Essays on Inequality in Education, Labor Market, and Health

Zhan Shi

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Economics
University of Strathclyde

August 2024

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date: 26/08/2024

Statement of Contribution of Co-Authors

Chapters 2 to 4 consist of papers submitted for publication and under editorial review or working papers during the period of candidature. An earlier version of Chapter 2 has been released for working paper and is currently under editorial review. The undersigned hereby certify that:

1. they meet the criteria for co-authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the working paper in their field or expertise;

2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the working paper;

3. there are no other authors of the publication according to these criteria; and

4. there are no potential conflicts of interest that require to be disclosed to (a) granting bodies, (b) the editor or publisher of any journal or other publication, and (c) the head of the responsible academic unit.

In the case of Chapter 2, contributions to the work involved the following:

Contributor	**Statement of contribution**	**Publication title and date of publication or status**
M. Fongoni Marco Fongoni Signature	Conceptualization; Theoretical framework; Co-writing;	Income Inequality and Peer Effects in Education. This paper previously appeared under a different title as IZA Discussion Paper No. 15785: "Reference Dependent Aspirations and Peer Effects in Education."
J. Norris JNorris Signature	Conceptualization; Empirical analysis; Supervision; Co-writing;	
A. Romiti Agnese Romiti Signature	Conceptualization; Supervision; Co-writing;	
Z. Shi Zhan Shi Signature	Empirical analysis; Co-writing;	

In the case of Chapter 3, contributions to the work involved the following:

Contributor	**Statement of contribution**	**Publication title and date of publication or status**
C. Boinet	Data cleaning; Data analysis; Co-writing;	Beliefs on Children's Human Capital Formation and Mothers at Work
 Signature		
J. Norris	Conceptualization; Survey design; Survey implementation; Data analysis; Supervision; Co-writing;	
 Signature		
A. Romiti	Conceptualization; Survey design; Supervision; Co-writing;	
 Signature		
Z. Shi	Survey coding; Survey implementation; Data collection; Data cleaning;	
 Signature		

In the case of Chapter 4, contributions to the work involved the following:

Contributor	**Statement of contribution**	**Publication title and date of publication or status**
Z. Shi	Conceptualization; Empirical analysis; Co-writing;	One-Child Policy in China and the Intergenerational Effects on Health
 Signature		
T. Ta	Conceptualization; Empirical analysis; Co-writing;	
 Signature		

Acknowledgements

I would first like to thank my supervisors, Dr Jonathan Norris and Dr Agnese Romiti, for their guidance and support throughout my PhD studies. Their insights, encouragement and constructive feedback have been invaluable in shaping this research and helping me to grow in the process. This thesis would not have been possible without their enthusiasm and patience.

I would also like to thank the Director of Postgraduate Research, Professor Julia Darby, for organizing catch-ups with Economics students and for providing any advice I may require. I would also like to extend my gratitude to Professor Alex Dickson, Associate Dean (PGR), for his support and efforts in providing a better PGR space.

I am also grateful to the Department of Economics at the University of Strathclyde for providing me with the resources and environment to pursue my research. Thank you for all staff members who allocated time to give me feedback and support on my research such as Dr Sharada Davidson, Dr Marco Fongoni, Dr Markus Gehrsitz, Dr Otto Lenhart, Professor Stuart McIntyre, Dr Paul Telemo and Dr Ping Wu.

I would like to thank my PhD colleagues Annie, Arnold, Baland, Bethany, Césarine, Geoff, Grant, Iswat, Jon, Lateef, Praveen, Rory, Sabin, Salman, Sam, Slawek and Yuri for their friendship, and for making my time here enjoyable and memorable.

I gratefully acknowledge the funding received towards my PhD from the University Research Excellence Award Studentship.

Finally, I would like to thank my family and friends Haoqi, Jiasheng, Tingting, Xingjian, Yiwen and Zhejian for their unwavering love and support. Their encouragement and belief in me have been a constant source of strength and motivation.

Abstract

This thesis examines the unintended consequences of social and political factors – namely income inequality, gender beliefs, and family planning policies – on educational outcomes, labor market dynamics, and health across generations. The thesis is structured into three distinct but interconnected essays, offering empirical evidence on how these factors shape individual and group behaviors.

The first essay (Chapter 2) investigates the long-run effects of income inequality within school peer compositions on educational attainment. I and my co-authors show that an increase in the share of low-income peers within school cohorts benefits low-income students by increasing their probability of graduating from the university, while at the same time disadvantaging high-income students. We propose a novel theoretical framework based on reference-dependent preferences and social comparison to explain these patterns, emphasizing the role of frustration or motivation depending on students' relative positions in the income distribution. This chapter also highlights that better social connections within schools can mitigate these unintended consequences of income inequality.

In the second essay (Chapter 3), I joined with another set of co-authors to examine gendered beliefs about the effects of mothers working long hours relative to fathers on children's skill development. Using a novel survey design linked to an experiment conducted among parents in England, this chapter elicits initial beliefs and tests the effect of an information treatment on these beliefs. The results show that parents, especially men and conservative voters, believe that mothers working longer hours negatively affect children's future outcomes. However, providing accurate information about children's outcomes when mothers work full time leads to more positive and accurate beliefs, especially among those who initially hold more positive but uncertain views about maternal work. This chapter highlights the potential of targeted information to update gendered beliefs and reduce labor market inequalities.

In the third essay (Chapter 4), I worked with a fellow PhD student studying the spillover effects of China's one-child policy on the health outcomes of the next generation. Using data from the China Family Panel Studies and employing a regression discontinuity design, we find that children born to parents directly affected by the policy show significant improvements in both physical and mental health. These results are attributed to increased parental investment and improved parental health, contributing to the literature on the quantity-quality trade-off and the intergenerational transmission of health. The results provide valuable insights into how family planning policies can have profound and lasting effects on population health.

Taken together, this thesis provides a comprehensive analysis of how income inequality, gender perceptions, and policy interventions interact with individual behavior to produce complex outcomes in education, labor markets, and health. The research highlights the importance of considering these dynamics in the design and implementation of policies aimed at promoting equity and well-being across generations.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis examines the multifaceted effects of socioeconomic factors and policies on inequalities in education, labor market and health outcomes. In three distinct essays, we employ applied econometric methods to explore the nuances of income inequality, gender norms, and family planning policies. Together, these essays provide comprehensive insights into the relationship between socioeconomic dynamics and individual outcomes, with valuable implications for policy design and social interventions.

Chapter 2 is titled *"Income Inequality and Peer Effects in Education"*. In this chapter, we study the long-run effects of income inequality on educational outcomes within peer compositions. Specifically, this chapter examines how the share of low-income peers, which we construct within school cohorts, affects the academic performance of students across income distribution. The focus of this chapter is twofold. First, it aims to show how changes in peer income composition affect long-term educational outcomes differently across the income distribution. To do so, we use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to examine the effects of the share of low-income peers in students' school cohorts. We compare students with similar family incomes and characteristics but different shares of low-income peers, using a within-school, across-cohort design. This design aims to isolate the effect of peer income composition while controlling for other variables.

Our main results show a distinct heterogeneous pattern. For students in the bottom 20th percentile of the income distribution, a standard deviation increase of 20% in the share of low-income peers increases the probability of college completion by 3.6 percentage points (pp). Conversely, for students in the top 20th percentile of income, the same increase decreases college completion by 4.1 pp. For middle-income students, null effects are observed. We further provide evidence that these patterns are not solely attributable to commonly observed peer effect mechanisms such as nonlinear peer ability effects, teacher and parent responses, or disruptive peer behavior.

Second, we propose a theoretical framework to explain our empirical findings. This model posits that students' effort choices are influenced by income-based social comparisons, leading to either motivation or frustration depending on their relative income positions. Drawing on the literature about reference-dependent preferences and social comparison, the framework suggests that perceived inequality among peers can have dual effects – motivating some while discouraging others.

To support the theoretical framework, the chapter presents empirical evidence on short-term performance and motivation measures, including high school grades, self-esteem, perceived intelligence, mental health, and motivation. The results suggest heterogeneous effects: low-income students benefit from increased motivation and self-esteem, while high-income students show signs of decreased motivation and increased depressive symptoms.

The chapter concludes by discussing the broader implications of these findings. It highlights the potential benefits of integrating students from different income backgrounds, suggesting that this can help disseminate information more effectively. It also notes that when social cohesion within a school is strong, integrating students across income groups can mitigate the negative effects observed at the aggregate level. The findings suggest that better social integration within schools, through increased connectedness and friendships, can help mitigate the negative effects of income inequality. This highlights the importance of considering social dynamics in efforts to address educational inequality.

Chapter 3 is titled *"Beliefs on Children's Human Capital Formation and Mothers at Work"*. This chapter presents the results of a pilot survey experiment conducted with a sample of parents in England. This chapter explores gendered beliefs about the impact of maternal employment on children's skill development compared to paternal involvement. The aim is to understand the persistence of gender gaps and gender norms that limit women's opportunities in the labour market.

The study employs a novel survey design that combines hypothetical scenarios and incentivized beliefs, complemented by an information treatment about children's skill development when mothers work long hours. Participants, recruited through the online platform Prolific, were asked to report their beliefs about children's future outcomes – such as the probability of university graduation and earnings at age 30 – under scenarios where either the mother or the father works full-time, with the other working part-time hours. This approach captures within-person differences in beliefs and how these beliefs vary with hypothetical family income levels.

The initial findings reveal two key insights: first, there is a belief that mothers working longer hours negatively affects children's future outcomes; second, this negative belief diminishes only at very high levels of family income. The study also finds clear heterogeneity in these beliefs, with men and conservative voters holding more negative views. These initial beliefs serve as a benchmark for assessing the effects of the subsequent information treatment.

In the information treatment, participants are randomly presented with actual data from the Millennium Cohort Study (MCS) on GCSE pass rates. Participants are first asked to estimate these pass rates under conditions where mothers work full time given the rates where mothers work less than full time, with monetary incentives provided for accurate estimates. After randomly receiving the actual information, participants update their posterior beliefs about children's behavioral problems using the same incentivized approach. The results suggest that providing factual information about GCSE pass rates when mothers work full-time leads to more positive and accurate beliefs about children's behavioral problems.

We further link the information treatment to the initial beliefs from the hypotheticals we collect. Interestingly, the response to the information about GCSE pass rates varies according to participants' initial beliefs. Those who already hold more positive views about children graduating from university when mothers work longer hours are more likely to update their beliefs accurately.

In addition to belief updating, the chapter explores changes in self-reported gender norms following the information treatment. Although the results of the pilot study are inconclusive due to insufficient statistical power, there is some indication that information can lead to polarization. Participants with more positive initial beliefs about maternal employment became more supportive of gender equality, while those with negative initial beliefs tended to revert to more traditional views.

Overall, this chapter provides valuable insights into the strength and variability of beliefs about the impact of maternal employment on children's human capital formation. It highlights the potential for real information to shift these beliefs, albeit heterogeneously, and underscores the importance of understanding these dynamics for designing policies to support working mothers. The findings lay the groundwork for further research into how gender norms and beliefs shape labor market outcomes and the effectiveness of information interventions in promoting gender equality.

Chapter 4 is titled *"One-Child Policy in China and the Intergenerational Effects on Health"*. This chapter studies the spillover effects of the one-child policy in China on the health outcomes of subsequent generations whose parents were born right before and after the implementation of the policy. Although formally relaxed and abolished in 2016, the one-child policy has had long-lasting and profound effects on at least two generations of people. To date, only a limited number of studies have examined the health effects of the policy and the intergenerational effects of family planning policies on the health of subsequent generations.

Using data from the China Family Panel Studies (CFPS), we examine how the policy that drastically altered family size and dynamics in China has affected both physical and mental health outcomes in the subsequent generation. We focus on three measures of physical health – the probability of being sick, self-rated health, and interviewer-rated health – and one measure of mental health, distress, for children born to urban Han parents. Our identification strategy relies on the fact that participants were unable to sort themselves on either side of the policy implementation date, allowing us to use a regression discontinuity (RD) design that exploits the policy cut-off in 1980 to estimate the causal effect of the policy on child health outcomes. This natural experiment setting provides a robust framework for isolating the local average treatment effect of the policy on children's health.

Our results show that children born to mothers affected by the one-child policy have significant improvements in both physical and mental health outcomes compared to children born to mothers not affected by the policy. These results are robust to a number of sensitivity checks and suggest that the policy has had lasting positive effects on the health of subsequent generations. Similarly, although less statistically significant, improvements are observed for children of fathers affected by the policy.

The chapter emphasizes the focus on maternal data, given the strong evidence that urban Han daughters, who were primarily affected by the one-child policy, benefited from demographic changes such as increased resource allocation. This focus is supported by literature showing better educational and health outcomes for children, especially urban daughters, in smaller families. Our analysis suggests several mechanisms, including increased investment in child health, such as increased access to health insurance, and the intergenerational transmission of health and household characteristics. NotaIn particular, better health outcomes are observed for mothers born after the policy's implementation. In addition, parenting practices and parent-child interactions in these families, where mothers were born after the policy, contribute to lower levels of child distress.

Overall, this chapter provides a comprehensive analysis of the intergenerational spillover health effects of the one-child policy, highlighting significant improvements in both the physical and mental health of children born to parents affected by the policy. The findings highlight the complex interactions between family planning policies, resource allocation and long-term health outcomes. It also highlights the importance of considering intergenerational effects when assessing the outcomes of such policies, and offers insights for policymakers and researchers interested in the long-term consequences of population policies.

Chapter 5 concludes the thesis with a summary of the overall objectives and the key contributions of each chapter.

Chapter 2

Income Inequality and Peer Effects in Education

2.1 Introduction

The effects of exposure to income inequality for students are less clear than the sources of inequality. We study changes in peer income compositions on students' long-run educational attainment and their short-run performance. A wide literature indicates that students' outcomes are influenced by peers (Sacerdote, 2014), but peer income inequality is less well understood. It could work through well-known channels in the literature that income may capture, such as the ability distribution, behavior, teachers, or other characteristics (Billings and Hoekstra, 2023; Booij et al., 2017; Carrell et al., 2018; Duflo et al., 2011; Feld and Zölitz, 2017). Alternatively, income inequality may draw students' attention to disparities in opportunity leading to unintended consequences. This could generate frustration among low-income students who have fewer opportunities, or motivation to get head among high-income students when surrounded by those with similar opportunities.[1]

[1]As an anecdotal example, consider a story told by the "This American Life" radio program about a group of high school students attending school in one of America's poorest congressional districts taken to visit a nearby elite private school (Episode 550: Three Miles available at https://www.thisamericanlife.org/550/three-miles). Their reactions, described by a teacher, tell a powerful story (Greenbaum, 2015). "They felt like everyone was looking at them. And one of the students started screaming and crying. Like, this is unfair. This is – I don't want to be here. I'm leaving."

In this paper, we make two main contributions. First, we empirically show that changes in peer income compositions affect educational attainment heterogeneously across adolescent students from lower to higher income families. Further, we show evidence that this heterogeneous pattern is not likely to be explained by a range of mechanisms discussed in the literature. Second, to help rationalize our results, we propose a novel theoretical framework of students' choice of effort where students make income-based social comparisons. Subsequently, we provide some empirical evidence in support of the key mechanisms highlighted by our model: depending on a student's relative position in the income distribution, social comparison based on income can be motivating or, alternatively, lead to frustration and discouragement. Finally, we close the paper with an extension highlighting a path through social cohesion and integration that may help mitigate the consequences of income inequality.

Empirical analysis. Our first contribution is to empirically test the long-run effects of changing adolescents' peer income distributions on eventual university completion heterogenous to students' family income. To capture changes in income distributions, we use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) and the leave-one-out share of low-income peers in students' school-cohort. We choose this broad reference group level partly based on our motivation that income inequality may be about the general environment students are exposed to. In Section 2.2, we discuss more on this choice, and later, in Section 2.7 we compare it against more refined peer reference groups.

We use a within school, across cohort design and effectively compare students in the same school, who have similar family incomes, face similar school-cohort variances in the income distribution, and have similar characteristics, but who face differences in the share of low-income peers across their cohorts. The key assumptions are that unobserved selection factors into schools are fixed at the school level and that our flexible own-income controls fully capture the link between students' family income and their outcomes. Based on these assumptions we avoid contamination of the peer income effect which is split across students' position in the income distribution. We discuss in detail our identification strategy and assumptions in Section 2.3. Moreover, because part of our motivation here is that income inequality can change environments, we condition on the leave-one-out standard deviation in school-cohort incomes so shifts in the share of low-income peers capture real differences in distributions. These stronger differences may represent more salient changes in the environment.[2]

Our main results reveal a clear heterogeneous pattern. Among students in the bottom 20th percentile of the income distribution, a standard deviation increase of 20% in the share of low-income peers *increases* the propensity to complete university by 3.6 percentage points (pp). For students in the top 20th percentile of income, this same change *decreases* university completion by 4.1pp, while middle income students have estimated *null* effects. Furthermore, we confirm that these results are robust to a wide range of checks. In Section 2.4.4, we then turn to assess whether this pattern is explained by common mechanisms within peer effects, such as non-linear peer ability effects, teacher and parental responses, and disruptive peer behavior. We find no evidence that the effects we observe from income inequality are explained by these mechanisms, suggesting that students' responses to income inequality can be significant but not adequately addressed by the current literature.

Theoretical framing and mechanisms. In our second contribution, we advance a novel theoretical model of student effort choice that offers a lens to rationalize the patterns we observe. We consider students who have different capacities for translating effort into an educational outcome, where capacity is a broader construct than just raw ability, encompassing a combination of factors, such as differences in opportunity, enabling a student to achieve outcomes. Importantly, we consider income as one such factor that is both salient and observable in school. A central component of our theory is then the idea that social comparison among students based on income can generate both frustration and motivation depending on a student's relative position in the income distribution.

In our model, students compare realized outcomes in relation to a reference point for

[2]This is what is meant above on facing similar school-cohort variances in the income distribution.

educational attainment, which we assume to be influenced by the capacity distribution of their peers: an indicator for what others can achieve. We show that for students with sufficiently high capacities (and therefore income), an increase in the share of low-income peers implies an increase in perceived inequality which leaves them further ahead of their peers. This generates loss of motivation, lower effort, and ultimately lower educational attainment. On the other hand, those students with sufficiently low capacities will see this as a reduction in the inequality of opportunities and will feel less frustrated as they are now less far behind their peers, leading to greater motivation, effort, and educational attainment. Middle-capacity students might experience both situations, rationalizing an average null effect for this income group.

