052 (0.098)	-0.074 (0.102)	-0.038 (0.097)	0.027 (0.090)	0.017 (0.090)	0.031 (0.092)
	[0.803]	[0.803]	[0.803]	[0.959]	[0.959]	[0.959]
Searching for a job	0.071 (0.069)	0.104 (0.072)	0.051 (0.068)	-0.011 (0.064)	-0.005 (0.064)	-0.013 (0.067)
A + Searching for a job × A	0.005 (0.064)	-0.036 (0.066)	0.026 (0.064)	-0.017 (0.062)	-0.025 (0.064)	-0.012 (0.063)
	[0.776]	[0.776]	[0.776]	[0.996]	[0.996]	[0.996]
M + Searching for a job × M	0.051 (0.060)	0.032 (0.063)	0.059 (0.061)	0.027 (0.064)	-0.012 (0.069)	0.047 (0.064)
	[0.776]	[0.915]	[0.776]	[0.996]	[0.996]	[0.996]
A + M + Searching for a job × A + M	-0.022 (0.062)	-0.025 (0.064)	-0.020 (0.062)	0.040 (0.062)	0.014 (0.063)	0.052 (0.064)
	[0.958]	[0.958]	[0.958]	[0.996]	[0.996]	[0.996]
Control mean: Not searching for a job	0.128	0.100	0.137	-0.076	-0.057	-0.082

Dependent variables in each column are standardised values of corresponding SES (sub)aggregates derived from self-reported measures at a corresponding survey wave. Each self-reported SES score is standardised relative to the values of the control group, separately for each survey wave. Panel A refers to estimates midline, while Panel B refers to estimates at the endline. For each survey wave, we interact with the treatment dummy with the baseline value of the specified covariate. We fit our regressions separately for the men and women sub-samples. This allows us to test the presence of treatment effects heterogeneity at various levels of the baseline covariate for both men and women. Regressions also include enumerator and randomisation strata fixed effects. Randomisation strata fixed effects included street, gender, and whether the respondent was older than the street-level median age. Heteroskedasticity robust standard errors in parentheses. Q-values adjusted for multiple hypothesis shown in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.B.11: Review of Selected SES Interventions

Program (Country)	Targetted SES	Length & Delivery	Study Design	Major Findings
SEED: Soft vs Hard Skills (Uganda) — Chioda et al. (2021)	Persuasion, negotiation, teamwork, grit, self-efficacy	Year-long school modules; class + practice (hrs n/r)	3-arm RCT HS students: (T1) Soft-heavy; (T2) Hard-heavy; (C) Control	Both ↑ earnings; soft-skills ↑ self-efficacy/persuasion
Negotiation for Girls (Zambia) — Ashraf et al. (2020)	Negotiation, perspective-taking, self-advocacy	6×2-hr sessions (12h); in-school workshops	2-arm RCT Grade-8 girls: Negotiation vs Control	↑ school continuation; improved bargaining
Grit/Patience Curriculum (Turkey) — Alan, Boneva, and Ertac (2019)	Grit, perseverance, patience	Full academic year; teacher-led	2-arm RCT Primary pupils: Grit–patience vs standard curriculum	↑ perseverance, task choice; ↑ test scores
Soft-skills Training (Jordan) — Groh et al. (2016)	Communication, teamwork, reliability, presentation	45 hrs; classroom, group	2-arm RCT Female graduates: Soft-skills vs no training	No employment impact; modest attitudinal gains
Soft-skills for Entrepreneurs (Jamaica) — Ubfal et al. (2022)	Personal initiative, perseverance, communication	Multi-week intensive course; hrs n/r	3-arm RCT micro-entrepreneurs: (T1) Soft; (T2) Soft + business; (C) Control	Short-run ↑ profits (men); fade by 12m; durable perseverance
On-the-job Soft Skills (India) — Adhvaryu, Kala, and Nyshadham (2023)	Communication, problem-solving, stress/time mgmt	80 hrs total; 2h/week × 12m; factory-based	2-arm RCT Female garment workers: training vs business-as-usual	↑ productivity 13–20%; spillovers; modest wage gains
VET: Hard vs Soft (Colombia) — Barrera-Osorio, Kugler, and Silliman (2023)	Communication, teamwork, self-efficacy	≈4m; class + practicum	3-arm RCT disadvantaged VET youth: (T1) Soft-heavy; (T2) Technical; (C) None	↑ formal jobs; soft-heavy arm more durable
Personal Initiative (Togo) — Campos et al. (2017)	Proactivity, opportunity recognition, planning	Multi-week class + feedback; hrs n/r	3-arm RCT Microentrepreneurs: (T1) PI; (T2) Business; (C) Control	↑ profits (30%) at 2y; durable 7y gains
Soft-skills, Networking & Workforce Entry (Rwanda) — Brudevold-Newman and Ubfal (2024)	Communication, networking, teamwork, empathy, initiative	13d; 160h (102h lectures + 58h practice); in-person	2-arm RCT Tertiary graduates: soft-skills training vs control	↑ job entry speed; mechanism: network use; improved SES scores
Skills Training & Business Outcomes (Liberia) — Dammert and Nansamba (2023)	Customer service, interpersonal, client focus, comm. (+)	(G1) 1d management; (G2) +1d interpersonal; local center	3-arm RCT: (T1) Biz mgmt; (T2) Biz mgmt + interpersonal; (C) Control	↑ customer focus & revenues; ↓ attrition; combined arm not stronger on profits

Notes: Arrows (↑/↓) denote increases/decreases. Hours = total training contact time.

Chapter 5

General Conclusion

Africa's youthful population presents a demographic dividend opportunity. Realising this dividend depends on efforts to equip young people with the skills to seize emerging opportunities. This dissertation examined three human capital interventions in East Africa, highlighting the complex ways such investments can shape—or fail to shape—economic and social outcomes.

The first empirical study evaluates the long-term intergenerational impacts of Kenya's Free Secondary Education (FSE) reform. While the policy successfully increased educational attainment, it did not generate discernible improvements in the next generation's health outcomes. The findings suggest that in contexts where schools drive access to education at the bottom of the distribution, the anticipated returns to education may be muted. This highlights the need to accompany supply expansion efforts with measures to strengthen the education quality.

The second empirical study investigates the influence of psychosocial constraints, including aspirations and gendered beliefs, on educational and marriage decisions among Somali youth. Exposure to successful and relatable role models, particularly female role models, led to more gender-equitable attitudes, enhanced transition to secondary school, and delays in early marriage. These results demonstrate the effectiveness of light-touch interventions in enhancing demand for education, particularly in fragile and highly constrained settings.

The third empirical study examines the effectiveness of socio-emotional skills training among young people not in education, employment, or training (NEET) in Tanzania. Training interventions substantially improved self-assessed socio-emotional skills but had limited impacts on behaviorally measured skills and labour market outcomes. Nonetheless, the SES training made men more overconfident about their abilities, facilitating positive labour market outcomes. These findings highlight the limitations of youth upskilling programs and underscore the need for greater attention to gender dynamics in designing and evaluating such interventions.

Bibliography

Abadie, Alberto and Guido W Imbens (2011). "Bias-corrected matching estimators for average treatment effects". *Journal of Business & Economic Statistics* 29.1, pp. 1–11.

Abdullahi, Arays, Mohamed Ali, Elijah Kipchumba, and Munshi Sulaiman (2023). "Supporting micro-enterprise in humanitarian programming: impact evaluation of business grants versus unconditional cash transfer". *Journal of African Economies* 32.4, pp. 415–437.

Abebe, Girum, A. Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, and Simon Quinn (2020). "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City". *The Review of Economic Studies* 88.3, pp. 1279–1310.

Abebe, Girum, Stefano A. Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, Simon Quinn, and Forhad Jahan Shilpi (2023). *Matching frictions and distorted beliefs: Evidence from a job fair experiment*. Working Paper 958. Queen Mary University of London, School of Economics and Finance.

Abel, Martin, Rulof Burger, and Patrizio Piraino (2020). "The Value of Reference Letters: Experimental Evidence from South Africa". *American Economic Journal: Applied Economics* 12.3, pp. 40–71.

Acevedo, Paloma, Guillermo Cruces, Paul Gertler, and Sebastian Martinez (2020). "How vocational education made women better off but left men behind". *Labour Economics* 65, p. 101824.

Adamecz-Völgyi, Anna and Nikki Shure (2022). "The gender gap in top jobs – The role of overconfidence". *Labour Economics* 79, p. 102283.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham (2023). "Returns to On-the-Job Soft Skills Training". *Journal of Political Economy* 131.8, pp. 2165–2208.

Agüero, Jorge M. and Mindy S. Marks (2008). "Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks". *The American Economic Review* 98.2, pp. 500–504.

Agüero, Jorge M. and Mindy S. Marks (2011). "Motherhood and Women's Earnings in Developing Countries: Evidence from Infertility Shocks". *Journal of Human Resources* 46.4, pp. 800–826.

Agurto, Marcos, Muchin Bazan, Siddharth Hari, and Sudipta Sarangi (2023). *To Inspire and to Inform: The Role of Role Models*. Tech. rep. SSRN 4595079.

Ajayi, Kehinde, Smita Das, Clara Delavallade, Tigist Ketema, and Lea Rouanet (2022). *Gender Differences in Socio-Emotional Skills and Economic Outcomes: New Evidence from 17 African Countries*. Policy Research Working Paper 10197. World Bank.

Alan, Sule, Teodora Boneva, and Seda Ertac (2019). "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit". *The Quarterly Journal of Economics* 134.3, pp. 1121–1162.

Alan, Sule and Seda Ertac (2018). "Fostering patience in the classroom: Results from randomized educational intervention". *Journal of Political Economy* 126.5, pp. 1865–1911.

Alderman, Harold and Derek D. Headey (2017). "How Important is Parental Education for Child Nutrition?" *World Development* 94 (2017), pp. 448–464.

Almond, Douglas, Janet Currie, and Valentina Duque (2018). "Childhood Circumstances and Adult Outcomes: Act II". *Journal of Economic Literature* 56 (4), pp. 1360–1446.

Anderson, Cameron, Sebastien Brion, Don A. Moore, and Jessica A. Kennedy (2012). "A status-enhancement account of overconfidence". *Journal of Personality and Social Psychology* 103 (4), pp. 718–735.

Angelucci, Manuela and Daniel Bennett (2024). "The economic impact of depression treatment in India: Evidence from community-based provision of pharmacotherapy". *American economic review* 114.1, pp. 169–198.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Appadurai, Arjun (2004). "The Capacity to Aspire: Culture and the Terms of Recognition". *Culture and Public Action*. Ed. by Vijayendra Rao and Michael Walton. Palo Alto, California: Stanford University Press. Chap. 3, pp. 59–84. ISBN: 9780804747868.

Arendt, Jacob Nielsen, Mads Lybech Christensen, and Anders Hjorth-Trolle (2021). "Maternal education and child health: Causal evidence from Denmark". *Journal of Health Economics* 80, p. 102552.

Ashraf, Nava, Natalie Bau, Corinne Low, and Kathleen McGinn (2020). "Negotiating a better future: How interpersonal skills facilitate intergenerational investment". *The Quarterly Journal of Economics* 135.2, pp. 1095–1151.