These predictions are based on a theoretical framework that builds on the literature on reference-dependent preferences (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991) and social comparison (see e.g. Clark et al. (2010) and Card et al. (2012)). Additionally, our modelling of heterogenous effects on students' behavior from changes in the income distribution is in the spirit of Genicot and Ray's (2017; 2020) "dual nature" of socially determined aspirations. In our framework, a dual nature comes clear as being surrounded by peers with greater capacities can be either motivating or frustrating, depending on someone's position in the capacity distribution.

Our model provides understanding on how changes in the exposure to income inequality can generate unintended consequences for students' educational outcomes. Moreover, it also highlights a potential mechanism based on students' motivation (or frustration) when choosing effort. To investigate this further, in Section 2.6 we look at empirical evidence on performance in high school based on transcript data and measures of self-esteem, relative intelligence rating, mental health, and motivation. Once again, we find a heterogeneous pattern: low-income students experience a strong, positive effect on performance and improvements in self-esteem and relative self-intelligence rating, while higher income students exhibit an increase in depressive symptoms and decreases in motivation. Altogether with the main results on long-run educational attainment, our evidence is well explained by a model where disparities in income can create contextual effects in the school environment that are unintended.

Social cohesion and integration. Finally, mixing students of different backgrounds may be desirable for many reasons, for example to spread information (Jackson, 2021). In Section 2.7, we turn to an extension asking what may improve the ability of schools to support disadvantaged students facing inequality. Recent work shows that better connectivity (friendships) in school networks improves students' perception of school climate (Alan et al., 2021b) and improvements in social cohesion improve students' outcomes (Alan et al., 2021a). Thus, we propose that better connections in the school network can reduce the effects from changes in peer income distributions. Intuitively, better cohesion may work against inequality effects by allowing students to put less weight on peer income when forming reference points or by learning about their peers' true abilities, feeling involved, and thereby more competitive. Using friendship nominations, we show descriptive evidence in Section 2.7 that social integration through friendships – either better centrality or more cross income group links – moderates the effects from the share of low-income peers on university completion. This holds for both low- and high-income students. We view this extension as descriptive but an important area for further work and policy. Our findings suggest that attempts to expose students to different income backgrounds must be coupled with efforts to improve social integration. Doing so, may help avoid unintended consequences due to reference dependence from inequality in opportunity.

Related literature. Our study relates to a literature on the consequences of inequality for skill development. Much of this literature has focused on how environments during early life affect skill development (for a review see Heckman and Mosso, 2014) and how inequality leads to different incentives for skill investments across low and high SES families (Doepke et al., 2019; Doepke and Zilibotti, 2017a). Additionally, neighborhood inequality has long lasting effects on economic mobility (Chetty and Hendren, 2018a), and children gaining entrance just on the margin to higher quality middle schools in Mexico have been found to achieve lower conscientiousness scores and to shift aspirations away from academics toward vocational tracks (Fabregas, 2022). We contribute to this literature by highlighting the consequences of unabated inequality within peer groups in schools. Furthermore, our results offer an additional explanation for why the benefits of moving to a better quality neighborhood are diminished if a child moves at a later age (Chetty and Hendren, 2018b; Chetty et al., 2016).

Our study further relates to a growing literature on the effects of school environments and peer compositions. These include effects from teacher quality (Chetty et al., 2014; Rothstein, 2017), smaller classes (Angrist and Lavy, 1999; Angrist et al., 2019; Chetty et al., 2011; Krueger and Whitmore, 2001), school spending (Jackson et al., 2015), and tracking students by ability (Duflo et al., 2011; Guyon et al., 2012). Related to these, a recent study by Jackson et al. (2022) finds that the benefits of attending an effective high school for disadvantaged students runs through dimensions unrelated to test score value added. Our study can help shed light here, as this fits with our results on social cohesion representing where and when disadvantaged students may not be harmed by exposure to income inequality.

We also contribute to a broad literature on the effects of peers. A non-comprehensive summary of studies on short-run influences of peers includes the link between peers' persistence and academic achievement (Golsteyn et al., 2021), exposure to low-achieving peers in Kindergarten (Bietenbeck, 2020), spillovers in educational attitudes among friends (Gagete-Miranda, 2020; Norris, 2020), and the effects of peer gender compositions (Black et al., 2013; Borbely et al., 2023; Gong et al., 2021; Lavy and Schlosser, 2011). Studies on the long-run effects of peers include disruptive peers (Carrell et al., 2018), working mothers within peer groups (Olivetti et al., 2020), peer gender effects on university major (Anelli and Peri, 2019), peers' parental education (Bifulco et al., 2014, 2011), peer deprivation and risky behaviors (Balsa et al., 2014), and the effects of high school ability rank on mental health in adulthood (Kiessling and Norris, 2022). More relatedly, Cattan et al. (2023) find elite peers in Norway to positively affect enrollment in elite schools and externally assessed exams.

Our focus is distinct in the literature and demonstrates that peer inequality, or inequality of opportunity, can have important and very different effects across the distribution of students' family income. We examine this within the US context, where inequality can be high.[3] The role of peers in generating frustration or competition may be especially salient in a relatively unequal context, where students can be placed much further away from their reference point than in a more equal environment. Thus, context may matter in shaping peer effects. Evidence from group based games in psychology is suggestive of such contextual responses, showing that when informed about the degree of income inequality in a group individuals with a low socioeconomic status (SES) take more risks and report less satisfaction (Payne et al., 2017). Frustration as a response to income inequality, particularly among low-income students, could additionally help explain why programs relocating adolescents from disadvantaged to advantaged areas have not always found success.[4] Such interventions involve exposing students to a different distribution of income both for the lower- and higher-income students. Our results are consistent with peer inequality creating contextual effects on students, our theory rationalizes this, and our evidence on mechanisms adds further support. Moreover, our evidence on social cohesion suggests that better connections in the school can mitigate the detrimental effects we find from income inequality further emphasizing that context can shape peer mechanisms.

2.2 Data and Variables

We use restricted data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is a longitudinal study representative of middle and high schools in the United States in the mid-1990s. Add Health has several useful features. First, it covers multiple cohorts within schools, which we need for our empirical strategy of exploiting variation within schools across cohorts. Second, a representative set of students from each cohort was first interviewed in 1994/95, when the majority of students were between 12 and 18 years old, and followed for five waves until 2016-

[3] Higher prevalence and salience of inequality in the US is particularly true when compared to countries such as Norway where the ratio between the top and bottom decile of the disposable income distribution is twice as big in the US than in Norway (6.3 vs 3.1, OECD 2018).

[4] For instance, in the Moving to Opportunity experiment adolescent movers experienced on average null or even negative effects (Chetty et al., 2016), while the integration of poor students into elite schools in Delhi improved some pro-social outcomes among existing students but appears to have harmed performance (Rao, 2019).

2018. Third, it includes students' household income, allowing us to observe within school inequality. Our measure to capture changes in the income composition of peers is the share of low-income peers within each student's school-cohort. We then compare long-run educational outcomes and short-run mechanisms as peer income composition changes relative to a student's position in the income distribution.

2.2.1 Income and Capacity

Before proceeding, we provide some descriptive patterns around income and capacity. We refer to capacity as a term to capture the variety of ways that income allows effort to be translated into outcomes. This can be about ability but also about opportunity holding ability constant.

In Figure 2.1, we look at ability, depressive symptoms, and parental investments. We measure cognitive ability using the Add Health Picture Vocabulary Test (PVT) score.[5] Consistent with evidence in the literature on skill trajectories and income (Doepke et al., 2019; Falk et al., 2021), we observe a positive relationship between PVT scores and income that persists when conditioning on school fixed effects (Figure 2.1a). Next, low-income may restrict capacity to achieve holding ability constant through mental health. Adolescents exposed to multiple stressors are at greater risk of experiencing higher depressive symptoms (Thapar et al., 2012), and the conditions of poverty increase uncertainty, adding greater stress (Haushofer and Fehr, 2014; Lichand and Mani, 2020; Mani et al., 2013). Low-income conditions may then expose a student to more stressors, leading to more depressive symptoms, which can reduce motivation and beliefs about the returns to effort (De Quidt and Haushofer, 2019). Supportive of this assertion we see that lower income students tend to score higher on depressive symptoms than do wealthier students (Figure 2.1b) using the Center of Epidemiologic Studies Depression Scale (CES-D, Radloff, 1977a).[6] Finally, we see that lower income

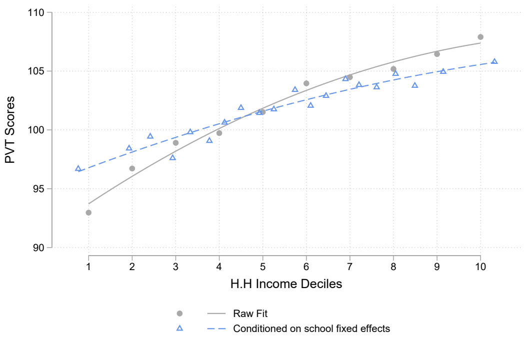
[5] PVT scores in Add Health have been used for ability in a range of papers. Kiessling and Norris (2022) provide discussion on what it measures and show evidence that it is a stable measure of ability.

[6] The CES-D is a often used measure of depression in psychiatric epidemiology. This is a scale measure based on self-reported items that are 1-5 with higher values implying more depressive experiences. AddHealth contains 19 of the 20 items on the full scale for which we follow the literature and collect these into a sum. See Kiessling and Norris (2022) for more description and a lengthy discussion about the CES-D score in AddHealth and see the Appendix Table A.3 for a list.

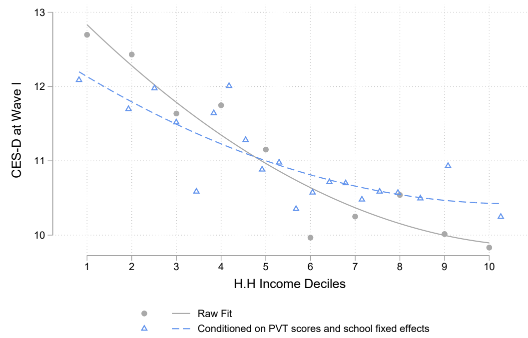
students receive fewer monetary investments from parents (Figure 2.1c), which connects to opportunity. This pattern holds even after conditioning on school fixed effects and PVT scores, implying they are not simply reflecting endogenous school sorting or ability.

Figure 2.1. PVT Scores, CES-D, and Parents' Monetary Investment by Household Income Deciles

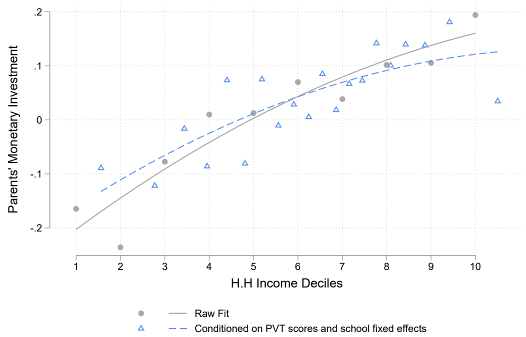
(a) PVT Scores Wave I



(b) CES-D Wave I



(c) Parents' Monetary Investment Wave I



Notes: For each household income decile, we report bin scatter plots with a quadratic fit line of PVT scores in panel (a), CES-D scores in panel (b) and parental monetary investment in panel (c). The bin scatter plot in panel (a) presents a quadratic fit line before and after conditioning on school fixed effects. Bin scatter plots in panel (b) and (c) present quadratic fit lines before and after conditioning on PVT scores and school fixed effects.

The patterns we find are consistent with a multi-dimensional interpretation of what income captures. Exposure to inequality may then signal to adolescents their relative opportunity, leaving an open question on how they will respond educationally across the income distribution.

2.2.2 Definition of Low-income Peers

We define low-income households at wave I of the survey when students were in grades $7-12$ and the majority (72%) in grades $9-12$. We will refer to grades as cohorts. To define low-income households, we first include households below the 1994 poverty threshold for a given family size. Second, we additionally include households who are not below the poverty threshold but who are in the bottom third of the income distribution for each family size.[7] We use this definition to balance sample size for the low-income category against miss-classification and to make sure our peer measure has good support. In robustness checks, we explore alternative definitions and provide more discussion.

Next, we define our peer measure as the leave-one-out share of low-income peers at the school-cohort level. On average, this measure has a 35% share of low-income peers, and it provides near full support (see Appendix Figure A.1a). Additionally, after the inclusion of school and cohort fixed effects, we still maintain considerable variation to identify our effects of interest (see Appendix Figure A.1b). We use this definition to efficiently capture shifts in the distribution of peer incomes based on being around a larger share of lower-income peers versus medium to higher income peers. The mean itself may not capture sufficient variation if what matters is how far one is from their peers, something we discuss more in Section 2.4.[8]

We use the school-cohort as the peer reference group, because we want to define the general environment of income inequality that students are exposed to. Later, in Section 2.7, we compare this against more refined peer reference groups. At that point, we provide intuition and expectations on why the broader environment captures effects that refined groupings may not capture, and we then provide evidence for these expectations.

[7]Family sizes of 8 or more people are grouped together.

[8]Interacting the peer mean with the peer standard deviation of income is another possibility to capture strong changes in the distribution, but this will considerably strain our data, because we need to disaggregate effects across students' own position in the income distribution. We did check results using this approach and found consistent, though less efficient evidence.

2.2.3 Educational Outcomes

To assess the long-run consequences of exposure to income inequality during adolescence, we focus on whether or not a student has completed at least a university bachelor's degree or higher by wave IV of the survey when respondents are on average 28 years old (range: 24-34).[9] We focus on the long-run educational outcome for most of our results, but later we also assess some short-run outcomes on performance in high school. For participants who agreed, Add Health collected their full high school transcript data at wave III. We calculate cumulative GPA excluding courses taken in years prior to the survey year of our treatment. We also construct indicators for whether the student took advanced courses in Math, Science, and English.

2.2.4 Sample Selection and Summary Statistics

Summary statistics for our sample are reported in Columns (1) - (4) of Appendix Table A.1. We first drop observations with missing household income, missing school and cohort identifiers, missing family size, individuals older than 19 at wave I, and individuals from schools with fewer than 20 students in total and 5 students per cohort (6,433 observations).[10] These steps leave us with complete information on the share of low-income peers. Next, we drop those missing information on education level at wave IV (3,174 observations), leaving us 11,165 students in our analytic sample. For all other controls, we impute them to either 0 for discrete variables or to the mean for continuous variables and control for corresponding missing indicators in all specifications.

In our analytic sample, 52% are female and the average age is 15.5 years old in wave I. The majority of students are white (59%), about 17% report at least one foreign born parent, 38% of all students come from university-educated households, and students have on average 34% of peers from low-income families. Moreover, the mean university graduation rate by wave IV (collected in 2008) in our sample is 33%, which is similar to

[9]While there is a wave V, attrition at this wave was much more severe. Our results, though, are very similar if using the wave V sample and education information.

[10]Family size is important for how we define low-income peers thus we drop those missing family size. The restrictions on school and cohort size are standard in the literature using Add Health for peer effect analysis (see Elsner and Isphording, 2018; Kiessling and Norris, 2022).

the national average of 29.4% at the time of the survey (U.S. Census Bureau, 2022). To give a sense of how full sample compares to our analytical sample, we compare means in the Appendix Table A.1 for each variable before and after our sample selection criteria. Though most of the mean differences are statistically significant from zero, we observe relatively small absolute mean differences. We interpret our analytic sample as representative of the full population. Additionally, we provide summary statistics for outcomes that we use in later analyses in the Appendix Table A.2. These include our measures taken from the high school transcript data and measures of self-esteem, beliefs, and mental health that we later use to assess frustration and motivation as mechanisms.

2.3 Empirical Strategy

We need to surmount two hurdles to identify effects from the share of low-income peers. One, selection into schools would likely bias our estimates, if unaccounted for. Two, responses to peer income compositions may be heterogeneous to own-income given the stark differences in opportunity that income can create. Thus, we need to disaggregate effect estimates for the share of low-income peers over the household income distribution and avoid contamination from any non-linear effects that stem from income. We address these problems through (i) using a within school, across cohort design with school and cohort fixed effects commonly deployed in the peer effects literature (e.g., see Sacerdote, 2014) and (ii) highly flexible controls for own-income.

2.3.1 Main Specification

We begin with the following specification:

$$\begin{aligned} Y_{ics} = {} & SLP_{-ics} \times \sum_{k=1}^{10} \mathbb{1}\{IncDecile = k\}\alpha_k \\ & + SD(ln(Inc))_{-ics} \times \sum_{k=1}^{10} \mathbb{1}\{IncDecile = k\}\beta_k \\ & + f(ln(Inc_{ics})) + \boldsymbol{X}_i^{'}\gamma_1 + \boldsymbol{X}_{-i}^{'}\gamma_2 + \boldsymbol{XSD}_{-i}^{'}\gamma_3 + \theta_{ics} + \epsilon_{ics}, \end{aligned} \tag{2.1}$$

where Y_{ics} denotes the university graduation of student i in cohort c and school s and SLP_{-ics} denotes the leave-one-out percentage share of peers from low-income households in cohort c and school s. The coefficients α_k are the marginal effects of SLP_{-ics} at each income decile. We take this as our starting point based on both classical reasons to expect non-linear peer effects (e.g., differential responses to peer ability or by teachers, etc.) and the motivation discussed in the introduction around the potential for inequality to induce frustration among lower income students.

We further include as a control a measure for the dispersion of income in peer groups, the leave-one-out standard deviation of peers logged household income, which we also disaggregate across income deciles. We include this as Tincani (2018) shows that higher order moments of peer distributions can exert separate effects. Including this dispersion measure may capture a ranking mechanism if part of the effect from exposure to the peer income distribution stems from rank concerns in ability and income is correlated with ability. In expanded specifications, we will additionally include ability rank disaggregated over the income distribution and later assess a wide range of checks showing our results are not simply reflecting non-linear peer ability effects. Moreover, controlling for income dispersion may also capture behavioral mechanisms separately from our share (SLP_{-ics}) effects, if those mechanisms correlate with the peer standard deviation of the income distribution. Later, we directly add peer disruption measures as further checks that income picks up other mechanisms.

Next, controls for own-income and school and cohort fixed effects are important for identification. We flexibly control for non-linear effects from own-income by including a cubic polynomial in logged household income. We use this polynomial approach to maintain efficiency rather than including deciles indicators. However, we show in robustness checks that our results are not sensitive to higher order polynomials in income nor are they sensitive to going beyond decile fixed effects by controlling for income ventile fixed effects. To focus on within school, across cohort variation we have school and cohort fixed effects given by $\theta_{ics} = \mu_c + \delta_s$.