Bandura, Albert (1997). *Self-Efficacy: The Exercise of Control*. New York: W. H. Freeman.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell (2016). "Heaping-Induced Bias In Regression-Discontinuity Designs". *Economic Inquiry* 54 (1), pp. 268–293.

Barrera-Osorio, Felipe, Andreas de Barros, and Deon Filmer (2024). "Long-term impacts of primary school scholarships: Evidence from Cambodia". *Journal of Policy Analysis and Management* 43 (1), pp. 10–38.

Barrera-Osorio, Felipe, Adriana Kugler, and Mikko Silliman (2023). "Hard and Soft Skills in Vocational Training: Experimental Evidence from Colombia". *The World Bank Economic Review* 37.3, pp. 409–436.

Barron, Kai and Christina Gravert (2022). "Confidence and career choices: an experiment". *The Scandinavian Journal of Economics* 124 (1), pp. 35–68.

Bassi, Vittorio and Aisha Nansamba (2021). "Screening and Signalling Non-Cognitive Skills: Experimental Evidence from Uganda". *The Economic Journal* 132.642, pp. 471–511.

Batista, Catia and Julia Seither (2019). *Aspirations, expectations, identities: behavioral constraints of micro-entrepreneurs*. Tech. rep. Universidade Nova de Lisboa, Nova School of Business and Economics, NOVAFRICA.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova (2012). "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India". *science* 335.6068, pp. 582–586.

Beg, Sabrin, Waqas Halim, Adrienne M Lucas, and Umar Saif (2022). "Engaging teachers with technology increased achievement, bypassing teachers did not". *American Economic Journal: Economic Policy* 14.2, pp. 61–90.

Bénabou, Roland and Jean Tirole (2002). "Self-Confidence and Personal Motivation". *The Quarterly Journal of Economics* 117.3, pp. 871–915.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1, pp. 289–300.

Berger, Eva M., Henning Hermes, Guenther Koenig, Felix Schmidt, and Daniel Schunk (2022). "Self-regulation training and job search input: A natural field experiment within an active labor market program". *Journal of Behavioral and Experimental Economics* 98, p. 101858.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, Alemayehu Seyoum Taffesse, and Giulio Schinaia (2023). *The Future in Mind: Aspirations and Long-term Outcomes in Rural Ethiopia*. CEPR Discussion Paper No. 18492. CEPR Press.

Bhalotra, Sonia and Samantha B. Rawlings (2011). “Intergenerational persistence in health in developing countries: The penalty of gender inequality?” *Journal of Public Economics* 95 (3-4), pp. 286–299.

Bjorvatn, Kjetil, Alexander W Cappelen, Linda Helgesson Sekei, Erik Ø Sørensen, and Bertil Tungodden (2019). “Teaching through television: Experimental evidence on entrepreneurship education in Tanzania”. *Management Science* 66.6, pp. 2308–2325.

Black, Robert E., Cesar G. Victora, Susan P. Walker, Zulfiqar A. Bhutta, Parul Christian, Mercedes De Onis, Majid Ezzati, Sally Grantham-Mcgregor, Joanne Katz, Reynaldo Martorell, and Ricardo Uauy (2013). “Maternal and child undernutrition and overweight in low-income and middle-income countries”. *The Lancet* 382 (9890), pp. 427–451.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2008). “Staying in the Classroom and Out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births”. *Economic Journal* 118.530, pp. 1025–1054.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Christopher Rockmore, Brian Stacy, Jakob Svensson, and Waly Wane (2017). “What Do Teachers Know and Do? Does It Matter? Evidence from Primary Schools in Africa”. *Journal of Human Resources* 52.3, pp. 559–594.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2019). “Beliefs about gender”. *American Economic Review* 109.3, pp. 739–773.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel (2008). “The Economics and Psychology of Personality Traits”. *Journal of Human Resources* 43.4, pp. 972–1059.

Brar, Rajdev Kaur, Niklas Buehren, Sreelakshmi Papineni, and Munshi Sulaiman (2023). *Rebel with a Cause: Effects of a Gender Norms Intervention for Adolescents in Somalia*. Policy Research working paper WPS 10567. The World Bank.

Breierova, Lucia and Esther Duflo (2004). *The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?* NBER Working Paper. NBER.

Brudevold-Newman, Andrew (2021). “Expanding access to secondary education: Evidence from a fee reduction and capacity expansion policy in Kenya”. *Economics of Education Review* 83 (2021), p. 102127.

— (2025). “Impacts of Large School Grants and Admissions Algorithms on Student Achievement and Admissions Diversity: Evidence from High-Performing Kenyan Secondary Schools”. *Journal of African Economies* 34 (3), pp. 297–323.

Brudevold-Newman, Andrew and Diego Javier Ubfal (2024). "Soft-skills, networking, and workforce entry: Impacts of a training program for recent graduates in Rwanda". *Labour Economics* 91.2024, p. 102650.

Bruhn, Miriam and David McKenzie (2009). "In pursuit of balance: Randomization in practice in development field experiments". *American economic journal: applied economics* 1.4, pp. 200–232.

Burde, Dana and Leigh L Linden (2013). "Bringing education to Afghan girls: A randomized controlled trial of village-based schools". *American Economic Journal: Applied Economics* 5.3, pp. 27–40.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini (2013). "Overconfidence and Social Signalling". *The Review of Economic Studies* 80.3, pp. 949–983.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik (2015). "Optimal Data-Driven Regression Discontinuity Plots". *Journal of the American Statistical Association* 110 (512), pp. 1753–1769.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs". *Econometrica* 82 (6), pp. 2295–2326.

Cameron, A Colin and Douglas L Miller (2015). "A practitioner's guide to cluster-robust inference". *Journal of Human Resources* 50.2, pp. 317–372.

Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary C. Johnson, David McKenzie, and Mona Mensmann (2017). "Teaching personal initiative beats traditional training in boosting small business in West Africa". *Science* 357.6357, pp. 1287–1290.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber (2015). "Inference on Causal Effects in a Generalized Regression Kink Design". *Econometrica* 83 (6), pp. 2453–2483.

Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin (2022). "Job Search and Hiring with Limited Information about Workseekers' Skills". *American Economic Review* 112.11, pp. 3547–83.

Cassidy, Rachel, Smita Das, Clara Delavallade, Elijah Kipchumba, and Julietha Komba (2024). *Do Men Really Have Greater Socio-emotional Skills than Women? Evidence from Tanzanian Youth*. Policy Research Working Paper 10873. World Bank.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma (2020). "Simple Local Polynomial Density Estimators". *Journal of the American Statistical Association* 115 (531), pp. 1449–1455.

Chiappori, Pierre-André, Murat Iyigun, and Yoram Weiss (2009). "Investment in Schooling and the Marriage Market". *American Economic Review* 99.5, pp. 1689–1713.

Chioda, Laura, David Contreras-Loya, Paul Gertler, and Dana Carney (2021). *Making Entrepreneurs: Returns to Training Youth in Hard Versus Soft Business Skills*. NBER Working Paper 28845. NBER.

Crawfurd, Lee and Aisha Ali (2022). "The Case for Free Secondary Education". *Schooling for All: Feasible Strategies to Achieve Universal Education*. Ed. by Justin Sandefur. Washington, DC: Center for Global Development. Chap. 3, pp. 55–73.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". *Econometrica* 78 (3), pp. 883–931.

Cunha, Flavio and James J. Heckman (2007). "The Technology of Skill Formation". *American Economic Review* 97.2, pp. 31–47.

Cunningham, Wendy V. and Paula Villaseñor (2016). "Employer Voices, Employer Demands, and Implications for Public Skills Development Policy Connecting the Labor and Education Sectors". *World Bank Research Observer* 31.1, pp. 102–134.

Currie, Janet and Tom Vogl (2013). "Early-Life Health and Adult Circumstance in Developing Countries". *Annual Review of Economics* 5, pp. 1–36.

Dahl, Gordon B, Andreas Kotsadam, and Dan-Olof Rooth (2021). "Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams". *The Quarterly journal of economics* 136.2, pp. 987–1030.

Dalton, Patricio S, Sayantan Ghosal, and Anandi Mani (2015). "Poverty and aspirations failure". *The Economic Journal* 126.590, pp. 165–188.

Dammert, Ana C. and Aisha Nansamba (2023). "Skills training and business outcomes: Experimental evidence from Liberia". *World Development* 162, p. 106120.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall (2014). "Household Choices and Child Development". *Review of Economic Studies* 81.1, pp. 137–185.

Delavallade, Clara, Smita Das, Lea Rouanet, Aidan Clerkin, Chris Gonzales, and Julian Jamison (2025). "Socio-emotional Skills in Sub-Saharan Africa: Validating and Comparing Self-reported and Behavioral Measures". *Mimeo*.

Delavallade, Clara, Lea Rouanet, and Smita Das (2020). *Unpacking Socio-Emotional Skills for Women's Economic Empowerment*. Tech. rep. World Bank.

Deming, David J. (2017). "The Growing Importance of Social Skills in the Labor Market". *The Quarterly Journal of Economics* 132.4, pp. 1593–1640.

Demiral, Elif E. and Johanna Mollerstrom (2024). "Signaling confidence". *Journal of Economic Behavior & Organization* 226, p. 106691.

Dhar, Diva, Tarun Jain, and Seema Jayachandran (2019). "Intergenerational transmission of gender attitudes: Evidence from India". *The Journal of Development Studies* 55.12, pp. 2572–2592.

— (2022). "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India". *American economic review* 112.3, pp. 899–927.

Di, Aurelia (2024). *Inspirational or Demotivational: Experimental Evidence on the Impacts of Role Models*. Tech. rep. SSRN 4749305.

Doss, Cheryl (2013). "Intrahousehold Bargaining and Resource Allocation in Developing Countries". *The World Bank Research Observer* 28.1, pp. 52–78.

Duckworth, A.L. and P.D. Quinn (2009). "Development and validation of the Short Grit Scale (GRIT–S)". *Journal of personality assessment* 91.2, pp. 166–174.

Duflo, Esther (2001). "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment". *American economic review* 91.4, pp. 795–813.

— (2003). "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa". *The World Bank Economic Review* 17.1, pp. 1–25.

— (2004). "The medium run effects of educational expansion: Evidence from a large school construction program in Indonesia". *Journal of Development Economics* 74 (1), pp. 163–197.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2024). *The Impact of Secondary School Subsidies on Career Trajectories in a Dual Labor Market: Experimental Evidence from Ghana*. CEPR Discussion Paper. CEPR.

Duflo, Esther, Pascaline Dupas, Elizabeth Spelke, and Mark P. Walsh (2024). *Intergenerational Impacts of Secondary Education: Experimental Evidence from Ghana*. NBER Working Paper. National Bureau of Economic Research.

Dweck, Carol S. (2006). *Mindset: The New Psychology of Success*. New York: Random House.

Eberwein, Julia Dayton, Jakub Kakietek, Davide de Beni, Grainne Moloney, Audrey Pereira, Jonathan Kweku Akuoku, Marjorie Volege, Sicily Matu, and Meera Shekar (2016). *An Investment Framework for Nutrition in Kenya: Reducing Stunting and Other Forms of Child Malnutrition*. Health, Nutrition and Population Discussion Paper. World Bank.