We then control for a set of exogenous demographics and characteristics in $\boldsymbol{X}_i$.[11]

[11] These are gender, age and age squared, indicators for race (Asian, Black, Hispanic, White, Other),

In our preferred specification, we supplement these controls by adding peer leave-one-out means for some of these characteristics ($\boldsymbol{X}'_{-i}$), as a way to capture other potential mechanisms that may run through peer compositions.[12] We also add peer leave-one-out standard deviations ($\boldsymbol{XSD}'_{-i}$) for continuous characteristic controls (age and family size) to further capture potential effects from second moments of peer compositions. The error term is ϵ_{ics}.

We could restrict our data further and estimate our effects on sub-samples of own-income. This would allow all covariates to vary by each sub-sample, but the sample sizes would prevent efficient estimation. Thus, we begin with the analytic sample and in a later robustness check consider sub-sample restrictions.

2.3.2 Identifying Assumption

In order to identify the causal effects from the share of low-income peers over the income distribution, α_k, the share has to be as good as randomly assigned. Our assumption, shared with all school-cohort based designs, is that we have exogeneity conditional on a rich set of controls and fixed effects, implying that[13]

$$E\big[\epsilon_{ics}|SLP_{-ics} \times \sum_{k=1}^{10} \mathbb{1}\big\{IncDecile = k\big\}, \boldsymbol{X_i'}, \boldsymbol{X}'_{-i}, \boldsymbol{XSD}'_{-i}, \theta_{ics}\big] = 0. \tag{2.2}$$

Note that while we begin with the disaggregation across deciles of income, based on results from this we then turn to a more parsimonious specification disaggregating over income groups defined as the bottom two deciles, the middle, and top two deciles. In this case, we replace the by decile interaction with $SLP_{-ics} \times \sum_{k=1}^{3} \mathbb{1}\big\{IncGroup = k\big\}$.

In either case, our assumption really rests on two critical components. One, that we adequately capture the relationship between our outcome and own-income, and two,

an indicator for being the child of an immigrant, the family size, indicators for parents' highest degree (less than high school, high school/GED, some college, college degree, postgraduate degree), and an indicator for being raised in a single parent household.

[12] Note that we exclude peer controls in parental education as these could create collinearity problems with our share of low-income peers. We have included them (indicators for whether parents have completed high school, some college education, or post graduate education) in unreported results and they did not change our baseline result but we believe they over-control.

[13] We could also include $ln(Inc)\_SD_{-ics} \times \sum_{k=1}^{10} \mathbb{1}\big\{IncDecile = k\big\}$ in this expectation. We do not show it here to keep things concise.

that we cut any link between determinants of selection into schools and our treatment. For the first, we use a flexible specification in own-income with a cubic polynomial. In later checks, we expand this up to a sixth degree polynomial or replace the polynomials with income ventile fixed effects.

For the second, we select factors likely correlate with SLP_{-ics}. We show evidence of this in the Appendix Figure A.2. This is a scatter plot of SLP_{-ics} against school mean income sorted from low to high among those in the bottom two income deciles (panel (a)), the middle deciles (panel (b)), and the top two deciles (panel (c)). In each case, we see that the raw, uncontrolled correlation is clearly negative. We then show these same scatter plots after removing school fixed effects. Though mechanical, as mean school income is a fixed factor, the plots illustrate our identification strategy showing that with school fixed effects this link is now cut and will also be cut for all other unobserved factors common at the school level. Moreover, we can see that in each segment of the income distribution there remains variation in the residual SLP_{-ics} that we leverage to identify our effects.

Our assumption here implies that parents select into schools based on fixed school factors thereby the school fixed effects remove all unobserved selection factors. We also relax this assumption in some specifications in case parents select schools partly based on school trends, adding these via $\delta_s \times c$ or in other specifications adding school specific income trends. Moreover, we explore an extensive set of robustness checks demonstrating that our results are insensitive and unlikely to be spurious (via placebo testing).

2.3.3 Balancing Test

We now test for evidence consistent with our identification assumptions using balance tests presented in Table 2.1. Each cell in columns (1) - (4) presents a regression of our treatment variable of interest on each row variable. In each test, we control for a cubic in logged household income and school and cohort fixed effects, as these are crucial to our identification. In columns (2) - (4), we restrict the sample around the bottom 20th, the middle, and the top 20th of own-household income to check that our identification assumption is still reasonable within these important groups. Finally, in columns (5) - (7), we repeat this but use the peer standard deviation of logged household income to show that even our additional peer income controls are reasonably exogenous.

Consistent with quasi-random assignment of peers, we observe that most characteristics are not related to our treatment variables. Only the indicator for whether a student is the first-born child seems to be associated with a higher share of low-income peers. Yet, given the number of tests performed is relatively high and the coefficient is small (amounting to less than one percentile score) we interpret the balancing check as strongly consistent with quasi-random assignment of peers.[14]

Table 2.1. Balancing test

	SLP_{-ics}	SLP_{-ics} × B20	SLP_{-ics} × M	SLP_{-ics} × T20	Log(Inc)SD × B20	Log(Inc)SD × M	Log(Inc)SD × T20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.001	0.001	-0.000	0.005*	-0.000	0.001	0.001
	(0.001)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)	(0.009)
White	-0.000	-0.003	0.000	-0.003	0.011	0.001	-0.012
	(0.002)	(0.005)	(0.003)	(0.005)	(0.009)	(0.005)	(0.010)
College-educated Parents	-0.002	-0.006	-0.002	0.002	-0.002	-0.001	-0.005
	(0.001)	(0.006)	(0.002)	(0.004)	(0.009)	(0.003)	(0.006)
Raised by a Single Parent	0.000	0.004	0.000	-0.005	0.000	-0.002	-0.009
	(0.002)	(0.003)	(0.002)	(0.005)	(0.005)	(0.003)	(0.008)
Birth weight (ounces)	0.000	0.000	0.000	0.000	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First-born child	0.003**	0.002	0.003*	0.006**	0.001	0.007**	0.005
	(0.001)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)	(0.005)
Child of an Immigrant	-0.001	0.004	-0.002	-0.001	-0.011	-0.003	0.005
	(0.002)	(0.004)	(0.003)	(0.004)	(0.007)	(0.005)	(0.011)
Household receives food stamps	-0.001	0.003	0.003	-0.013	0.006	-0.000	0.013
	(0.002)	(0.003)	(0.005)	(0.024)	(0.006)	(0.006)	(0.038)
Household size	0.001	0.001	0.001	0.000	-0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)
Function of Log Household Income	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School and Grade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11165	2180	6920	2065	2180	6920	2065

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SLP_{-ics} denotes the leave-one-out percentage share of peers from low-income households in cohort c and school s. Standard errors are in parentheses and clustered at the school level. Columns (1) use the analytic sample; columns (2)-(4) and columns (5)-(7) split the analytic sample by the bottom 20th percentile of household income, the 20th-80th percentiles (endpoints are not included), and the top 20th percentile of households income.

[14]The significant, positive estimate on first-born does show up both on the average of SLP_{-ics} and on SLP_{-ics} for the top-20 income group. We do not think that this is a concern. First-born children often get better resources (Black et al., 2018), thus if anything, we may have predicted an opposite sign effect here. Again, the magnitudes are small, go against our effects and are not persistently significant in columns (2), (5), or (7). Finally, we have confirmed that including or excluding it from our controls does not change our baseline nor mechanism results.

2.4 Results: Long-run Effects on Educational Attainment

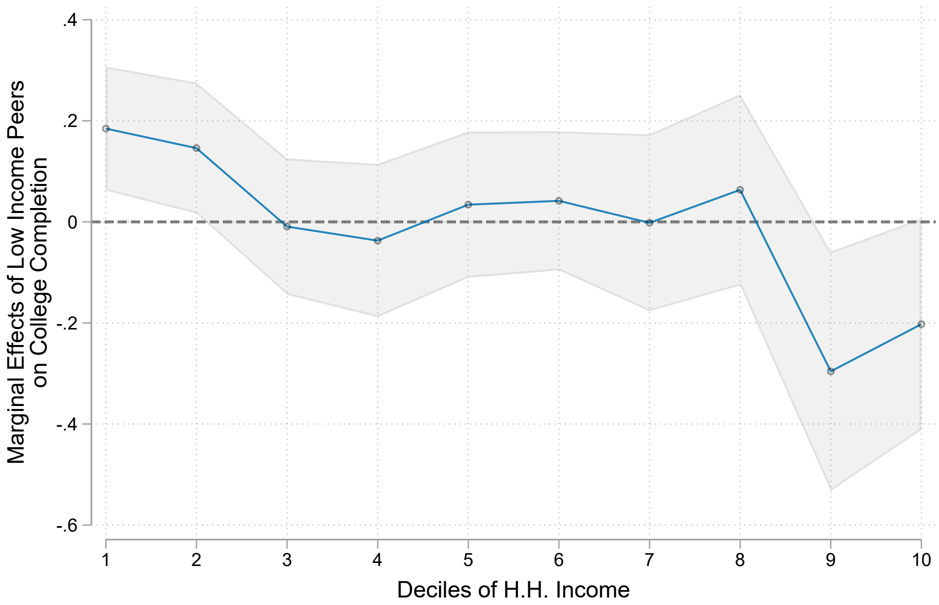
We test the effect from a shift in the share of low-income peers on the probability a student completes a university degree or higher by the wave IV survey across students' positions in the income distribution. We then turn to robustness checks followed by tests for whether our results on income compositions are explained by standard peer effects in the literature.

2.4.1 Baseline Results

We begin by studying the marginal effects from a student's share of low-income peers at wave I on their probability of completing university by wave IV. We use our preferred specification, as discussed in Section 2.3, to calculate the marginal effects (α_k) at each decile of the own-household income distribution at wave I. Figure 2.2 reports the results. We cluster standard errors at the school level here and in all results to follow. We find positive and significant effects for lower-income students (bottom two deciles), null effects over the middle, and negative and significant effects for higher-income groups (top two deciles). These effects are consistent with the idea that shifts in the degree of income inequality create different responses across the income distribution.

To empirically shed light on the sharp cutoffs in effects, we calculate the gap between the individual logged household income and the school-cohort peer mean of logged household income to give the percentage difference (gap: $ln(Inc_{ics}) - ln(Inc_{-ics})$). In the Appendix Figure A.4, we present plots of the interquartile range, median, and mean for this gap over household income deciles. We see that students in the first two deciles are much further behind their peers than better off students. Even students in the third decile are considerably less far behind their peers than those in the first two deciles. Next, for the top household income deciles, we see that those in the ninth and tenth deciles are consistently much further ahead of their peers. Overall, we think the patterns on distance discussed above support our findings in Figure 2.2, suggesting that being further away from one's peers drives the effects as it is predominately the bottom and top two income deciles that have much distance. This does not, however,

Figure 2.2. The share of low-income peers and effects on university completion over deciles of the own-household income distribution



Notes: This figure presents the marginal effects on the probability of completing university by wave IV of the survey from the leave-one-out mean (share) of low-income peers in the same-high school and cohort (wave I). The effects are calculated at each decile of the own-household income distribution at wave I.

explain why being further away to either side should matter.

In the next section we turn to a more parsimonious specification, and after a series of robustness checks we also show, in Section 2.4.4, that our empirical results are not explained by a variety of mechanisms that can be drawn from the literature. Hence, in Section 2.5 we develop a model of social comparison and students' effort choice that can rationalize our findings, and subsequently provide some evidence in support of its main mechanisms.

2.4.2 Baseline Results: Parsimonious Specification

Based on the by decile results, we group the distribution of own-household income into the bottom 20th, middle, and top 20th. We then use these groups to disaggregate the effect from the share of low-income peers. In Table 2.2, we present the results across multiple specifications in columns (1) - (6), finding stable results across specifications.

Interpreting our preferred specification (column 2), we find that for the bottom 20th of the household income distribution in high school, a 100% shift in the share of low-income peers yields a 18 percentage point (pp) increase in the likelihood of holding at least a four year degree by wave IV. For the middle group, we find null effects, and for the top 20th of household income the marginal effect is a 25pp decrease. A 100% shift, however, is not realistic. Interpreting these in standard deviation shifts (20%) translates the effect for the bottom 20th into a 3.6pp increase and for the top 20th into a 4.1pp decrease.

The estimates for the bottom and top 20th groups are significantly different across all specifications. One concern is that multiple hypothesis testing within and across specifications could lead to false rejections of the null (Clarke et al., 2020). To account for this, in the Appendix, Table A.11, we report Romano Wolf p-value adjustments across all specifications based on a block cluster bootstrap around schools. Although we obtain higher p-values, our results remain statistically significant at the 5% level for the bottom 20th group and at the 10% level for the top 20th group.

To give some context to the effect estimates, we compare them to the average probability of having at least a four year degree split across income groups. The overall average in our analytic sample is 33%, which breaks into 15% for those in the bottom 20th of the household income distribution in high school, 31% for the middle group, and 59% for the top 20th group. Thus, for the bottom 20th students the effect from the share of low-income peers amounts to a 24% increase from the group mean, whereas the effect within the top 20th group is only about 7%. We also compare these effects to conditional university completion gaps over gender and socioeconomic differences, detailed in the Appendix, Figure A.5.[15] These results are sizeable for low-income students and of similar magnitude to other interventions targeting low-income families and their children. For perspective, the magnitude of our effects is comparable to financial assistance programs, such as the Social Security Student Benefit Program,

[15] The standardized effect for the bottom 20th group amounts to about half of the gap between females and males, around 40% of the gap between university and non-university parents, and is similar in size to the gap between single and two-parent homes. Comparisons are similar looking at the top 20th group.

Table 2.2. Baseline effects on university completion: Share of low-income peers

	University Graduate					
	(1)	(2)	(3)	(4)	(5)	(6)
SLP_{-ics} × Bottom 20	0.18**	0.18**	0.16**	0.19***	0.27***	0.22**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)	(0.10)
SLP_{-ics} × Middle	0.01	0.02	0.01	-0.00	0.07	-0.02
	(0.07)	(0.07)	(0.06)	(0.06)	(0.09)	(0.07)
SLP_{-ics} × Top 20	-0.25**	-0.25**	-0.27**	-0.27**	-0.19	-0.29**
	(0.11)	(0.11)	(0.11)	(0.11)	(0.13)	(0.13)
Peer Log(Inc) (SD)	Yes	Yes	Yes	Yes	Yes	Yes
Own-Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School and Cohort FEs	Yes	Yes	Yes	Yes	Yes	Yes
Peer Effects (means)	No	Yes	Yes	Yes	Yes	Yes
Peer Effects (SD)	No	Yes	Yes	Yes	Yes	Yes
Own-Ability Polynomials	No	No	Yes	Yes	Yes	No
Ability Rank × Income Position	No	No	No	Yes	Yes	No
School-specific Cohort Trends	No	No	No	No	Yes	No
School-specific Income Trends	No	No	No	No	No	Yes
Mean University Graduation	0.33	0.33	0.33	0.33	0.33	0.33
Observations	11,165	11,165	11,165	11,165	11,165	11,165
R^2	0.241	0.243	0.263	0.264	0.273	0.253
Difference between B20 and T20	<0.001	<0.001	<0.001	<0.001	<0.001	0.002

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SLP_{-ics} denotes the leave-one-out percentage share of peers from low-income households in cohort c and school s. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Estimates of marginal effects of SLP_{-ics} are for those in the bottom 20th percentile of household income, in the middle, and finally in the top 20th percentile of household income. Peer Log(Inc) (SD) denotes the standard deviations of peer log income. We always include a 3-degree polynomial of log household income in the own characteristics control. Ability rank means the ability rank within school cohorts.

a large financial assistance program paid to children of deceased, disabled, or retired Social Security beneficiaries in the US to finance post-secondary education (Dynarski, 2003). Effect estimates suggest that an offer of $1,000 in grant aid corresponds to an increase in the probability of high-school students attending university by about 3.6pp (Dynarski, 2003).

Next, in columns (3) - (4), we check our results against the inclusion of flexible controls for own-ability and rank. We include a quartic polynomial in the PVT scores and control for the peer (school-cohort) leave-one-out mean as well as the standard deviation in PVT scores (column 3). We also check that our effects are not driven by a rank mechanism, as a wide literature illustrates the importance of relative ability (Bertoni and Nisticò, 2019; Denning et al., 2021; Elsner and Isphording, 2017; Murphy and Weinhardt, 2020). Thus, we next add the PVT school-cohort rank, which we disaggregate by students' position in the bottom 20th, middle, or top 20th income group (column 4).[16] Our key results on the bottom and top 20th groups remain consistent and significant.

In columns (5) - (6), we now add school-specific trends to relax the assumption that selection factors are captured by the school fixed effects. First, in column (5), we include the expansive specification from column (4) and allow for a linear trend within schools. This specification allows for selection based on linear-trends, but is the most restrictive on the data. Second, in column (6), we use our preferred specification as in column (2) but allow for school specific trends across our defined income groups. In both cases, we find very similar results to those in our simpler specifications.

Finally, we consider a different outcome by using the natural log of individual income at wave IV. These results are reported in the Appendix Table A.4. We find that wave IV income improves for the bottom 20th household income group at high school in response to an increase in the share of low-income peers, while for the top 20th group, we see null effects on wave IV income. Note, that for top income students, the effect size on university completion as a percent of the mean is much smaller than it is for lower-income students. Also, it may be that those from higher parental income backgrounds are better positioned to maintain higher-income regardless of their university completion status. This question is beyond the scope of our paper. Nevertheless, the pattern of results suggests strong effects for the bottom 20th group that are different from the experience of the top 20th group.

Altogether, the results here suggest the presence of strong, heterogeneous effects stemming from peer income inequality. We further examine additional heterogeneity within each income group across student characteristics using a causal forest (Athey et al., 2019). See the Appendix Section A.6 for a detailed description of the method and

[16] In an additional checks against alternative mechanisms (see Section 2.4.4), we will go even further and allow for a wide range of non-linear peer ability effects and also consider income rank effects.

results. We find that our pattern of results on the effects from the share of low income peers across income groups remains consistent when estimated with a causal forest (see the Appendix Figure A.9a). Additionally, we see that these peer effects within income groups are generally persistent across other student characteristics, including across the ability distribution.

2.4.3 Robustness Checks

In this section, we report a series of additional analyses to probe the robustness of our results.

Definitions for the share of low-income peers. We define low-income households as those whose household income is either below the 1994 poverty threshold or in the bottom third of the income distribution for a given family size. We then calculate the leave-one-out share of low-income peers at the school-cohort level based on this definition. Yet, other definitions of low-income households are conceivable for assignments of the share of low-income peers for those students who are in the same school cohort and have the same household income. For instance, we could define low-income households based on (i) the bottom 20th percentile of the income distribution for a given family size, (ii) below the median of the income distribution for a given family size, or (iii) the bottom third of the household income distribution based on school region, school urbanicity, and family size (grouping households whose family size is equal or larger than 5).