Exley, Christine L and Judd B Kessler (2022). "The gender gap in self-promotion". *The Quarterly Journal of Economics* 137.3, pp. 1345–1381.

Exley, Christine L. and Kirby Nielsen (2024). "The Gender Gap in Confidence: Expected but Not Accounted For". *American Economic Review* 114 (3), pp. 851–85.

Feld, Jan, Jan Sauermann, and Andries de Grip (2017). "Estimating the relationship between skill and overconfidence". *Journal of Behavioral and Experimental Economics* 68, pp. 18–24.

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti (2004). "Mothers and sons: Preference formation and female labor force dynamics". *The Quarterly Journal of Economics* 119.4, pp. 1249–1299.

Frese, M., D. Fay, T. Hilburger, K. Leng, and A. Tag (1997). "The concept of personal initiative: Operationalization, reliability and validity in two German samples". *Journal of occupational and organizational psychology* 70.2, pp. 139–161.

Fujimoto, Junichi, David Lagakos, and Mitchell Vanvuren (2023). *Macroeconomic Effects of 'Free' Secondary Schooling in the Developing World*. NBER Working Paper. NBER.

Galasso, Emanuela and Adam Wagstaff (2019). "The aggregate income losses from childhood stunting and the returns to a nutrition intervention aimed at reducing stunting". *Economics and human biology* 34, pp. 225–238.

Genicot, Garance and Debraj Ray (2017). "Aspirations and inequality". *Econometrica* 85.2, pp. 489–519.

— (2020). "Aspirations and Economic Behavior". *Annual Review of Economics* 12.2020, pp. 715–746.

Glennerster, Rachel, Michael Kremer, Isaac Mbiti, and Kudzai Takavarasha (2011). *Access and Quality in the Kenyan Education System: A Review of the Progress, Challenges and Potential Solutions Office of the Prime Minister of Kenya*. Tech. rep. Office of the Prime Minister of Kenya.

Glewwe, Paul (1999). "Why does mother's schooling raise child health in developing countries? Evidence from Morocco". *Journal of Human Resources* 34 (1), pp. 124–159.

Golan, Jennifer and Jing You (2021). "Raising Aspirations of Boys and Girls through Role Models: Evidence from a Field Experiment". *The Journal of Development Studies* 57.6, pp. 949–979.

Grépin, Karen A. and Prashant Bharadwaj (2015). "Maternal education and child mortality in Zimbabwe". *Journal of Health Economics* 44, pp. 97–117.

Groh, Matthew, Nandini Krishnan, David McKenzie, and Tara Vishwanath (2016). "The impact of soft skills training on female youth employment: evidence from a randomized experiment in Jordan". *IZA Journal of Labor & Development* 5.1, p. 9.

Grossman, Michael (1972). "On the Concept of Health Capital and the Demand for Health". *Journal of Political Economy* 80 (2), pp. 223–255.

Gruijters, Rob J., Mohammed A. Abango, and Leslie Casely-Hayford (2024). "Secondary School Fee Abolition in Sub-Saharan Africa: Taking Stock of the Evidence". *Comparative Education Review*, pp. 000–000.

Gulesci, Selim, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman, and Peyton Young (2025). "A stepping stone approach to norm transitions". *American Economic Review* 115.7, pp. 2237–2266.

Hanushek, Eric A. and Ludger Woessmann (2008). "The Role of Cognitive Skills in Economic Development". *Journal of Economic Literature* 46.3, pp. 607–668.

Harris, J. Andrew and Daniel N. Posner (2019). "(Under What Conditions) Do Politicians Reward Their Supporters? Evidence from Kenya's Constituencies Development Fund". *American Political Science Review* 113 (1), pp. 123–139.

Hasan, Amer, Nozomi Nakajima, and Marcos A Rangel (2020). *Mama Knows (and Does) Best Maternal Schooling Opportunities and Child Development in Indonesia*. Policy Research Working Paper. World Bank.

Heath, Rachel and Seema Jayachandran (2017). "The causes and consequences of increased female education and labor force participation in developing countries". *The Oxford Handbook of Women and the Economy*. Oxford University Press, pp. 345–367.

Heckman, James J. and Tim Kautz (2012). "Hard evidence on soft skills". *Labour Economics* 19.4, pp. 451–464.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior". *Journal of Labor Economics* 24.3, pp. 411–482.

Heß, Simon (2017). "Randomization inference with Stata: A guide and software". *The Stata Journal* 17.3, pp. 630–651.

Hoffman, Mitchell and Stephen V. Burks (2020). "Worker overconfidence: Field evidence and implications for employee turnover and firm profits". *Quantitative Economics* 11.1, pp. 315–348.

Hossain, Mobarak and Matthew C. H. Jukes (2024). "Gender Differences in Socioemotional Skills among Adolescents and Young Adults in Ethiopia, India, Peru and Vietnam". *The Journal of Development Studies* 60.10, pp. 1–19.

ILO (2018). *Child labour and the youth decent work deficit in Tanzania*. Tech. rep. Geneva: International Labour Office (ILO).

Imai, Kosuke, Gary King, Clayton Nall, et al. (2009). "The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation". *Statistical Science* 24.1, pp. 29–53.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Jackson, C. Kirabo (2010). "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago". *Economic Journal* 120 (549), pp. 1399–1429.

Jales, Hugo, Jun Ma, and Zhengfei Yu (2017). "Optimal bandwidth selection for local linear estimation of discontinuity in density". *Economics Letters* 153, pp. 23–27.

Jayachandran, Seema (2021). "Social Norms as a Barrier to Women's Employment in Developing Countries". *IMF Economic Review volume* 69, pp. 576–595.

Jensen, Robert (2010). "The (perceived) returns to education and the demand for schooling". *The Quarterly Journal of Economics* 125.2, pp. 515–548.