Of these, we expect most results to be similar except for the below median definition to introduce measurement error by misclassifying a larger share of students as low-income peers when they are not, implying it should return smaller and less precise effects. Moreover, definitions that shrink the size of the low-income peer groupings have another tradeoff in that they reduce the degree of variation available within schools thereby potentially yielding less efficient results. In Appendix Table A.5, we compare results from these different definitions. We find similar effects across definitions except for the below median definition where we find weaker effects, as expected, and some less efficient results where the definitions are more stringent. Importantly, the results – absent the definition by the median – are stable. Generally, our current definition of low-income households seems reasonable to capture the stratification of household income.

Non-linearity in household income. In our main specification, we adopt a cubic polynomial in logged household income to take the relation of university graduation and own-income into account. Yet, one might worry that we have not captured all the relevant non-linearity between our outcome and logged household income. In Appendix Table A.6, we therefore examine different polynomials up to the sixth order. We find that our results are highly robust regardless of the degree we control for. Moreover, we include a specification with indicators for each ventile level of the logged household income, which non-parametrically controls for different own-income levels, and find our results remain unchanged.

Subsample by income groups. In our main specification, we disaggregate our results by own-household income groups for being in the bottom 20th, the middle, and the top 20th. While we gain efficiency from this specification, we do not allow all covariates to vary by each subsample. In Appendix Table A.7, we examine the consistency of our results by splitting the sample over each of the income groups we use. We start from our baseline specification and then add a quartic own-ability polynomial and the school-cohort ability rank as an additional check. We find that our subsample results for the share of low-income peers are consistent with our main results. While the results slightly lose some efficiency, we find the point estimates are quite stable and robust.

Placebo tests. Our identification strategy assumes that the share of low-income peers is as good as randomly assigned conditional on own income and school and cohort fixed effects. One way to test against failures of this assumption is with placebo tests. In Appendix Table A.8, we reproduce our main specification results with an alternative treatment variable and then with an alternative outcome variable. As for the placebo treatment, we take the share of low-income peers within the same school but from a different cohort with a 1-year or 2-year time gap. As for the placebo outcome, we use an indicator for ever repeating a grade in the past. This is a pre-treatment placebo outcome. Given our identification assumptions hold, we would not expect a link to past repetition of school grades. For the bottom 20th group of own-household income, both placebo tests yield an expected zero. For the top 20th group, we do find some correlation between the placebo treatment and our university graduation outcome, but this effect disappears once we control for school-specific income trends. As is shown in column (6) of Table 2.2, our point estimates stay consistent when we control for school-specific income trends. These placebo tests are highly consistent with our identifying assumptions and suggest that our main model is well identified.

Attrition. In wave IV, approximately 14 years after the treatment in wave I, about 78% of the baseline sample remains.[17] Appendix Table A.9 shows that attrition patterns do not differ by treatment status across own-household income groups regardless of the school and cohort fixed effects we control for. We further assess the robustness of our results to accounting for attrition in two ways. First, we calculate treatment effects using inverse probability weighting (IPW), where the weights are calculated as the predicted probability of being in the wave IV follow-up sample based on the main specification controls and an additional variable for whether the family was willing to move.[18] Second, we use the wave IV sampling weights provided by Add Health to adjust for non-response in longitudinal models. Our results survive parametric corrections for attrition using either IPW or sampling weights in Wave IV.

Random sampling of students per school. The impact of sampling error on estimates is not entirely clear in a nonlinear model with group means. We assess the consequences of observing a random sub-sample of students per school using a simulation. The data generating process (DGP) is specified in Appendix A.3 and is

[17]Note that the baseline sample is defined after our initial set of sample selection criteria but before dropping those missing information on education level at wave IV.

[18]We then replicate our results with IPW weights using the specifications in column (2) and column (6) of Table 2.2.

based on our estimated values for the share of low income peers. We simulate 500 schools of 240 students and decrease the share of students in our sample from full saturation, where all students in a school are sampled, to a situation where we observe only 10% of all students.[19] Our simulations in Appendix Figure A.6 show that it leads to attenuation for the bottom and top 20th income group students, while the middle income group shows a small upward bias. We then repeat the simulations based on estimates for subgroups by income in Appendix Figure A.7 and find that sampling variation attenuates the estimates when the true coefficient is non-zero for both the bottom and top 20th income groups and has no effect on the middle income group where the true effect is set to 0.

Measurement error in income. We then turn to measurement error in our income measure. Specifically, we assess how our estimates change when we introduce noise to the measure of income, i.e., we measure $ln(\tilde{Inc}) = ln(Inc) + \phi \cdot v$, where $ln(Inc) \sim \mathcal{N}(3.5, 0.85)$, $v \sim \mathcal{N}(0, 0.85)$, and $\phi \in [0, 1]$. Thus $\phi = 0$ corresponds to situations where we have no measurement error, while $\phi = 1$ corresponds to situations where we allow as much measurement error as noise in our income measure. This measurement error then translates into measurement error in the share of low income peers that each student faces. We then combine the measurement error in income with random sampling of students, taking the situations with 100% and 30% sampling, respectively. In Table A.10, we report estimates of the effect for the share of low income peers among the bottom 20th, middle, and top 20th income groups along with the ratio of the estimated effect to the true coefficient in parentheses. For the middle group, the ratio is not reported because the true coefficient is 0. Our simulation results show that for the bottom 20th and top 20th income groups the effects are attenuated, resulting in an underestimation of the true effects.

[19]Our approach is adapted from the designs used by Elsner and Isphording (2017) and Kiessling and Norris (2022) to assess measurement error for ability rank effects.

2.4.4 Results Explained by Common Peer Effects in the Literature?

Effects from shifts in the share of low-income peers could be rationalized by responses from students, teachers, and parents. We now turn to whether commonly observed mechanisms in the literature explain the patterns we observe on income inequality in peer groups. We investigate whether our heterogeneous peer income effects seems to pick up the following set: non-linear peer ability and ability rank effects; responses from teachers to changes in peer income compositions; changes in disruptive behaviors; and parental responses to changes in the peer income composition. We provide a thorough discussion of each of these and report our results in the Appendix, Section A.5.1

We find no evidence that the heterogeneous effects from the share of low income peers documented in Sections 2.4.1 and 2.4.2 are explained by this wide range of plausible mechanisms. Next, we propose a novel explanation to rationalize our findings. It is based on reference dependence and the idea that social comparison among students can generate both frustration and motivation depending on a student's relative position in the income distribution.

2.5 A model of Social Comparison and Student Effort

We propose a theoretical model of student effort choice towards the achievement of an educational outcome. The model provides a lens to understand how exposure to income inequality may generate the patterns we have observed.

Our model is based on two premises. First, we think of income as a salient and observable characteristic signaling students' capacity, in-line with the evidence we present in Section 2.2. Moreover, capacity here can be thought of as differences in opportunity even holding raw ability fixed. Second, we consider a possible non-monotonicity in the effect that social comparison on this capacity dimension can have on students' behavior. Our model provides a novel perspective on how exposure to income inequality, and therefore to inequality of opportunities, may affect students' effort and their educational attainment. Thus, we consider how income inequality may affect the contextual environment students live within.

To capture this we build on reference dependence (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991) and social comparison (see e.g. Clark et al. (2010) and Card et al. (2012)). More precisely, we assume that a student's reference point for educational attainment is, at least in part, determined by the capacity distribution of their peers—a proxy for what others can achieve. Further, in the spirit of Genicot and Ray's (2017; 2020) model of socially determined aspirations, we provide a framework in which changes in the capacity distribution of peers can have heterogeneous effects across students depending on their relative position in the income distribution.

We place particular emphasis on the effects of changes in the composition of peers' family income on students' behavior. The theoretical framework explains a contextual effect, but alternatively, our framework can also be considered a reduced-form model of students' best response with non-linear peer effects in which students use peers' income as a salient and observable indicator of what others can achieve.[20]

In fact, the reference point for educational attainment in our model is an artifact which, together with our assumption of reference-dependent preferences, enables us to capture the effect of inequality in opportunities on a student's utility.

2.5.1 Preferences

Students are endowed with initial capacity θ defined as the combined set of factors that enable a student to translate effort into educational outcomes. In particular, we assume that capacity is a strictly increasing function of both ability s and income I. That is, $\theta \equiv \theta(s, I) > 0$, with $\theta_s > 0$ and $\theta_I > 0$, and that the only source of heterogeneity in capacity in our model is income.[21] Denote the distributions of income and capacity

[20] An extension of our framework could also incorporate the possibility that students' reference point for social comparisons is influenced by their beliefs about peers' effort (in addition to peers' capacities). This would then generate strategic complementarities between students' effort, as in game-theoretic foundations of social interaction peer effects (see e.g. Boucher et al. (2024)). Such an extension would, however, require a more specific form of the students' reference point than the one we consider below, as well as the characterization of an equilibrium in which students' beliefs are consistent. We believe this extension to be interesting, but also beyond the scope of the model developed in this paper. Moreover, it can be deduced that the predictions we derive to rationalize our empirical findings, which are based on changes in the composition of peers' capacities, will also hold in a more elaborated model with strategic complementarities in effort.

[21] This assumption is made for simplicity and to capture the fact that income is more salient and easily observable than ability. For instance, in the absence of complete information about peers' abilities, this

by F^I and F^θ respectively. Our assumption implies that F^θ is a transform of F^I: the distribution of capacities a student faces in school captures within-school income inequality.

Students choose effort e to achieve an educational outcome y, realized attainment, given by $y \equiv y(e, \theta) = \theta e$. Further, to capture the effect of social comparison and inequality of opportunity on behavior, we assume that students compare their realized outcome in relation to a reference outcome r which is influenced by the capacity distribution they face. In particular, we assume r to be positively related to someone's own capacity as well as to the capacity of their peers. More formally, $r \equiv r(\theta, F^\theta)$ with the following properties: i) $r_\theta > 0$; ii) $r(\theta, \hat{F}^\theta) > r(\theta, F^\theta)$ when $\hat{F}^\theta$ first-order stochastically dominates F^θ; and iii) $r(\lambda\theta, F^{\lambda\theta}) = \lambda r(\theta, F^\theta)$ for all $\lambda > 1$, where $F^{\lambda\theta}$ denotes the distribution of θ when all capacities increase by λ. This last assumption implies that if all capacities increase by the same proportion, then r increases by the same proportion.[22]

Students' preferences are characterized by the following additively separable utility function:

$$u(e, y, r) = b(y) - c(e) + \mu(y - r), \tag{2.3}$$

where $b(y) = y^\alpha$, $\alpha \in (0, 1)$, captures the benefit from achieving the outcome y; $c(e) = e^2/2$, is the cost of effort (where the marginal cost is normalized to e); and $\mu(y - r)$ captures the effect of social comparison over outcomes on a student's utility. We assume μ to be a reference-dependent gain-loss function, such that $\mu''(y - r) < 0$ if $y > r$ (i.e. concavity over gains) and $\mu''(y - r) > 0$ if $y < r$ (i.e. convexity over losses):

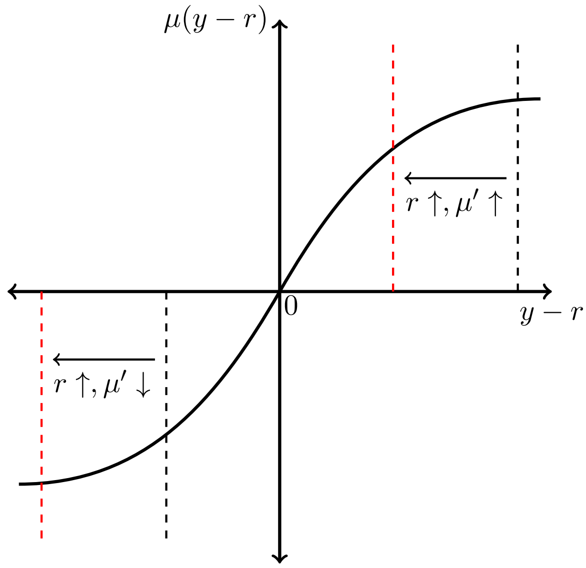
$$\mu(y - r) = \begin{cases} [y - r]^\beta & \text{if } y \geq r \\ -[r - y]^\beta & \text{if } y < r; \end{cases} \tag{2.4}$$

assumption would imply that students' use income as a proxy for capacity. This set up also enables a closer matching between our theoretical model and the empirical analysis, in which we control for peers' ability.

[22]Our formulation of the properties of r is inspired by the model of socially determined aspirations in Genicot and Ray (2017, 2020). In particular, $r(\theta, F^\theta)$ satisfies both "scale-invariance" and "social monotonicity".

where $\beta \in (0,1)$.[23] The properties of μ, combined with our assumptions on r, are a central component of our model, capturing the effect of inequality of opportunity, due to income inequality, on students' behavior. Figure 2.3 plots μ as a function of $y - r$ when θ and e are fixed, and shows the partial effect of an increase in the reference outcome r on the slope of μ: the marginal returns of effort that stem from reference dependence. For instance, consider a student with a relatively high θ' and such that $y > r$, implying they perceive additional satisfaction from achieving the educational outcome y (the student is in the gain domain, the upper-right panel of Figure 2.3). In this case, an increase in peer income, and therefore peer capacity, generates an increase in the marginal returns to effort (r increases and the slope of μ becomes steeper), and an increase in effort will increase utility. As we will formally establish later, this effect can be interpreted as greater motivation stemming from a reduction in inequality of opportunity between the student and their peers.

Figure 2.3. The Gain-loss Function



[23]This formulation is in the spirit of Kahneman and Tversky (1979) value function under riskless choice (Tversky and Kahneman, 1991). In particular, our function μ displays both "reference dependence" and "diminishing sensitivity", but it does not feature "loss aversion". Note that while both reference dependence and diminishing sensitivity are crucial ingredients of our model, the consideration of loss aversion—despite adding one additional parameter and layer of complexity—would not affect our qualitative predictions. Moreover, while there is ample evidence of the existence of loss aversion in the evaluation of monetary/material payoffs, less is known about its role in less tangible domains such as that of educational outcomes.

Instead, consider a student with a relatively low θ'' and such that $y < r$, implying they perceive a sense of frustration, which negatively affects utility (the student is in the loss domain, the lower-left panel of Figure 2.3). Here, an increase in peer income generates a decrease in the marginal returns to effort (r increases, but the slope of μ becomes flatter), implying that decreasing effort will increase utility: as the inequality in opportunity between the student and their peers widens, higher frustration dampens the incentive to exert effort.

In the remainder of this section we will formally characterize the consequences of these changes in the reference outcome on a student's choice of effort. Subsequently, we will establish how shifts in peers' income can affect effort differently depending on the student's relative position in the income distribution.

2.5.2 Capacities, Peers, and Students' Effort

Consider a student endowed with capacity θ, facing capacity distribution F^{θ}, and with reference outcome r, that needs to choose effort e to maximize their utility as given by (2.3). The first-order conditions characterizing this maximization problem are given by:

$$\alpha[\theta e]^{\alpha-1}\theta + \beta[\theta e - r]^{\beta-1}\theta = e \qquad \text{if } y > r, \tag{2.5}$$

$$\alpha[\theta e]^{\alpha-1}\theta + \beta[r - \theta e]^{\beta-1}\theta = e \qquad \text{if } y < r, \tag{2.6}$$

where the left-hand side captures the marginal benefit of exerting effort, while the right-hand side is the marginal cost. The solution, denoted by $\tilde{e}(\theta, r)$, is the level of effort at which these are equal.[24]

Because the marginal benefit of effort crucially depends on the gap $y - r$, we know from the preceding discussion that the properties of $\tilde{e}(\theta, r)$ might differ depending on whether the student is experiencing frustration $y < r$, or greater motivation, $y > r$, in

[24] Our assumptions over μ imply that there may be at most two solutions when $y < r$. To proceed, we only consider the one according to which a student's effort would be decreasing in its marginal cost: a student with higher marginal cost would exert less effort than a student with lower marginal cost (note, however, that marginal cost is normalized to e in this model for simplicity). See Appendix A.1 for more details.

achieving the educational outcome y.

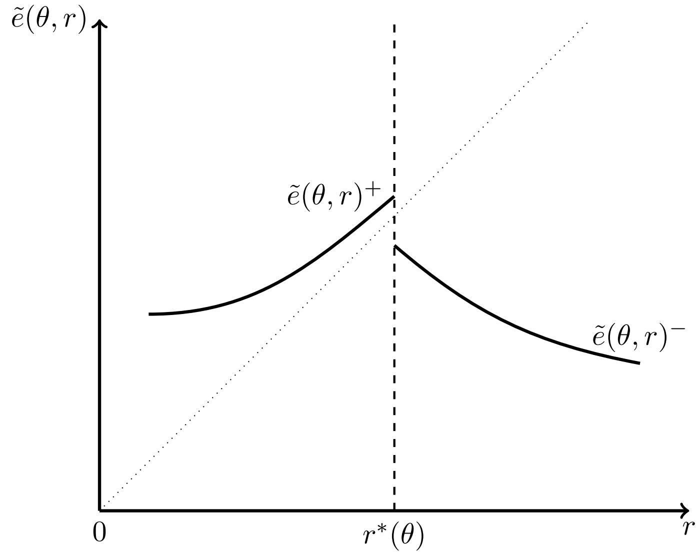
To see this, consider a student that is currently perceiving additional satisfaction so that $y > r$ at the optimal effort, which is the solution to (2.5) and denoted by $\tilde{e}^+ \equiv \tilde{e}(\theta, r)^+$. To understand how changes in the reference outcome can affect behavior in this case, suppose that r increases. For example, this could stem from the student being exposed to peers with higher income, and therefore higher capacities. In this case, the marginal benefit of increasing effort is higher, implying that the student will exert more effort to achieve a better educational outcome. However, higher effort is increasingly costly and the marginal benefit of achieving better outcomes decreases. As we formally establish below, this implies there exists a threshold reference outcome r^* beyond which utility is maximized by the solution to (2.6), denoted by $\tilde{e}(\theta, a)^-$. In this case, $y < r$, the student perceives frustration, and further increases in r will decrease the marginal benefit of effort, resulting in lower effort and worse educational outcomes.

Proposition 1. *For a given θ, there exists a unique reference outcome r^* such that: if $r < r^*$, then $y(\tilde{e}^+, \theta) > r(\theta, F^\theta)$ and optimal effort $\tilde{e}(\theta, r)^+$ is increasing in r; and if $r > r^*$, then $y(\tilde{e}^-, \theta) < r(\theta, F^\theta)$ and optimal effort $\tilde{e}(\theta, r)^-$ is decreasing in r. Moreover, $r^* \equiv r^*(\theta)$ is increasing in θ.*

Proposition 1 establishes that the effect of changes in a student's reference outcome is non-monotonic: effort and educational outcomes are increasing in r for all $r < r^*$ and decreasing in r for all $r > r^*$. This relationship is plotted, for a given θ, in Figure 2.4. In fact, since students are heterogeneous in θ, due to differences in income, there exists a distribution of r^*: each student has a different reference threshold depending on their capacity, and the greater is their capacity, the higher this threshold will be. Intuitively, the greater is a student's capacity, the larger the increase in their peers' capacity, and therefore r, will have to be before they feel frustrated by their peers having greater opportunities to achieve higher outcomes.