Jensen, Robert and Emily Oster (2009). "The power of TV: Cable television and women's status in India". *The Quarterly Journal of Economics* 124.3, pp. 1057–1094.

Juma, Amina A. (2025). "Bridging skills gaps: evaluating VET reskilling and curriculum reforms for enhancing graduate employability in Tanzania". *International Journal of Training Research* 0.0, pp. 1–14.

Karlsson, Omar, Jan Walter De Neve, and S. V. Subramanian (2019). "Weakening association of parental education: analysis of child health outcomes in 43 low- and middle-income countries". *International Journal of Epidemiology* 48 (1), pp. 83–97.

Karszen, Lucienne N, Seth Richard Gitter, and Melissa A Groves (2025). "Nudges to encourage female and racial minority students to enroll in additional economics courses". *Economic Inquiry* 63.4, pp. 1386–1408.

Kazianga, Harounan, Dan Levy, Leigh L Linden, and Matt Sloan (2013). "The effects of "girl-friendly" schools: Evidence from the BRIGHT school construction program in Burkina Faso". *American Economic Journal: Applied Economics* 5.3, pp. 41–62.

Kazibwe, Douglas and Jinhu Li (2025). "Universal secondary education, schooling and women 's empowerment: Evidence from Uganda". *Journal of Development Economics* 174, p. 103464.

Kiss, Andrea, Robert Garlick, Kate Orkin, and Lukas Hensel (2023). *Jobseekers' Beliefs About Comparative Advantage and (Mis)Directed Search*. IZA Discussion Paper Series 16522. IZA.

KNBS (2012). *2009 Kenya Population and Housing Census Analytical Report on Education*. Tech. rep. Kenya National Bureau of Statistics (KNBS).

— (2023a). *Economic Survey 2023*. Tech. rep. Kenya National Bureau of Statistics (KNBS).

— (2023b). *Kenya Demographic and Health Survey 2022*. Tech. rep. Kenya National Bureau of Statistics (KNBS), Ministry of Health, and ICF.

Kreemers, Loes M., Edwin A. J. van Hooft, and Annelies E. M. van Vianen (2022). "If you want a job, don't just search hard, search systematically: A field study with career starters". *European Journal of Work and Organizational Psychology* 31.3, pp. 317–330.

Kremer, Michael, Conner Brannen, and Rachel Glennerster (2013). "The challenge of education and learning in the developing world". *Science* 340.6130, pp. 297–300.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea (2012). "Soap operas and fertility: Evidence from Brazil". *American Economic Journal: Applied Economics* 4.4, pp. 1–31.

Lavy, Victor, Olmo Silva, and Felix Weinhardt (2012). "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools". *Journal of Labor Economics* 30 (2), pp. 367–414.

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin (2021). "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools". *American economic review* 111.7, pp. 2213–2246.

Leight, Jessica, Daniel O Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet (2024). "Aspiring to more? New evidence on the effect of light-touch aspirations interventions in rural Ethiopia". *Journal of Development Effectiveness* 16.4, pp. 485–498.

LeVine, Robert A., Sarah LeVine, Beatrice Schnell-Anzola, Meredith L. Rowe, and Emily Dexter (2012). *Literacy and Mothering: How Women's Schooling Changes the Lives of the World's Children*. Oxford University Press. ISBN: 9780195309829.

Lim, Jaegeum and Jonathan Meer (2019). "Persistent effects of teacher-student gender matches". *Journal of Human Resources* 55.3, pp. 809–835.

Lindeboom, Maarten, Ana Llena-Nozal, and Bas van der Klaauw (2009). "Parental education and child health: Evidence from a schooling reform". *Journal of Health Economics* 28 (1), pp. 109–131.

Lubega, Patrick, Frances Nakakawa, Gaia Narciso, Carol Newman, Archileo N Kaaya, Cissy Kityo, and Gaston A Tumuhimbise (2021). "Body and mind: Experimental evidence from women living with HIV". *Journal of Development Economics* 150, p. 102613.

Lucas, Adrienne M. and Isaac M. Mbiti (2012). "Access, sorting, and achievement: The short-run effects of free primary education in Kenya". *American Economic Journal: Applied Economics* 4 (4), pp. 226–253.

— (2014). "Effects of school quality on student achievement: Discontinuity evidence from Kenya". *American Economic Journal: Applied Economics* 6 (3), pp. 234–263.

Lundberg, Shelly and Robert A. Pollak (1993). "Separate Spheres Bargaining and the Marriage Market". *Journal of Political Economy* 101.6, pp. 988–1010.

Lybbert, Travis J. and Bruce Wydick (2018). "Poverty, Aspirations, and the Economics of Hope". *Economic Development and Cultural Change* 66.4, pp. 709–753.

Macharia, Peter M., Noel K. Joseph, Benn Sartorius, Robert W. Snow, and Emelda A. Okiro (2021). "Subnational estimates of factors associated with under-five mortality in Kenya: a spatio-temporal analysis, 1993–2014". *BMJ Global Health* 6 (4), p. 4544.

Maina, Joseph, Paul O. Ouma, Peter M. Macharia, Victor A. Alegana, Benard Mitto, Ibrahima Socé Fall, Abdisalan M. Noor, Robert W. Snow, and Emelda A. Okiro (2019). "A spatial database of health facilities managed by the public health sector in sub Saharan Africa". *Scientific Data 2019 6:1* 6 (1), pp. 1–8.

Marsh, Vic, Clara Delavallade, Smita Das, Lea Rouanet, Tricia Koroknay-Palicz, and Dawn McDaniel (2025). "Assessing Workplace Social, Emotional, and Behavioral Skills in Africa: The ESTEEM Framework". *Mimeo*.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani (2019). "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania". *The Quarterly journal of economics* 134.3, pp. 1627–1673.