The results established in Proposition 1 characterize a mapping between effort and reference outcomes, by taking as given an individual student's capacity θ. However, both the reference outcome $r(\theta, F^\theta)$ and the reference threshold $r^*(\theta)$ are functions of θ. This suggests that for a given distribution of capacities F^θ, whether a student

Figure 2.4. Optimal Student Effort



perceives satisfaction, or a sense of frustration, crucially depends on their capacity in relation to the ones of their peers, that is, it depends on their position in the income distribution. The following proposition formally establishes this role for a student's initial capacity endowment.

Proposition 2. *For a given F^θ, there exists a unique threshold θ^* such that, for all $\theta < \theta^*$ then students are frustrated, while for all $\theta > \theta^*$ then students are satisfied.*

Proposition 2 states that students with lower capacities are more likely to be in the frustration zone than students with higher capacities. This result bears important implications for the effect of changes in the composition of the peer capacity distribution on students' behavior and educational outcomes. For instance, being exposed to peers with higher capacities and opportunities may be beneficial for students at the highest end of the capacity distribution, but detrimental for students at the lowest end of the capacity distribution.

2.5.3 Predictions and Empirical Counterpart

At this stage, it is possible to use the results just established to illustrate how our the model can rationalize our empirical findings. First, note that our assumptions on the determinants of students' capacities immediately imply that, for a given income distribution F^I, there exists a unique threshold income, which we denote by I^*_F, such that students with income $I < I^*_F$ are frustrated, while students with income $I > I^*_F$ are in the satisfaction zone. This also implies that we can express the reference outcome in terms of income: $r = r(I, F^I)$. Next, we can classify students in terms of their relative position in the income distribution F^I. For a given $\varepsilon > 0$, where ε is large enough, denote with $I^l_F \equiv I^*_F - \varepsilon$ and with $I^h_F \equiv I^*_F + \varepsilon$, and define "low income" students those endowed with $I \leq I^l_F$, "high income" students those endowed with $I \geq I^h_F$, and "middle income" students those endowed with $I \in (I^l_F, I^h_F)$.

In the next proposition, we characterize the response of students to a change in the composition of the income distribution they face, which mimics our empirical analysis. In particular, we will do a comparative statics exercise across the income groups defined above, where we shift the income distribution from F^I to G^I such that G^I contains a larger share of low income peers. Hence, we consider a distribution G^I that is first-order stochastically dominated by F^I. In our model this implies that students' reference outcome will be lower, with heterogeneous effects across the income distribution. For simplicity, we will assume that even the richest of the low income students remains frustrated.

Proposition 3. *Consider a shift in the income distribution from F^I to G^I, such that $G^I > F^I$ and $r(I^l_F, G^I) > r^*$. Low capacity students will increase effort and achieve better educational outcomes, high capacity students will decrease effort and achieve worse educational outcomes, while the effect on middle income students is ambiguous: while those endowed with $I \in (I^*_F, I^h_F)$ will decrease effort, those endowed with $I \in (I^l_F, I^*_F)$ will increase effort, only as long as $r(I, G^I) > r^*$.*

Proposition 3 establishes the existence of heterogeneous effects of a change in the composition of peers' income on students' educational attainment, which is conditional on their relative position in the income distribution. Through the lens of our model we can interpret this result as follows. An increase in the share of low income peers will reduce the inequality of opportunity from the perspective of low income students, who will now feel less frustration and greater motivation (as the marginal benefit of effort is greater), leading to an increase in effort and higher educational outcomes. On the other hand, from the perspective of high income students there is an increase in the inequality of opportunity which leaves them even "further ahead of their peers". This generates a loss of motivation (as the marginal benefit of effort is smaller) and a drop in effort, which ultimately translates into lower educational attainment. For middle income students, the effect is qualitatively ambiguous: some of these students will see a reduction in the inequality of opportunity and feel less frustrated, while others although feeling satisfied to be ahead, will loose motivation and decrease their effort.

This result rationalizes our empirical finding that, controlling for students' ability, an increase in the share of low-income peers has positive effects on low-income students, negative effects on high-income students, and null effects on middle-income students (see Figure 2.2 and Table 2.2 in Section 2.4). Moreover, our theoretical model suggests a potential mechanism based on student effort and generated by heterogeneous effects on students' motivation and frustration depending on their relative position in the income distribution. In the next section, we investigate the empirical plausibility of this mechanism.

2.6 Results: Effort, Frustration, and Motivation

We look now at short-run measures in terms of high-school performance and then at measures related to frustration and motivation.

2.6.1 High-school Performance

Although we lack a good measure of pure effort, Add Health has excellent measures of high-school performance from transcript data which we use to proxy effort. Our baseline results on university graduation and our model predictions suggest there should be non-linear effects on performance. We start with self-reported grades and then use high school transcripts collected by Add Health for all participants in the wave III survey, who agreed, and for whom the transcripts were accessible. To overcome attrition at wave III and from wave III into the transcript sample, Add Health constructed specific non-response weights for the education transcript data, which we use in the following analysis.[25] We calculate each person's cumulative GPA from the year of their wave I survey (time of our treatment) to the end of high school.[26] Also, we construct separate indicators for whether someone chose to take an advanced course in Math, in Science, or in English anywhere from the time of their wave I survey to the end of high school.[27]

In Table 2.3, we report effect estimates for a shift in the share of low-income peers using our baseline specification. With self-reported GPA (column 1), we observe null effects, but with transcript cumulative GPA (column 2), we observe a strong, positive increase in GPA for the bottom 20th group. We also see a positive, but smaller, effect for the middle-income group and a null for the top 20th.

We then look at the choice to take advanced courses (columns 3 - 6). The bottom 20th income group continues to respond positively to an increase in the share of low income peers. They are significantly more likely to take advance Math and to take more than one advanced subject. We see no change for the middle income group, and the top 20th group have mainly null results with a marginally significant negative effect on taking Advanced Sciences. We also repeat the Romano Wolf p-value adjustment conducted at the baseline to check that our inference is not driven by multiple hypothesis test bias (see Table A.12 in the appendix). We find that the key results here for the bottom 20th group survive this adjustment.[28]

While the results here point toward effort responses, they could rather be explained by grading on a curve. If low-income students tend to have lower grades than high income students, then having more low-income students in a cohort means that these students are on average in classes with lower overall grades. In this case, grading on

[25]Wave III was collected over 2001 and 2002 with participants in young adulthood aged roughly 18-24.

[26]For example, this means that for someone in 10th grade at the wave I survey we calculate their GPA from 10th-12th, while for someone in 12th at the wave I survey we use only their 12th grade scores.

[27]Core required credits for graduation are set by each state, but advanced courses are often at the choice of the student in an effort to pursue University entrance. For Math, this is defined by taking pre-calculus or calculus. For Science, it is whether one took advanced science or physics. For English, it is whether one took an honors English class.

[28]We have restricted the sample to those present in our baseline analysis, meaning we drop those who are missing data for university completion. In the Appendix Table A.17, we also report the results where we include even those who are not present in the baseline analytic sample. These are generally similar to our results in Table 2.3 and qualitatively yield similar conclusions.

Table 2.3. GPA and advanced Courses

	GPA		Advanced Courses				GPA		
	Self	Transcript	Math	Science	English	More than one	Transcript	Transcript	Transcript
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SLP_{-ics} × Bottom 20	0.05	0.81***	0.36***	0.25	0.13	0.47***	0.75***	0.96**	0.60**
	(0.15)	(0.25)	(0.13)	(0.16)	(0.22)	(0.17)	(0.24)	(0.43)	(0.28)
SLP_{-ics} × Middle	-0.07	0.49**	0.08	0.01	0.05	0.14	0.49**	0.33	0.51**
	(0.13)	(0.21)	(0.12)	(0.14)	(0.23)	(0.14)	(0.20)	(0.27)	(0.25)
SLP_{-ics} × Top 20	-0.18	0.04	0.10	-0.30*	0.23	-0.00	0.02	0.23	-0.03
	(0.17)	(0.29)	(0.15)	(0.18)	(0.25)	(0.17)	(0.28)	(0.40)	(0.39)
$Peer\ \ PVT_{-ics}$							-0.01**	-0.00	-0.02**
							(0.01)	(0.01)	(0.01)
$Peer\ \ PVTSD_{-ics}$							0.01	0.00	0.01
							(0.01)	(0.01)	(0.01)
Edu non-response weights	NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ability Tracking Split	NA	NA	NA	NA	NA	NA	NA	Yes	No
Mean Dep Var	2.77	2.41	0.40	0.45	0.23	0.59	2.41	2.41	2.40
Observations	11074	7297	7309	7277	5183	7318	7297	4409	2771
R^2	0.20	0.28	0.25	0.22	0.26	0.24	0.32	0.31	0.33

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. Column (1) shows the effects of share of low-income peers on self-reported GPA from Wave I In-Home data while column (2) shows the effects on average GPA calculated from the first interviewed year to the end of the high school from Wave III high school transcript data. Columns (3) - (6) show the effects of share of low-income peers on the taking rate of advanced courses of Math, Science, English, and if ever took more than one advanced course. Columns (7) - (9) control for the distribution of peer ability where PVT is Picture Vocabulary Scores and SD is standard deviation. In columns (8) and (9), we stratify the sample by schools who report to use ability tracking for English and Language Arts. We use specific educational sampling weights constructed to adjust for transcript non-response as well as survey non-response in column (2) - (9). We trim our data to our analytic sample as in Table 2.2.

a curve would make these students appear to have higher grades. To check this, in columns (7) - (9) we compare students who face similar distributions of ability in their school-cohort – controlling for both the peer leave-one-out mean in PVT scores and its standard deviation – but who have variation in the share of low-income peers. The idea here is that they will on average face similar grade distributions thereby remaining effects from shifts in the share of low-income peers are unlikely to be due to such grade inflation.

The effect estimates for the share of low-income peers upon controlling for the peer ability distribution (column 7) remain essentially unchanged. We also see that an increase in the peer mean of ability (PVT scores) leads to weakly lower GPA (about 6.7 points lower for a standard deviation increase in peer ability). This negative effect on peer ability would be consistent with a grading on the curve mechanism. If so, then this effect on peer ability should disappear in schools which track by ability. In column (8), using schools who report to track by ability on English and Language Arts, we see this is indeed the case.[29] Moreover, the effects from the share of low income peers remain consistent across this stratification (columns 8 and 9). Overall, we see no evidence for a grading on a curve mechanism instead of an effort or motivation mechanism.[30]

2.6.2 Frustration and Motivation

Our model implies that shifts in income inequality can affect students through frustration and motivation. To proxy these, we use measures of self-esteem, relative intelligence (self) rating, depressive symptoms (the CES-D scale), and a scale we formed for motivation.[31] Details for these are reported in Table A.3 of the Appendix. We see self-esteem and depressive symptoms as particularly good proxies, because they relate to pessimistic beliefs on the returns to effort that once too low lead to withdrawal (De Quidt and Haushofer, 2019; Kiessling and Norris, 2022). This interpretation is consistent with exposure to income inequality generating frustration when inequality in opportunity is salient for those who are too far behind the opportunities of others. It is also consistent with generating competition for those with similar opportunities keeping effort higher.

In Table 2.4, we show that the effects from the share of low-income peers on these measures are non-linear across students' income groups. Students in the bottom 20th improve on self-esteem (column 1, significant) and self-perception of intelligence (column 2, weakly significant), while we continue to find null effects for middle-income students. Students in the top 20th see an increase in depressive symptoms (column 3, weakly significant) and a decrease in our measure of motivation (column 4, significant).

[29]The survey does not provide information about whether schools track by ability on other dimensions.

[30]We also explore outcomes on self-reported risky behaviors. The evidence on risky behaviors is inefficient, with a clearer suggestion of an increase in risky behavior for adolescents from higher income families as the share of low income peers increases, while we see null or negative effects on lower income adolescents. We describe these results in more detail in the Appendix Section A.5.3.

[31]Our motivation scale is an aggregate of two questions about how often the student has problems paying attention in school and getting their homework done. We scale these so that higher values imply less trouble. We recognize that this may also capture effort but it also can capture a degree of motivation.

Table 2.4. Frustration and motivation

	Self-Esteem	Intelligent Feeling	CES-D scale	Motivation
	(1)	(2)	(3)	(4)
SLP_{-ics} × Bottom 20	1.75**	0.34*	0.94	-0.10
	(0.85)	(0.20)	(1.44)	(0.17)
SLP_{-ics} × Middle	0.98	-0.01	0.74	-0.21
	(0.80)	(0.18)	(1.06)	(0.16)
SLP_{-ics} × Top 20	0.15	-0.00	3.11*	-0.52***
	(1.04)	(0.26)	(1.78)	(0.19)
Mean Dep Var	28.56	3.9	11.02	28.56
Observations	11134	11151	11154	11164
R^2	0.088	0.111	0.092	0.069

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the school level. School and cohort fixed effects are included in all specifications. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. *Self-esteem* is measured from 7-items that we base on discussions in (Rosenberg, 1989) and higher values imply better self-esteem. *Intelligent feeling* is a student's perception of their relative intelligence. The *CES-D scale* measures depressive symptoms where higher values imply worse mental health. Finally, *motivation* is composed of students' report on a 0-4 scale of how frequently they do not pay attention in school and a second 0-4 scale on not getting homework done. We reverse code these so that higher values imply they pay more attention and get homework done more frequently and then take the average of these two. Details of those variables can be found in the Appendix, Table A.3.

Interpreting in aggregate across all four measures in Table 2.4, the effects we observe here are consistent with the predictions of our model. Changes in income inequality generate non-linear patterns of frustration and motivation. Also, our model is unique in the non-linear predictions it makes. For instance, if changing income inequality only changed the structure of academic rank, then higher income students are likely to improve in rank as the share of low income peers increases. No mechanism in the literature on ranks (Elsner and Isphording, 2017; Kiessling and Norris, 2022; Murphy and Weinhardt, 2020) predicts worse outcomes among the top students.[32] Introducing social comparisons via reference-dependent preferences, as in our model here, brings this to focus and the empirical patterns are confirmatory.

[32] Also, as discussed earlier, we have controlled in several ways for rank effects and did not find them to explain our results for either the bottom or top income students.

2.7 Social Cohesion: Avoiding Harm from Inequality

In this last section, we ask what can be done to avoid harmful effects from exposure to inequality? Matching lower income students to better environments may in fact be desirable but only if it opens opportunities. Theoretical work on networks suggests that homophily can prevent the flow of information and opportunity across groups (for a review, see Jackson, 2021). For instance, a low-income student placed into a higher income school where the network is highly segmented by income groups, will be less likely to have network links with high income students and therefore not receive information nor experience any complementaries in effort. We view this as a low social cohesion environment consistent with an observed link between perceptions of school climate and network centrality (Alan et al., 2021b). Moreover, recent evidence shows that improving social cohesion improves student outcomes for both worse and better off students (Alan et al., 2021a).

In light of our model, better network integration could dampen the reference dependence mechanism that leads to our observed non-linear patterns. This could work through simply allowing a student to put less weight on their peers' income distribution to determine their reference point. It also may allow students who are unsure about the true abilities and opportunities of the peers to learn more and feel more involved and competitive. On the low-income side, students would then feel less frustrated and, on the high income side, less likely to lose motivation.

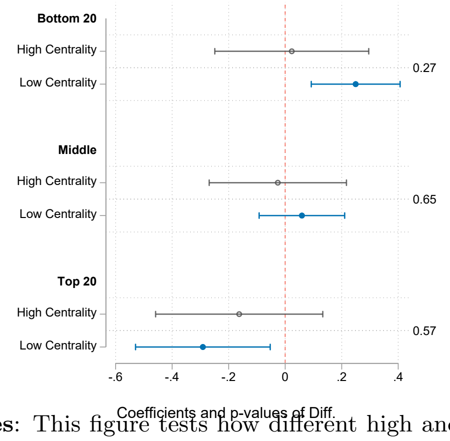
We test these implications by splitting our sample based on data from students' friendship nominations within the school. This is, of course, a descriptive exercise. It is beyond the scope of this paper to deal with the endogeneity of friendship nominations. Nonetheless, this is instructive for future work and points toward a hopeful direction. In Figure 2.5, we split the effects from the share of low income peers on university graduation based on having high versus low network centrality[33] and links outside of a student's own income group. We interpret having a high centrality and having a high number of links outside own income group as representing a high social cohesion

[33] We use Bonacich centrality, an index score that takes into account students' direct and indirect friendship links in the school (Bonacich, 1987).

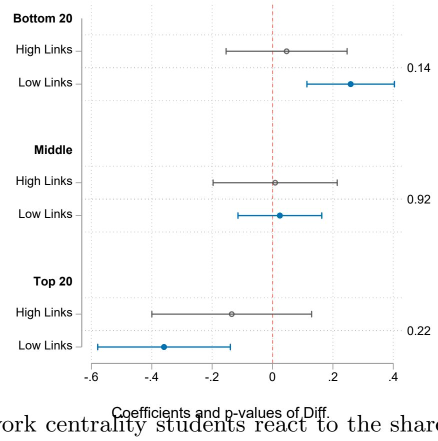
environment where information and opportunity is more likely to be shared.

Figure 2.5. University completion: heterogeneity by network centrality and school climate

(a) Sample split by centrality



(b) Sample split by links outside own income group



Notes: This figure tests how different high and low network centrality students react to the share of low-income peers where we split the sample by those above or below the median centrality in panel (a). In panel (b) we do the test over friendships nominated by the students outside their own income group. We always include school and cohort fixed effects as in column (2) of Table 2.2. P-values of differences are presented at the side.

Across both measures of social cohesion and integration, we observe a similar pattern. When a student has better network links, the effects from the share of low income peers are near zero and insignificant. However, when network links are poor, then the effects from the share of low income peers are large, significant, and aligned with our previous results for the bottom and top 20th income groups. We must be careful here because statistical efficiency does not allow making strong conclusions on the heterogeneity, but taken together, this descriptive evidence points to social cohesion as a moderator of the results we observe and potentially of harmful mechanisms from exposure to income inequality.