McCrary, Justin and Heather Royer (2011). "The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth". *American Economic Review* 101 (1), pp. 158–195.

McKelway, Madeline (2025). *Women's Self-Efficacy and Economic Outcomes: Experimental Evidence from India*. Tech. rep. SSRN 5158816.

Ministry of Education (2022). *Annual Statistics Yearbook 2021-2022*. Tech. rep. Ministry of Education, Culture and Higher Education.

Ministry of Education, Culture and Higher Education (2017). *Education Sector Strategic Plan 2018-2020*. Tech. rep. Ministry of Education, Culture and Higher Education.

Ministry of Health (2015). *Kenya Health Workforce Report: The Status of Healthcare Professionals in Kenya*. Tech. rep. Ministry of Health.

Morgenroth, Thekla, Michelle K Ryan, and Kim Peters (2015). "The motivational theory of role modeling: How role models influence role aspirants' goals". *Review of General Psychology* 19.4, pp. 465–483.

Morsy, Hanan and Adamon N. Mukasa (2020). "'Mind the mismatch?' Incidence, drivers, and persistence of African youths' skill and educational mismatches". *African Development Review* 32.S1, S5–S19.

Muchiri, Steve (2021). "Impact of Free/Subsidized Secondary School Education on the Likelihood of Teenage Motherhood". *Demography* 58 (4), pp. 1401–1421.

Munishi, Emmanuel J (2016). "Factors contributing to lack of employable skills among technical and vocational education (TVET) graduates in Tanzania". *Business Education Journal* I.2.

Muralidharan, Karthik and Ketki Sheth (2016). "Bridging education gender gaps in developing countries: The role of female teachers". *Journal of Human Resources* 51.2, pp. 269–297.

NACTE (2020). *Mapping Skills Gap and Skills Needs for Technician Graduates in the Selected Economic Sectors for Industrial Growth in Tanzania*. Tech. rep. Dar es Salam: National Council for Technical Education (NACTE).

Neve, Jan Walter De and S. V. Subramanian (2018). "Causal Effect of Parental Schooling on Early Childhood Undernutrition: Quasi-Experimental Evidence From Zimbabwe". *American Journal of Epidemiology* 187 (1), pp. 82–93.

Nguyen, Trang (2008). *Information, role models and perceived returns to education: Experimental evidence from Madagascar*. enGender Impact. World Bank.

Nicolai, Susan, Annalisa Prizzon, and Sébastien Hine (2014). *Beyond Basic: The growth of post-primary education in Kenya*. Tech. rep. Overseas Development Institute.

Nyqvist, By Martina Björkman, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott (2019). "Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda". *American Economic Journal: Applied Economics* 11 (3), pp. 155–92.

Ohba, Asayo (2009). *Does free secondary education enable the poor to gain access? A study from rural Kenya*. Tech. rep. Consortium for Educational Access, Transitions and Equity (CREATE).

Oketch, Moses and Anthony Somerset (2010). *Free Primary Education and After in Kenya: Enrolment impact, quality effects, and the transition to secondary school*. Tech. rep. Consortium for Educational Access, Transitions and Equity (CREATE), University of Sussex.

Okoroafor, Sunny C., Brendan Kwesiga, Julius Ogato, Zeinab Gura, Joel Gondi, Nakato Jumba, Teresa Ogumbo, Maureen Monyoncho, Annah Wamae, Mutile

Wanyee, Meldah Angir, Mona Ahmed Almudhwahi, Chagina Evalyne, Juliet Nabyonga-Orem, Adam Ahmat, Pascal Zurn, and James Avoka Asamani (2022). "Investing in the health workforce in Kenya: trends in size, composition and distribution from a descriptive health labour market analysis". *BMJ Global Health* 7 (Suppl 1), e009748.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou (2020). "Mothers, peers, and gender-role identity". *Journal of the European Economic Association* 18.1, pp. 266–301.

Osili, Una O. and Bridget T. Long (2008). "Does Female Schooling Reduce Fertility? Evidence from Nigeria". *Journal of Development Economics* 87.1, pp. 57–75.

Oyugi, Boniface, Olena Nizalova, Sally Kendall, and Stephen Peckham (2024). "Does a free maternity policy in Kenya work? Impact and cost–benefit consideration based on demographic health survey data". *European Journal of Health Economics* 25 (1), pp. 77–89.

Ozen, Efsan Nas, Stefan Hut, Victoria Levin, and Ana Maria Munoz Boudet (2020). *A Field Experiment on the Role of Socioemotional Skills and Gender for Hiring in Turkey*. Policy Research Working Paper 9154. World Bank.

Ozier, Owen (2018). "The Impact of Secondary Schooling in Kenya". *Journal of Human Resources* 53 (1), pp. 157–188.

Patnaik, Arpita, Gwyn Pauley, Joanna Venator, and Matthew Wiswall (2024). "The impacts of same and opposite gender alumni speakers on interest in economics". *Economics of Education Review* 102.October 2024, p. 102579.

Pop-Eleches, Cristian and Miguel Urquiola (2013). "Going to a Better School: Effects and Behavioral Responses". *American Economic Review* 103 (4), pp. 1289–324.

Porteous, Obie (2020). "Research Deserts and Oases: Evidence from 27 Thousand Economics Journal Articles on Africa". *Oxford Bulletin of Economics and Statistics* 84.6, pp. 1235–1258.

Porter, Catherine and Danila Serra (2020). "Gender differences in the choice of major: The importance of female role models". *American Economic Journal: Applied Economics* 12.3, pp. 226–54.