We finally turn to a more plausibly exogenous approach. Throughout this paper the peer reference group has been set at the same school-cohort level. We now compare this against more refined peer groupings where stronger friendship ties are likely to exist due to homophily. If social integration mitigates harm from peer inequality, then we should expect stronger peer inequality effects at the broader school-cohort level where exposure to income inequality signals more about the inequality in opportunity. Thus, we enrich our main specification and add a second share of low-income peers effects calculated (i) within school, cohort, and gender, (ii) within school, cohort, and race, or (iii) within school, cohort, gender, and race.[34] These results are reported in the Appendix, Figure A.8. In all cases, we find no effects on these more refined groupings, consistent with expectations based on more likely interactions in these groups, while our prior observed effects at the school-cohort level remain unchanged.

Our evidence throughout this paper contributes an important point to the literature, that income inequality may signal inequality of opportunity to students. This can then have adverse effects that our model rationalizes in theory and our evidence is consistent with. The results here on social cohesion then point toward a path forward that policy can take: attempting to expose students to different income backgrounds requires coupling this with efforts to improve social cohesion to avoid reference dependence from inequality in opportunity.

2.8 Conclusion

Exposure to income inequality among students may draw their attention to disparities in opportunity, in turn producing unintended consequences that are heterogeneous across the income distribution. Low-income students may realize they have fewer opportunities than their more fortunate peers, whereas high-income students can be motivated to do better if surrounded by students with similar opportunities. In this paper, we empirically examine the role of changes in peer income compositions on students' long-run educational attainment and their short-run performance, and how these change by own income.

We model the shift in school peer inequality using the share of low-income peers in a student's cohort within school. We then use this measure to examine how peer distributional shifts affect university completion and how these effects differ across the distribution of students own-household income. In order to identify these effects, we

[34] This is a "horse race", as we include both our baseline peer reference group definition and a more refined grouping in the same regression.

leverage within school, across cohort variation and flexibly control for students' household income. We also compare adolescents facing similar variances in the distribution of school-cohort income and additionally control for a rich vector of individual characteristics.

Our results show that low-income students benefit from an increase in the share of low-income peers, which positively affect their likelihood of university completion. Middle-income students experience on average null effects, and high-income students experience a reduced likelihood of university completion. These findings are robust to a rich battery of robustness checks. Our effects are sizable in magnitude: a 20% increase in the share of low-income peers raises the likelihood of completing university by 3.6pp for the bottom income students and decreases it by 4.1pp for the top income students. We also provide evidence that common mechanisms discussed in the peers literature do not explain our findings.

We then propose a novel theoretical framework that helps rationalize our results. We consider students with varying capacities for translating effort into educational outcomes, where capacity goes beyond the concept of raw ability and includes factors like opportunity and income. Students compare their outcomes to a reference point for educational attainment influenced by their peers' capacities. High-capacity (high-income) students perceive an increase in inequality when surrounded by a greater share of low-income peers, leading to lower motivation, effort, and attainment. Conversely, low-capacity (low-income) students see a reduction in inequality, resulting in higher motivation, effort, and attainment. Middle-capacity students may experience both situations, explaining an average null effect for this group. Hence, our model establishes that social comparison based on income can generate either frustration or motivation, depending on a student's relative position in the income distribution, and helps to understand potential unintended consequences of exposure to income inequality.

In further empirical analysis, we examine measures of performance, frustration, and motivation and find support for heterogeneous effects that are consistent with the theoretical predictions. Low-income students benefit from exposure to low-income peers in terms of short-term school performances, self-esteem and relative self-intelligent rating, whereas high-income students react with an increase in depressive symptoms and decreases in motivation.

Finally, we show descriptive evidence that the unintended effects of income inequality on students can be mitigated by social cohesion and a more integrated school environment. Social integration, measured through friendship nominations and cross-income group links, moderates the effects of low-income peers on university completion for both low- and high-income students. This suggests that policies fostering social cohesion can mitigate the consequences of exposure to peer group inequality. Overall, our evidence points to unintended consequences of exposure to peer income inequality that policy needs to take into account in order for students to benefit from this integration.

Chapter 3

Beliefs on Children's Human Capital Formation and Mothers at Work

3.1 Introduction

Gender gaps in labor markets, to the detriment of women, are well documented (Bertrand, 2011, 2020; Goldin, 2006), particularly harmful for new mothers (Kleven et al., 2019), and have remained stubborn in the face of a range of family policies to alleviate the cost for mothers to work (Kleven et al., 2023). One possible explanation put forward lies within gender norms via persistent beliefs about the dominant role of women in childcare that may constrain women's choices or affect their opportunity (Blau and Kahn, 2017; Cortés and Pan, 2023). This could mean that expectations about how well children will do are more pessimistic when a mother works longer hours relative to a father.[1] Yet, beliefs are difficult to gauge in standard surveys because different combinations of preferences and beliefs can be consistent with a given choice (Manski, 2004). Further, beliefs about how well children do when mothers work longer hours could be inaccurate opening a question on whether information about children's de-

[1] Even for mothers who are working these beliefs could affect not only a mother's decision to accept a more advanced and demanding role at work but also the decision to offer her the role.

velopment may shift beliefs and change support for policy that helps mothers go to work.

We address these questions in a novel survey design. We combine the elicitation of initial beliefs about children's skills and future human capital when mothers work longer hours relative to fathers with an information treatment about the actual relative development of children's skills when mothers work full-time. We advance four core contributions. *One*, we examine the beliefs people hold for children's future outcomes when mothers work longer hours relative to fathers. *Two*, we ask whether these beliefs vary over the population and what characterizes this variation. *Three*, we assess whether people respond to information about how well children do when mothers work full-time. And, *fourth*, we assess whether responses to information are homogeneous or polarized around initial beliefs. Answering these questions are particularly important to understand what drives variation in beliefs about mothers at work and to better understand how to message policies that aim to reduce the cost of mothers to work.

Sample and initial beliefs. We recruit participants for our pilot who are parents living in England through the online platform Prolific. The inclusion criteria and characteristics of our sample are discussed in Section 3.3.1. To elicit initial beliefs, we employ (i) hypothetical scenarios and (ii) incentivized beliefs about children's actual skill when mothers work full-time.

In the first step, we present participants with a hypothetical family of a mother and father with a primary school aged child. We then extract beliefs about child's future outcomes (graduation from university and earnings) across whether the mother or father works longer hours in the labor market while we hold constant the family's income. The design details are covered in Section 3.3.2. For each participant, we capture these beliefs at iterations of overall family income levels by varying the wage of the parent working the longer hours. This allows us to capture within-person differences in perceptions on the child's future outcomes when comparing a mother versus a father working longer hours at the same wage rate. It additionally allows us to capture how beliefs vary as these comparisons are made repeatedly along a hypothetical family income distribution. Importantly, we can estimate individual level average perceived returns from mothers working longer hours in the labor market that we will use as a measure of initial beliefs.

From the hypothetically extracted beliefs, we can summarize two key results. One, there is a belief that mothers who work longer hours relative to fathers are harmful to their child's future outcomes in terms of both the likelihood to graduate university and earnings. And, second, these gendered beliefs of harmful effects when mothers work longer hours only disappear as the hypothetical family income becomes very high. Thus, even when a mother would command the same wage as a father, there remains a strong gendered belief in terms of the return to children's skills. Furthermore, it appears that among our participants men and those who voted more conservatively in the last election drive more negative perceptions. Finally, we find a high degree of heterogeneity in these beliefs at the individual level that we use later to assess differences in responses to information. We believe that this pilot provides the first evidence of strong gendered beliefs from parental time to children's skills when maternal and paternal time are compared ruling out potential income effects due to wage differentials.

In the second step, we draw incentivized beliefs. We use data from the Millennium Cohort Study (MCS) which follows a cohort of children born in the year 2000 in the UK. From these data, we calculate a statistic that is well known to English parents as it is relevant for being admitted to college: the share of children passing five or more of their GCSE tests.[2] We compare dual parent families equalized by income and education levels. Details are covered in Section 3.4 and Appendix Section B.4. We present participants with the pass-rate for the share of five or more GCSE tests when mothers work part-time or fewer hours. We then ask them what they believe this pass-rate to be when mothers work full-time or more hours where we have compared families with similar income and education levels. We incentivize their answers by offering £1.5

[2] In England, a GCSE is a qualification in a specific subject typically taken by students aged between 14 and 16. GCSEs are below A-levels. If each University has its own admission criteria, they require, in most cases, at least a grade C/4 in Maths and English at the GCSE level, along with three or more GCSEs at the same grade or higher. In England, students are expected to take 9 subjects in GCSEs, among which 3 of them are compulsory – Maths, English and Science. Maths gives you 1 GCSE, English 2 GCSEs (English Language and Literature) and Combined Science is worth 2 GCSEs.

if it falls within plus/minus 2 percentage points (pp) of the actual pass-rate. We show in Subsection 3.4.2 that participants' beliefs extracted from the hypotheticals are indeed related to their incentivized belief about the GCSE pass-rate.

Finally, we look at these measures of initial beliefs and their links with self-reported gender norms. The results are too inefficient to draw conclusions at this time. However, we do see a consistently positive relationship between each of our initial beliefs and an index constructed from a set of questions on liberal versus more traditional gendered norms.[3]

Information treatment and effects. Next, we ask whether participants respond to information and whether this is heterogeneous to their initial beliefs. We randomly allocate participants to receive the actual information about the GCSE pass-rate when mothers work full-time or more hours. We then collect another incentivized belief. We inform participants that among children aged 7, and of the same gender, around 36 out of 100 have more behavioral problems than the median child in families where the mother works on average less than 35 hours per week. We then ask them what they expect this to be in families with similar education and income levels where the mother on average works 35 hours or more per week.[4] Again, we offer £1.5 for an answer that is within 2 percentage points of the actual statistics.

We also collect a range of additional outcomes. We ask participants a set of questions aimed to capture more liberal versus traditional gender norms and a set of questions on channels related to beliefs around parenting. These we motivate in Section 3.2 where we discuss possible paths creating heterogeneity in initial norms that if moveable may explain belief updating due to information. Finally, we design an obfuscated follow-up, inviting participants back one week later but obscuring any connection to the original survey.[5] In this survey, we ask participants a range of policy questions some on unrelated topics and some on topics related to support for policies to help mothers go to work.

[3] Since we can only look at these relationships in the information treatment control group, we will need more data beyond the pilot to make any stronger conclusions.

[4] See Subsection 3.4.1 for the specific wording given to participants.

[5] Approximately, 95% of our pilot sample participated in the follow-up.

We now summarize some key results on belief updating. In Subsection 3.4.3, we find that information about GCSE pass-rates when mothers work full-time yields significantly more positive and accurate beliefs about children's behavior. Furthermore, this appears to be driven by both those who under-estimated the GCSE pass-rate and those who hold more positive views about the likelihood a child will graduate university when a mother works longer hours relative to a father. It appears that participants who respond to the information are those who, while under-estimating the GCSE pass-rate when mothers work full-time, nevertheless do hold some positive views about women working. We cannot from our pilot results draw any strong conclusions, but these results suggest there is potentially important heterogeneity by initial beliefs and that a single measure of initial beliefs may not fully capture this heterogeneity.[6]

We also analyze the information treatment effect on self-reported gender norms but lack the variation to interpret these in a meaningful way. However, the scale of gender norms capturing more liberal versus more traditional views on women and mothers provides the best variation on these that we can get in the pilot. While the results on this scale are not significant, we see that the information treatment effect is split across our measure of initial beliefs about the likelihood to graduate university based on the hypothetical vignettes. The scale moves toward more liberal views among those who receive the treatment and have pre-existing more positive beliefs when mothers work longer hours. Yet, we see this scale move toward more traditional views among those receiving the information who have pre-existing more negative beliefs. This pattern is suggestive of a polarizing effect; however, we do not draw conclusions at this time as we do not find a similar pattern when looking at initial beliefs split by the expected earnings dimension.

Our evidence from the pilot is *suggestive* that responses to our information treatment may be heterogeneous to initial beliefs. Additionally, our evidence is quite clear that there is substantial heterogeneity in the initial beliefs we elicit. These suggest that indeed there are strong norms about the impact of mother's working on their

[6]It is important to note that treatment effects split by participants expected returns to children's future earnings when mothers work longer hours give less clear results than those split by expectations on children's likelihood to graduate university. See Table 3.6.

children's skill development consistent with suggestions in the literature that beliefs, or norms, may shape women's labor market behavior. Thus, whether beliefs respond to information and whether this is heterogeneous to initial beliefs is important for policy that aims to provide information about working mothers. Our pilot indicates that these initial beliefs are present and that information may lead to belief updating albeit heterogeneously.

We are not able to make strong or further conclusions at this time. Moreover, at the one-week follow-up, we see no link between our information treatment and support for policies to help mothers be in the labor market. Nevertheless, we expect that both on self-reported gender norms and policy views a much larger sample is required to obtain sufficient variation, which we will explore following adjustments to our pilot design.

Related literature. Our study relates to the literature on gender gaps in the labor market and their potential drivers. Gender gaps and the under-representation of women in the labor market are economically important as they are costly in terms of economic efficiency (Hsieh et al., 2019). The literature has highlighted that, despite substantial progress of women in terms of human capital investment, where women caught up and even surpassed men in many rich countries (Bertrand, 2020), gender earnings gaps are still persistent (Bertrand, 2020; Blau and Kahn, 2017). Recent studies examine the role of children in the gender earnings gap (Angelov et al., 2016; Cortés and Pan, 2023; Kleven et al., 2019). One common finding from these studies is that gender earnings gaps are mostly driven by gender differences in both the extensive and intensive margin of the labor supply. Women tend to reduce their working time after the birth of the first child where we see the gap begins to open. We contribute to this literature by examining the role of gendered beliefs in returns to parental time investment with children that ultimately can affect specialization within the couple, and labor supply decisions.

Our study is also related to the literature examining gender differences in preferences for job attributes as another potential driver of gender earning gaps. Job amenities are an important factor that women (but not men) consider in the decision making about their jobs (Hotz et al., 2018; Wasserman, 2022). Women relative to men are typically observed sorting into less demanding jobs in terms of working time, with a larger share of female co-workers, part-time workers, and female co-workers with young children, a pattern confirmed by experimental studies (Maestas et al., 2023; Wiswall and Zafar, 2017). Our study speaks to this literature insofar as the beliefs that mothers working longer hours relative to fathers can be harmful for children's development leads women to sort into more flexible jobs requiring shorter hours.

We further contribute to the growing literature on parental time investment and parental beliefs about returns to parental time in terms of children's skill development and future success in the labor market (Attanasio et al., 2020; Boneva et al., 2022; Boneva and Rauh, 2018; Kiessling, 2021). Parental time with children is increasing in many countries (Aguiar and Hurst, 2007; Borra and Sevilla, 2019), due partially to increasing returns to education and competition in the education market (Ramey and Ramey, 2009). One recent study examines beliefs about effects of mothers' decision to work on children's skill development (Boneva et al., 2022). They find that beliefs on children's skills and family outcomes are improving for mothers moving from no work to part-time work, effects that are partially driven by increases in income, but declining relative to part-time work when moving into full-time work. Our paper sets out to explore a related though different mechanism, by focusing on beliefs about the comparison between mothers working longer versus shorter hours *relative* to fathers, abstracting from mechanisms operating via income effects. And, we turn attention to the updating of beliefs based on information provision heterogeneous to initial beliefs.

Finally, we also relate to the literature on gender norms and the role that they can play in constraining women's behavior and shaping individual preferences for work (Blau and Kahn, 2017; Cortés and Pan, 2023). Recent evidence from Norway (Andresen and Nix, 2022) shows that child penalties differ substantially between women in heterosexual couples and same-sex couples, suggesting gender norms potentially play an important role. Moreover, these beliefs can be incorrect, as evidence from a US sample shows that people under-estimate the progressiveness of peers in their state about women working but update their beliefs in response to information about their peers' beliefs (Cortés et al., 2022). Thus far, beliefs about women working appear substantially heterogeneous but to some degree malleable. We turn attention to understanding the strength of these beliefs, what characterizes them, and whether responses to information are homogeneous or polarizing.

3.2 Skills Development and Belief Distributions

Skill development. Suppose a simple model of a household with one child. The parents jointly maximize utility over consumption, leisure, and the child's uncertain future human capital (or skill) accumulation. The child attains future skills based on initial skills (s_0), purchased investments (I_0^{pu}), and time investments from the mother (I_0^M) and from the father (I_0^F). The production function for skill accumulation at adulthood then follows:

$$S_1 = f_0(s_0, I_0^{pu}, I_0^M, I_0^F). \tag{3.1}$$

We make standard assumptions on the technology of skill formation $f(\cdot)$, assuming it is continuous, monotonically increasing, and concave in the inputs. We do not assume, however, that people necessarily hold accurate beliefs on the technology of skill formation. Specifically, we outline three mechanisms that could vary beliefs heterogeneously and in turn be important for gendered differences in the selection of time investments and hours worked in the labor market.

Our empirical interest is on how people perceive the returns to children's skill accumulation to differ across whether a mother or father work longer hours in the labor market when we hold constant the budget constraint. We can summarize some relevant points to consider even when both the mother and father command the same wage offer.

1. *Beliefs on marginal productivity.* Differential beliefs on the marginal productivity of mother's and father's time investments can form incentives for heterogeneous sorting on time investments into the child and hours worked in the labor market. The presumed values of the productivity parameters need not match reality, and there can be believed differences across gender leading to differential beliefs on

$\frac{\partial f_0}{\partial I_0^M}$ versus $\frac{\partial f_0}{\partial I_0^F}$. Also, wider societal beliefs on these productivity differences may act to put further pressure on the partner presumed to have the higher productivity to sort out of the labor market.

2. *Beliefs on differences in preferences.* Differential beliefs on the preferences for leisure between mothers and fathers, or alternatively differences in the weight each places on the child's future skills in the utility function, can lead to differential sorting on hours worked through different choices on the level of time investments. People may form beliefs about their partner's preferences and these can then impact their own choices such that, for example, let ϕ_F^M represent the belief the mother forms about the father's preference for leisure. If $\frac{\partial I_0^M}{\partial \phi_F^M} > 0$, then the mother increases her time investment in compensation as the belief on the father's leisure preference increases. Wider societal expectations about men's preferences could lead to beliefs that women working longer hours will harm children through less overall time investments.
3. *Beliefs on resource allocation.* When the partner who works longer hours, and earns a greater share of the household budget, has greater control of allocating household resources, then different preferences between the mother and father for consumption versus purchased investments (I_0^{pu}) imply a different allocation of resources depending on who works longer hours holding constant the overall budget. Beliefs on who gains greater control over the household budget can then imply different beliefs on children's skill production.