Prada, María F, Graciana Rucci, and Sergio Urzúa (2019). *Training, soft skills and productivity: evidence from a Field Experiment in Retail*. Tech. rep. IDB Working Paper Series.

Pritchett, Lant (2001). "Where Has All the Education Gone?" *The World Bank Economic Review* 15 (3), pp. 367–391.

— (2013). *The Rebirth of Education: Schooling Ain't Learning*. Washington, D.C.: Center for Global Development.

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth (2018). "Measuring and Bounding Experimenter Demand". *American Economic Review* 108.11, pp. 3266–3302.

Ray, Debraj (2006). "Aspirations, Poverty, and Economic Change". *Understanding Poverty*. Ed. by Abhijit V. Banerjee, Roland Benabou, and Dilip Mookherjee. Oxford: Oxford University Press, pp. 409–421.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar (2017). "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender". *The Economic Journal* 127.604, pp. 2153–2186.

Riley, Emma (2024). "Role models in movies: The impact of Queen of Katwe on students' educational attainment". *Review of Economics and Statistics* 106.2, pp. 334–351.

Ringdal, Charlotte and Ingrid Hoem Sjursen (2021). "Household Bargaining and Spending on Children: Experimental Evidence from Tanzania". *Economica* 88 (350), pp. 430–455.

Romano, Joseph P and Michael Wolf (2005). "Stepwise multiple testing as formalized data snooping". *Econometrica* 73.4, pp. 1237–1282.

Save the Children (2018). *Review of Save the Children Norway's Education Programmes and Partnerships with National Government Institutions at Country Level*. Tech. rep. Save the Children Norway.

Schutte, N. S., J. M. Malouff, L. E. Hall, D. J. Haggerty, J. T. Cooper, C. J. Golden, and L. Dornheim (1998). "Development and validation of a measure of emotional intelligence". *Personality and individual differences* 25.2, pp. 167–177.

Schwardmann, Peter and Joël van der Weele (2019). "Deception and self-deception". *Nature Human Behaviour* 3 (10), pp. 1055–1061.

Schwarzer, R. and M. Jerusalem (1995). "Generalized self-efficacy scale". *Measures in health psychology: A user's portfolio. Causal and control beliefs*, pp. 35–37.

Selman, Robert L, William Beardslee, Lynn Hickey Schultz, Michael Krupa, and Donna Podorefsky (1986). "Assessing adolescent interpersonal negotiation strategies: Toward the integration of structural and functional models". *Developmental Psychology* 22.4, p. 450.

Serra, Danila (2025). "Role models in developing countries". *Handbook of experimental development economics*. Edward Elgar Publishing, pp. 123–135.

Soldà, Alice, Changxia Ke, William von Hippel, and Lionel Page (2021). "Absolute Versus Relative Success: Why Overconfidence Creates an Inefficient Equilibrium". *Psychological Science* 32 (10), pp. 1662–1674.

Spaull, Nicholas (2013). "Poverty & Privilege: Primary School Inequality in South Africa". *International Journal of Educational Development* 33.5, pp. 436–447.

Spence, Michael (1973). "Job Market Signaling". *The Quarterly Journal of Economics* 87.3, pp. 355–374.

Tamir, Maya and Brett Q Ford (2009). "Choosing to be afraid: preferences for fear as a function of goal pursuit". *Emotion* 9.4, pp. 488–497.

Tan, Hong, Sajitha Bashir, and Nobuyuki Tanaka (2016). *Skill Use, Skill Deficits, and Firm Performance in Formal Sector Enterprises: Evidence from the Tanzania Enterprise Skills Survey, 2015*. Policy Research Working Paper. The World Bank.

Tanzania National Bureau of Statistics (2021). *Integrated Labour Force Survey 2020/21*. Key Labour Market Indicators. Dar es Salam: Tanzania National Bureau of Statistics.

Thurstans, Susan, Charles Opondo, Andrew Seal, Jonathan Wells, Tanya Khara, Carmel Dolan, André Briend, Mark Myatt, Michel Garenne, Rebecca Sear, and Marko Kerac (2020). "Boys are more likely to be undernourished than girls: a systematic review and meta-analysis of sex differences in undernutrition". *BMJ Global Health* 5 (12), e004030.

Ubfal, Diego, Irani Arráiz, Diether W. Beuermann, Michael Frese, Alessandro Maffioli, and Daniel Verch (2022). "The impact of soft-skills training for entrepreneurs in Jamaica". *World Development* 152, p. 105787.

UNDP (2019). *Human Development Report 2019*. Tech. rep. United Nations Development Program.

— (2023). *Human Development Report 2023/2024*. Tech. rep. United Nations Development Programme Somalia.

UNESCO (2022). *Somalia Education Sector Analysis, Assessing opportunities for rebuilding the country through education*. Tech. rep. Federal Government of Somalia, IIEP-UNESCO Dakar.

UNICEF (2017). *Somalia Education Cluster Annual Report 2016*. Tech. rep. UNICEF.

Usawa (2024). *Is our secondary school system unjust by design? Usawa 2nd Secondary School Survey Report*. Tech. rep. Usawa Agenda.

World Bank (2018). *World Development Report 2018: Learning to Realize Education's Promise*. Tech. rep. World Bank.

— (2019). *Somali Poverty and Vulnerability Assessment*. Tech. rep. Report No. AUS0000407. World Bank.

❧

This thesis was typeset using the LaTeX typesetting system created by Leslie Lamport and Donald Knuth, and the memoir class. The body text is set 11pt with Palatino designed by Hermann Zapf, which includes italics and small caps. Other fonts include `Monospace` from Young Ryu's TX Fonts family, Sans from TeX User Group Poland's Latin Modern family, and Helvetica by Max Miedinger and Eduard Hoffmann.