Belief distributions. We aim to empirically investigate the belief differences people hold based on whether a mother or father works longer hours in the labor market.[7] We define a belief distribution on the expected future outcomes of a child (H_{t+1}) – reflecting human capital accumulation which we assume directly reflect S_{t+1} – when a mother works the longer hours ($MWL = 1$) to be $B_i(H_{t+1}|k, MWL = 1)$, where we index this over individuals (i) and wage offers (k). Likewise, $B_i(H_{t+1}|k, MWL = 0)$ represents the belief distribution when a father works longer hours. The belief distributions we

[7] Note in our design this belief is not necessarily only for the individual but the belief each individual holds in general when a mother or father works longer hours.

propose share some similarities to the belief distributions that Wiswall and Zafar (2021) consider when they explore beliefs that college students hold about future returns to different choices of college majors in that our distributions (i) reflect uncertainty, (ii) can vary over individuals, and (iii) can be wrong, e.g., not reflect actual outcomes.[8]

Through collecting expectations on children's future outcomes in hypothetical scenarios we will investigate:

$$\Delta_{B,i,k}(MWL) = B_i(H_{t+1}|k, MWL = 1) - B_i(H_{t+1}|k, MWL = 0). \tag{3.2}$$

We allow for differences in these belief distributions to be heterogeneous over individuals (i) and wage offers (k). These ex ante beliefs are important for determining labor market choices between mothers and fathers and our previous summary points have relevance now to expectations on the sign of $\Delta_{B,i,k}(MWL)$. We consider this from the likely direction of beliefs under what may be thought of as traditional gender norms. From (1), if people hold different beliefs on the marginal productivity of time investments across mothers and fathers such that mothers are expected to be more productive, then this would press $\Delta_{B,i,k}(MWL)$ downward. From (2), if people expect fathers to have different preferences where consequentially they put in less time investments, then again this presses $\Delta_{B,i,k}(MWL)$ downward. Finally, from (3), if people believe women have stronger preferences for purchased investments than do men, and they also believe that when a woman earns a higher share of the household budget she has greater allocative control, then this would press $\Delta_{B,i,k}(MWL)$ upward.

We could draw different hypotheses based on different expectations about points (1), (2), and (3). Nevertheless, these are instructive, because even when gender norms form beliefs that put pressure on women to select fewer hours worked, point (3) could confound our ability to capture these beliefs as it can work in the opposite direction. Therefore, in our survey design we introduce information aimed at shutting down this channel. An empirical estimate of the average over N individuals and K wage iterations, expressed as $(\sum_{i=1}^{N}\sum_{k=1}^{K} \Delta_{B,i,k}(MWL))$, removing point (3) will provide evidence

[8]Further, for simplicity, we do not index these beliefs by time, but we can easily allow for learning by allowing multiple periods where people can update beliefs.

on gendered beliefs but not on whether the important channel is either from beliefs over marginal productivity (point 1) or differences in preferences for time investments (point 2). Thus, in our survey we introduce questions on these two points to investigate mechanisms.

3.3 Hypothetical Belief Elicitation: Design and Results

3.3.1 Sample

We conducted our experiment on the online platform Prolific, and considered 2 main inclusion criteria. We required participants to be (i) parents of at least one child aged 18 or below, and (ii) currently residing in England.[9] We then drew 249 participants meeting those criteria, among which 133 (47%) receive the information treatment (see Subsection 3.4.1 for further details). Table B.1 in the Appendix Section B.1 provides summary statistics on our pilot respondents, as well as a balance check for treated and control participants. Overall, our sample demographics indicate a balanced sample, where the differences in characteristics between treated participants and the control group are mainly non-significant. The only significant differences worth mentioning are about the ages of the eldest and youngest child(ren). Indeed, the control group has overall younger children than treated participants.

3.3.2 Hypothetical Design

Framing. We use hypothetical scenarios in vignettes to elicit people's beliefs on the human capital accumulation of children in response to women vs. men working longer hours in the labor market. The following is the text participants see to set the stage for the scenarios:

> We are interested in your opinion on children's future outcomes, comparing families with different financial resources and time demands.

[9] We focus on England only, as we will focus, later in the survey, on a metrics (the GCSE pass rates, see Subsection 3.4.1) mainly known in England. In other countries of the United Kingdom, the name and content of the GCSEs, as well as the exam requirements may be different.

> **Setup:** For this purpose, we would like to ask you to imagine an average family in your community. Suppose this family consists of a father and a mother who are both employed, and they have a boy (girl, *randomized*) who is aged 10 (4, *randomized*). Suppose household expenditure decisions are made jointly by the father and the mother, and this hypothetical family spends 10% (20%, *randomized*) of their total income on the child's educational activities such as clubs, tutoring, music, sports, etc. Assume that there is no inflation (i.e. prices do not increase).
>
> More specifically, we will show you different scenarios, and ask your opinion about the likelihood that the child will be successful in education and the labor market. There are no clear right or wrong answers, and we know these questions are difficult. Please try to consider each scenario carefully and tell us what you believe the likely outcomes will be.

Randomization in the setup. We randomize several features in the setup, which are also presented in Table B.2 in the Appendix Section B.1. These are whether the participant reads that the family has a boy or a girl, the age of the child, and the share of income allocated by the family to the child's educational activities. The last one relates to our discussion in Section 3.2 on the potential for people to hold different expectations about the household allocation of resources to purchased investments when the mother versus the father contributes the most to the family budget. We aim to hold this allocation constant since it may divert attention from the mechanisms we are interested in eliciting. We randomize the level of this share that different participants see in order to provide a check that participants actually pay attention to this part of the setup.

Scenarios and outcomes. Next, for each participant we iterate through a set of scenarios (6 total) varying two components through 2 pages — presenting 3 scenarios on each page: (i) whether the father or mother works longer hours and (ii) the wage rate of the parent who works the longer hours. An example scenario is as follows:

The father works 36 hours per week at a wage rate of £25 per hour.

The mother works 15 hours per week at a wage rate of £11 per hour.

We then ask each participant for what probability they believe the child will eventually graduate from university (0-100 scale answered with a slider) and what they expect the earnings of the child to be at age 30 (textbox entry). We iterate on the scenarios and re-collect the expectations/beliefs. Example images of what the participants see here are presented in the Appendix Section B.1. Finally, we convert the outcomes to a 0 to 1 scaled probability of graduating from university and the log of expected earnings at age 30.

Randomization in the scenarios. Table 3.1 contains the design on how we iterate through scenarios. Participants work through 2 pages each containing 3 scenarios. To avoid order effects, we randomize whether each participant starts with the man or woman working longer hours. We also randomly draw the *kth* iteration of wages shown within each page so that participants do not move sequentially through wage changes. In all cases, we hold constant the wage of the parent working fewer hours. We further randomize whether the wage profile of the one working longer hours has a lower bound of either £11 or £15 and an upper bound of either £25 or £29. This allows across participants for the overall wage profile to range from £11 to £29.

Table 3.1. Design

	Man Works More		Woman Works More	
	w_m	w_f	w_m	w_f
$k = 1$	15£(11£)	15£(11£)	15£(11£)	15£(11£)
$k = 2$	22£(18£)	15£(11£)	15£(11£)	22£(18£)
$k = 3$	29£(25£)	15£(11£)	15£(11£)	29£(25£)

Note: The man works more scenario corresponds to 42 (36) hours vs. the woman with 20 (15) hours. The woman works more scenario repeats this switching the longer hours to the woman and the shorter hours to the man. w_m is the man's wage rate and w_f is the female's wage rate.

Finally, the number of hours worked is randomized across participants. Some in the "works more" category see 36 hours and others see 42 hours, while for the partner working part time some see 15 hours and others see 20 hours. We will use this later for heterogeneity. Additionally, further technical details on the operation of the survey are in the Appendix Section B.1.

Checks and confidence. We follow Haaland et al. (2023) to test participant attention to the survey and confidence in their answers. First, upon completing the hypothetical scenarios we provide participants with a paragraph of text wherein we ask them to report that their favourite colour is "turquoise" in the textbook below. Then at the textbook we simply ask "what is your favourite colour?." Later in the survey we use a second attention check that we will discuss in Section 3.4. Second, we ask participants to what extent they are sure about their answers to the hypothetical scenarios.[10]

We report summary statistics on both our attention and confidence checks, in Table B.4, in the Appendix, Section B.1. At least 99% of our participants passed the two attention checks, suggesting strong attention to our survey. On confidence, 63% of our participants report being at least "somewhat sure" of their answers to the hypothetical scenarios. We perform robustness checks later using these screeners, to highlight the reliability of our estimates.

Additionally, we look at the share spent on educational (SSE_i) activities (e.g., clubs, tutoring, music, sport, etc.) of the hypothetical family's total income and its association with our two main expected outcomes: 1) the expected probability of the child's graduation, and 2) the expected earnings at age 30 of the child. We include this randomized dimension (SSE_i) because it can constitute a potential channel where parents believe that, when mothers work longer hours, they allocate more resources to the child's education. Therefore, to investigate the relationship between SSE_i and our expected outcomes, we first run an OLS regression of each outcome on indicator for SSE_i, allowing random effects and controlling for the hypothetical child's characteristics, as well as the hypothetical monthly household income.[11] Second, we run

[10]In the Appendix Figures B.1 and B.2, we provide screenshots of the attention check and confidence questions that participants actually see. For the additional attention check, see Figure B.3 in the Appendix.

[11]In order to obtain the hypothetical monthly household income, we calculated the individual's

another OLS regression, as before, but by interacting the working profile of the mother (MWL) with the SSE_i. The results are displayed in Table B.3 in the Appendix Section B.1. They highlight that there seems to be no association between the SSE_i and our two main outcomes. Thus, at this time, we cannot rule out that beliefs on resource allocation matter in the beliefs we elicit.

3.3.3 Results: Hypothetical Beliefs Elicitation

We now address our first set of aims to elicit beliefs on children's future outcomes based on whether a mother versus a father in a family works longer hours. We do this around two main empirical assessments.

Empirical strategy on gendered beliefs. Our design collects expectations as the household budget constraint rises in exactly the same way when the mother or father works longer hours. If there are no gendered beliefs, then we expect that the average change in expectations as the household budget rises will be the same regardless of who works the longer hours.

Empirically, we want to aggregate the difference in beliefs defined in equation 3.2 as $\Delta_{B,i,k}(MWL)$ across the k wage iterations within individuals and across individuals. We define the average change in beliefs on the child's future outcomes as the mother's wages rise to be $\Delta_{B,M}$ and the average change in beliefs as the father's wages rise to be $\Delta_{B,F}$. When there are no gendered beliefs, we expect that $\delta = \Delta_{B,M} - \Delta_{B,F} = 0$. Because we can control for the effect from changes in the household budget, then an estimated $\hat{\delta} \neq 0$ will be consistent with gendered beliefs.

We estimate this for each expectation (e) that we collect based on our design with the following specification:

$$y^e_{i,j,k} = \alpha_0 + \delta MWL_j + \tau_k + \mu_i + \epsilon_{i,j,k}. \tag{3.3}$$

A vector of participant fixed effects is captured by μ_i, representing our preferred

(mother's and father's) income by multiplying the hourly wage rates and the number of hours worked per week, which we multiplied by four to be displayed monthly. We then summed individual incomes.

specification. In some specifications, we replace these with a vector of participant characteristics which are age, gender, and an indicator for whether they have at least a degree. Lastly, τ_k refers to household income fixed effects, which corresponds to the hypothetical monthly household income. As previously explained, we retrieved this information by multiplying the hourly wage rates and the number of hours worked per week, which we multiplied by four to be displayed monthly. We then summed individual incomes.

Thus, we will use variation within participants and across whether the mother works longer hours ($MWL_{j=1}$) or the father works longer hours ($MWL_{j=0}$) holding constant the effect from the change in the household income. An estimate of δ, then captures average differential beliefs based on whether the mother or father works longer hours. A $\hat{\delta} < 0$ will be consistent with gendered beliefs that suppose it is more harmful for women to work longer hours than men for children's human capital accumulation.

To investigate heterogeneity in gendered beliefs, we further estimate various versions of equation 3.3. First, we investigate heterogeneity in this effect based on hypothetical's family income, by interacting MWL_j with hypothetical household income tertiles. Second, we investigate heterogeneity in gendered beliefs by hypothetical features (e.g., child's gender and age, etc.). Lastly, we investigate heterogeneity by participant's characteristics.

Average estimates of gendered beliefs. Results of OLS regressions for equation 3.3 and for each outcome are presented in Table 3.2 below.

Result 1. *There is a belief that mothers who work longer hours are harmful to their child's future outcomes.*

Our estimates of $\hat{\delta}$ with individual fixed effects return both significant and negative effects for mothers working longer hours versus fathers on both of the hypothetical child's outcomes.[12] Specifically, our estimates indicate that when women work longer hours (compared to men working longer hours), participants on average reduce their

[12]Our estimates of $\hat{\delta}$ seem to be more efficient when we use individual fixed effects (columns 2 and 4) than when we simply control by participants' characteristics. The magnitude of the coefficients is the same for both specifications

Table 3.2. OLS Results – Gendered Beliefs

	(1) $\mathbb{P}$(graduate)	(2) $\mathbb{P}$(graduate)	(3) ln(earnings)	(4) ln(earnings)
$\text{MWL}_{\text{j=1}}$	-0.014***	-0.014***	-0.016**	-0.016*
	(0.005)	(0.005)	(0.008)	(0.009)
Mean Dep. Var	.566004	.566004	3.4526	3.4526
Individuals	249	249	249	249
Observations	1494	1494	1494	1494
Indiviual Controls	Yes	No	Yes	No
Individual Fixed Effects	No	Yes	No	Yes
Household Income Fixed Effects	Yes	Yes	Yes	Yes

Notes: * p <0.10; ** p <0.05, *** p <0.01. Standard errors are clustered on individuals. Individual controls include participant's age, gender, a dummy variable for having a degree or less, and a dummy for ethnicity (white vs. non-white). For the ease of interpretation, the probability to graduate has been recoded to be ranging from 0 to 1, i.e., we divided the original variable by 100.

expected probability of the child graduating University by 1.4 percent. For future earnings at age 30, they reduce their expectation by an average of 1.6 percent.

Heterogeneity by hypothetical design features. Here, we investigate heterogeneous effects of mothers working longer hours than fathers, by features of the hypothetical setup. We focus first on the role of household income in the scenarios in Table 3.3.[13]

At low and medium income categories, beliefs about the child's likelihood of graduating university are 1.7 percent (lower income) and 1.9 percent (medium income) lower for scenarios when mothers work longer hours than fathers. However, these beliefs are essentially identical at high income. For expected income, the point estimates again only drop to near zero for the high income scenarios, albeit our results here are only significant at the low income scenarios.

Overall, gendered beliefs do disappear at scenarios with high income families but

[13]Due to sample issues, here, we interact MWL_j with income tertiles, deduced from the hypotheticals. The low, medium and income categories encompass, respectively, 604 (40.43%), 498 (33.33%), and 392 (26.35%) observations. The average hypothetical annual household income for the low income group is £36,135. For the medium income group, the average is £51,171, and for the high income group, the average is £64,524. We acknowledge that these are quite generous bands. In particular, 40% of our sample earns £36,135 in the hypothetical scenarios while recent ONS estimates indicate a median household disposable income, for 2022, of £31,883.

Table 3.3. OLS Results of $\hat{\delta}$ – Gendered Beliefs and Hypothetical's Household Income

	(1) $\mathbb{P}$(graduate)	(2) ln(earnings)
Low income × $MWL_{j=1}$	-0.017**	-0.024*
	(0.007)	(0.014)
Medium income × $MWL_{j=1}$	-0.019***	-0.016
	(0.007)	(0.012)
High income × $MWL_{j=1}$	-0.003	-0.006
	(0.008)	(0.011)
Individuals	249	249
Observations	1494	1494
Individual Fixed Effects	Yes	Yes
Household Income Fixed Effects	Yes	Yes

Notes: * p <0.10; ** p <0.05, *** p <0.01. Standard errors are clustered on individuals. The low, medium and high categories are quantiles of the hypothetical household income.

these beliefs are persistent across a large range of potential family incomes. A possible explanation is that people hold heterogeneous beliefs and assume that with high enough income any negative channels from mothers working longer hours will be outweighed by income effects. Our results suggest this only occurs once income is quite high, suggesting these beliefs are fairly sticky.

Additionally, in Table B.5 in the Appendix, Section B.2, we report heterogeneity estimates by the remaining randomized hypothetical design features. These include the hypothetical child's gender (boy or girl) and age (4 or 10), whether participants see 36 or 42 hours for full time work and whether participants see an overall lower wage profile (starts lower and finishes lower) or higher wage profile (starts higher and finishes higher). We estimate equation 3.3, which we further stratify by the above-mentioned dimensions. For both outcomes, our estimates are mostly homogeneous across these features with the exception of across the child's gender. Here we find that mothers who work longer hours relative to fathers are believed to be particularly harmful for boys rather than for girls.

We can now summarize the key results for how beliefs vary over features of the hypothetical design.

Result 2. *Beliefs of harmful effects from a mother working longer hours only disappear as the hypothetical family income becomes very high and are the strongest when the child is a boy.*

Heterogeneity by participant's characteristics. Heterogeneous results by participant's characteristics are reported in Table B.6 in the Appendix Section B.2. We stratify the sample by participant characteristics and estimate $\hat{\delta}$. Our evidence here is relatively mixed.

We observe some evidence that men drive the gendered beliefs we mentioned in Result 1, especially concerning the expected probability of graduating from University[14]. Next, by participants education level and employment status we see some heterogeneity but no clear pattern across outcomes and significant results in places across all categories. Thus, we omit further interpretation at this time.

Finally, we split the results by participants' voting behavior in the last election. Those who voted for the conservative party and those who did not vote for any prelisted party hold stronger beliefs of harm for both of the future outcomes when mothers work the longer hours.

Robustness checks. In the Appendix, Section B.2, Table B.7, we test the robustness of our key finding (Result 1) by implementing different sample restrictions. First, we asked participants how confident they were in their answers to the hypothetical scenarios (cf. Figure B.2). Here we exclude participants who reported being unsure or very unsure about their answers – the results for those at least somewhat certain are displayed in column 1. Second, we exclude participants who did not pass at least one, out of the two, attention checks (see Figures B.1 and B.3). The results are reported in column 2. Third, and finally, despite our attention checks respondents may either give minimal attention and swiftly navigate through the survey or engage in activities

[14]Earlier we observed that scenario setups with "boys" are are where we observe the strongest beliefs. One concern for our heterogeneous result by participant gender is that men could have by random chance been allocated more (than women) "boy" setups as the hypothetical child's gender. We cross-tabulate participant's gender and the hypothetical child's gender to dismiss this concern and find that 49.17% of men got assigned "boys", while 47.62% of women got assigned "boys". The gendered belief we observe is therefore not due to an unbalanced distribution of our randomized design features.

unrelated to answering the survey, which would provide unreliable estimates of our gendered beliefs. Therefore, we exclude participants with the 5% lowest and highest response times.[15]

Our main result is robust to the checks we implement. The coefficients associated with $MWL_{j=1}$, in every column of Table B.7, are about the same magnitude as the ones we find in Table 3.2 without sample restrictions.

3.3.4 Individual Perceived Returns

Approach. We estimate individual level average perceived returns to mothers working longer hours versus men. To do this, we estimate equation 3.3 for each person in the sample dropping the household income fixed effects and individual fixed effects. This recovers each respondent's average gap between scenarios with mothers versus fathers working longer hours. We do this for each expectation outcome labelling the perceived returns to mothers working longer hours compared to fathers as $\theta_{graduate,i}$ (graduating university) and $\theta_{earnings,i}$. We winsorize the resulting returns at the 1 and 99% levels to account for outliers.[16] We then explore how these relate to each other (consistency check) and to individual respondent characteristics. Later, we will use these to assess how they relate to our incentivized beliefs and survey outcomes along with whether our information treatment is heterogeneous to these hypothetically elicited beliefs that we recover here.

Perceived returns consistency. We check the consistency of individual perceived returns over the probability of graduating University and earnings by age 30. We report in the appendix, Figure B.5, a strong and positive relationship between the two individual-level perceived returns significant at the $p < 0.01$ level. Reassuring for our empirical approach, both of our expectation outcomes garner comparable beliefs with some variation.

[15] Note here that, due to technical issues, we could not retrieve response times for 17 participants out of the original 249. After applying the sample restriction excluding participants with the 5% lowest and highest response times, we have 208 participants.

[16] See Figure B.4 in the Appendix for the histograms of both θs, with kernel density plot.

Individual characteristics and perceived returns. Finally, in Table B.8 of the Appendix, we check how these perceived returns relate to participant characteristics, by regressing both dimensions of θs on a set of individual characteristics. We lack the sample size to detect significant variation and only point out a few suggestive patterns. First, being older, above the median age, seems to be associated with a reduction in both perceived returns albeit stronger for earnings. Second, being from another ethnic background than white is also associated with a reduction in both perceived returns. Third, voting for a liberal party at the last UK General Elections, compared to voting for the Conservative party, is associated with an increase (more positive belief about women working longer hours) in both perceived returns.

We further investigate the relationship between individual's perceived returns and their actual behavior with their child(ren), and on the labor market. In particular, we focus on the the time they spend with their child(ren) to 1) develop their skills, 2) doing outdoors activities, and the log of their weekly hours worked. Results are provided in Appendix Table B.9 – for all participants (Panel A) and for the control group (Panel B). Additionally, we produce the same results by gender in Table B.10. However, we are again underpowered to detect relationships leaving very noisy estimates with no clear pattern in the pilot.

3.4 Information Experiment: Design and Results

So far we observe that women versus men working longer hours in the labor market is believed to be harmful for children's future outcomes. We now want to investigate the following set of objectives.

First, we assess whether the hypothetically extracted beliefs predict an incentivized belief about children's skill. We further assess how self-reported beliefs on gender norms and channels related to our discussion in Section 3.2 associate with both the hypothetically extracted beliefs and the initial incentivized belief about child skill.

Second, we ask whether beliefs are moveable in response to information about children's skill among families with similar education and income levels when mothers work full-time hours compared to less than full-time. We also test whether the effect of information is heterogeneous to the prior beliefs both from the hypotheticals and the incentivized initial belief. Thus, we ask are responses to information dependent on prior beliefs?

Third, and finally, we assess policy views related to promoting mother's labor market opportunities and the link between these and our initial beliefs as well as the impact of information.

3.4.1 Treatment Design and Outcomes

We use the Millennium Cohort Study (MCS) to draw some statistics on child development. The MCS is a longitudinal study following a nationally representative sample of families and children born in the year 2000. We calculate the share of children passing five or more of their secondary school GCSE's with a C/4 or higher.

This pass rate is a common metric in school league tables in England, which will likely be familiar to our sample of parents living in England.[17] We split this metric by families where the mother on average worked 35 hours or more when the child was aged 5 and 7[18] versus families where the mother worked less than 35 hours or did not work at all. We draw our calculations from dual-parent homes in England and compare families where parents have similar income and education levels. We provide more details on the data and our calculations in the Appendix, Subsection B.4.1.

Prior belief. We first inform participants of this pass rate for families where the mothers worked fewer hours[19]. We then collect their incentivized belief about this pass rate for families where the mother works longer hours. Below is the text participants read.

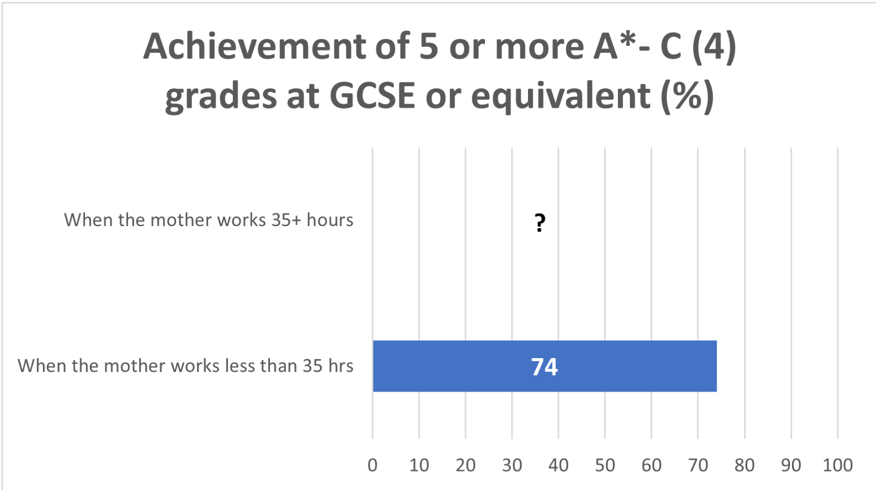
> We, as researchers at the University of Strathclyde, have calculated the

[17] See the UK Government website for further information about GCSEs results in 2023. Also see the subject content of GCSEs, by field.

[18] We use age 5 and 7 sweeps (respectively years 2006 and 2008) because we want to retrieve mother's working hours at primary school age of the child.

[19] We present unweighted versions but also provide in the Appendix, in Figure B.6, the weighted version, which are nearly identical.

share of children passing five or more of their GCSEs with a C/4 or higher. For dual-parent homes in England who have similar income and education levels, we obtain the following statistics.



This graph indicates that among families where the mother worked less than 35 hours per week, around **74%** of children passed five or more GCSEs with a C/4 or higher.

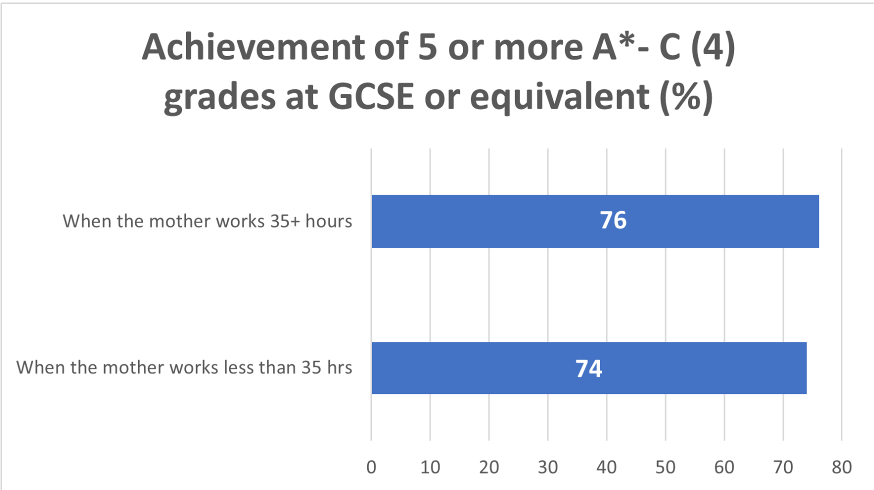
Among families with similar income and education levels but where the mother worked 35 hours or more per week, what percentage of children do you believe eventually passed five or more GCSEs with a C/4 or higher?

You will gain £1.50 if your answer is within 2 percentage points of what was actually found.

Participants respond by dragging a slider between 0 and 100% in increments of 1 percentage points.

Information treatment. Next, we randomize participants to either receive the actual pass rate when mothers work longer hours or to the control group with no information. Those who are assigned to the treatment are shown the statistics graphically, (see Figure below[20].) and told in text the following:

> Making comparisons among families with similar income and education levels and where the mother worked 35 hours or more per week while the child was aged 5 and 7, we found that around **76**% eventually passed five or more of their GCSE's with a C/4 or higher.
>
> Compared to families with similar income and education levels but where the mother worked less than 35 hours per week this means these children did on average about **2 percentage points** better.



Outcomes. We follow the information treatment by collecting a set of outcomes.[21] First, we collect another incentivized belief this time focusing on behavioral problems when mothers work part-time or less compared to full-time or more when the parents have similar education levels and income. Relative to our question on GCSE pass-rates we use a different response scale and change the direction of the scale that implies better perceptions to mitigate concerns over numerical anchoring. Below is what we show and ask of participants.

[20]For further details on the information treatment statistics (notably the weighted versions), see Figure B.6 in the Appendix Section B.4

[21]Refer to the Appendix, Subsection B.4.4 for descriptive statistics and more details about our final outcomes.

> The data that we used to calculate the share of children passing five or more GCSEs also provides information on the children's externalising behavioural problems at age 7 (e.g. conduct problems and hyperactivity/inattention).
>
> Comparing families with similar income and education levels in England, we calculate, by gender, the percentage of children who had more behavioural problems than the national average. We further split our statistics according to the mother's number of hours worked per week while the child was aged between 5 and 7.
>
> Among families where the mother worked on average less than 35 hours per week during this period, **out of 100 children** aged 7 and of the same gender, **around 36** of them had more behavioural problems than the median child.*
>
> Among families where the mother worked 35 hours or more per week, out of 100 children, how many of them do you believe had more behavioural problems within their own-gender than the middle (median) child?
>
> You will gain £1.50 if your answer is within 2 values of what was actually found.
>
> * The median child corresponds to the child in the middle of the distribution

Participants are asked to report their expectation in a textbox.[22]

Second, we collect self-reported gender norms and channels related to our discussion in Section 3.2 that may relate to beliefs about mothers working longer hours. We ask a set of questions to explore (i) beliefs about marginal productivity across mothers and fathers for investments in a child's skills (Q1 of Table 3.4); (ii) beliefs about resource allocation (Q2 and Q3 of Table 3.4); and (iii) beliefs about differences in preferences for child investments (Q4 and Q5 of Table 3.4).

Participants are asked the following questions (Table 3.4). For each of the first three questions, we define a variable to be equal to 1 if the participant replied "mother" as

[22]The acutal share of children with more behavioral problems than the median for when mothers work 35 hours or more per week making comparisons among families with similar income and education levels is 36.58 out of 100.

being the more efficient in the dimension considered, 0 if they either replied "father" or "both equally". For the last two questions, we simply use reported expected hours as the outcome.

Table 3.4. Beliefs on Channels Related to Gender Norms

	Answer modalities
Q1 – Productivity: In a family where the mother and father have the same education level, in time spent on helping a child with educational activities, if only one parent can be involved, which parent do you believe would be the most effective?	
Q2 – Resource Allocation: In a family where the mother and father have the same education level, in money spent on helping a child with educational activities, if only one parent is allowed to make the resource (money) allocation decisions, which parent do you believe would allocate more money to the child?	1. Mother 2. Father 3. Both equally
Q3 – Resource Allocation: In a family where the mother and father have the same income level, who do you believe would be more likely to make the resource (money) allocation decisions?	
Q4 – Preferences: In a typical family where the father works full time, how many hours per day on average do you think the father spends helping their children develop educational and social skills?	Drop down menu ranging 1 to 24
Q5 – Preferences: In a typical family where the mother works full time, how many hours per day on average do you think the mother spends helping their children develop educational and social skills?	

Notes: Descriptive statistics for the beliefs on channels related to gender norms available in the Appendix, Table B.13.

We use these self-reported beliefs to examine a link (in the information control group) between them and our elicited beliefs about women working longer hours on children's future outcomes. We then turn to whether our information treatment can shift these beliefs, and again, whether this treatment effect is heterogeneous to initial beliefs we have extracted.

We ask a second set of questions related to gender norms in Table 3.5. We reverse the scale for Q3, Q5 and Q6 from the variables presented in Table 3.5 so that the first category corresponds to more traditional attitudes. We then generate a gender norms score variable – ranging from 1 (traditional) to 5 (liberal) – corresponding to the within-individual average to these 6 questions. An histogram of this score is presented in Figure B.8 in Appendix. Additionally, we construct an indicator for being more liberal than the median.

Table 3.5. Gender Norms

	Answer modalities
Q1 – A pre-school child is likely to suffer if his or her mother works **Q2** – All in all, family life suffers when the woman has a full-time job **Q3** – Both the husband and wife should contribute to the household income **Q4** – A husband's job is to earn money, a wife's job is to look after the home and family **Q5** – Employers should make special arrangements to help mothers combine jobs and childcare **Q6** – Women are facing discrimination in the labor market	1. Strongly agree 2. Agree 3. Neither agree, nor disagree 4. Disagree 5. Strongly disagree

Attention check. As motivated in Subsection 3.3.2, we provide participants with a last attention check to test whether they are attentive to the survey. Upon completing the demographic information part of the survey, we provide participants with a paragraph of text wherein we ask them to report "none", in a textbox, as their current feeling. The textbox simply asks them "what is your current feeling?".[23]

Obfuscated follow-up. We invited participants back one week later and collected a set of self-reported policy views. We want to see whether information effects are persistent on policy views and whether they are heterogeneous to initial beliefs. Additionally, experimenter demand effects are a concern for our main survey. While recent evidence suggests that bias from experimenter demand effects is typically minimal (Bursztyn et al., 2022; De Quidt et al., 2018), we obfuscate the follow-up survey to alleviate this concern. Participants received a generic invitation from Prolific to take a five-minute survey which did not reveal the connection to the main survey. Among the 249 participants of the first part of the survey, we ended up with 227 (around 91%) of them taking part in the obfuscated follow-up. Further, we asked five questions but only two of these are related to our research questions and are about policies to lower the cost for mothers to work, e.g., on childcare policies and paternity leave policies. The additional questions serve to obscure a link between this survey and the original. These questions, as well as answer modalities, are presented in the Appendix, Figure B.14.

[23]See Figure B.3 in the Appendix for a screenshot of this attention check.

3.4.2 Heterogeneity in Beliefs about Children's Skills

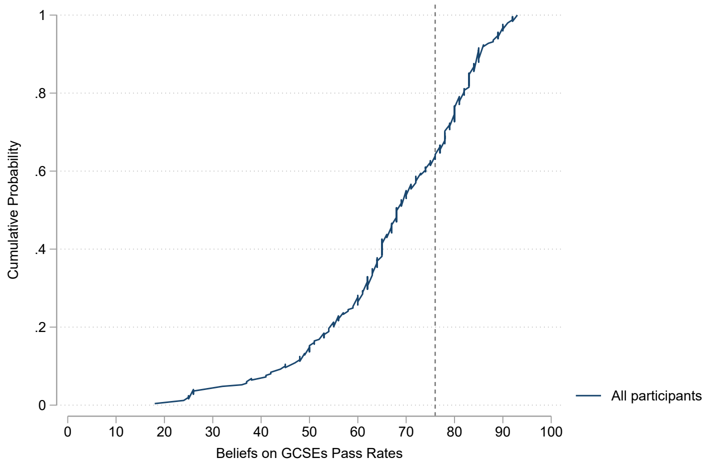
Distribution of incentivized beliefs. Figure 3.1 below provides representative evidence of participants' beliefs over two dimensions. Panel A (Subfigure 3.1a) shows the cumulative distribution function for beliefs about the passing rate of 5 or more GCSEs (with at least C/4) for when the mother works 35 hours or more per week. Panel B (Subfigure 3.1b), using only control group respondents, shows the cumulative distribution function for beliefs about the share of children out of 100 who had more behavioral problems than the median, when the mother works 35 hours or more per week. This quantitative belief elicitation allows us to assess the fraction of respondents who overestimate and underestimate GCSEs pass rates and behavioral problems. Moreover, we observe high degree of variation in these beliefs suggesting a significant degree of heterogeneity across participants.

Hypothetically elicited beliefs and incentivized beliefs. We now look at whether the elicited beliefs ($\theta_{i,k}$) from the hypothetical scenarios predict the initial incentivized belief (GCSE pass rates) of participants. In Figure 3.2a, we check this first with the perceived returns for the probability of the child to graduate from University (Panel A), and second with the perceived returns of the log of expected earnings when the child is 30 (Panel B). Although not significant, we observe there appears to be a strong positive association between both the perceived returns and the prior beliefs on the GCSE pass rates. In other words, when participants hold positive views about mothers working longer hours on both dimensions (University and earnings) – deduced from their hypothetical elicited beliefs – they tend to report higher GCSE pass rates for when the mother works full time.

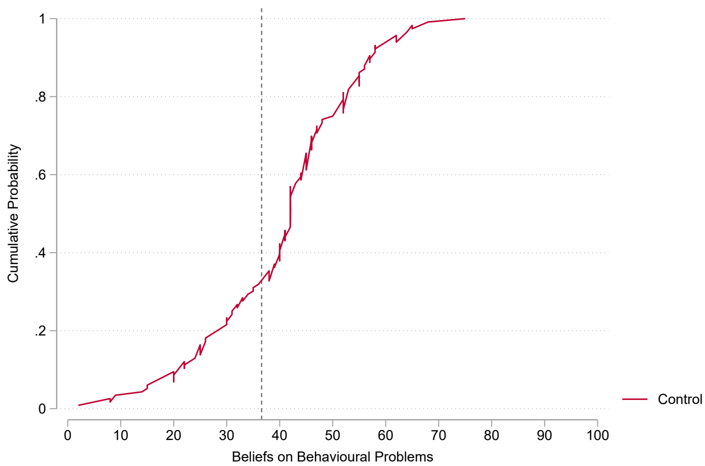
We explore this again in Figure 3.3 by reporting density plots of the GCSE beliefs split by $\theta_{graduate} \geq 0$ and $\theta_{graduate} < 0$ in Panel A and likewise in Panel B by $\theta_{earnings}$. In Panel A, we find that those with more positive perceptions of children's probability to graduate university when mothers work longer hours have a distribution of GCSE beliefs that is massed around the true GCSE pass-rate. For those with more negative views (dashed line in Panel A), we see a different story with the entire distribution

Figure 3.1. Beliefs about Children's Skills When The Mother Works Full-Time

(a) Panel A: Beliefs About GCSEs



(b) Panel B: Beliefs About Behavioral Problems

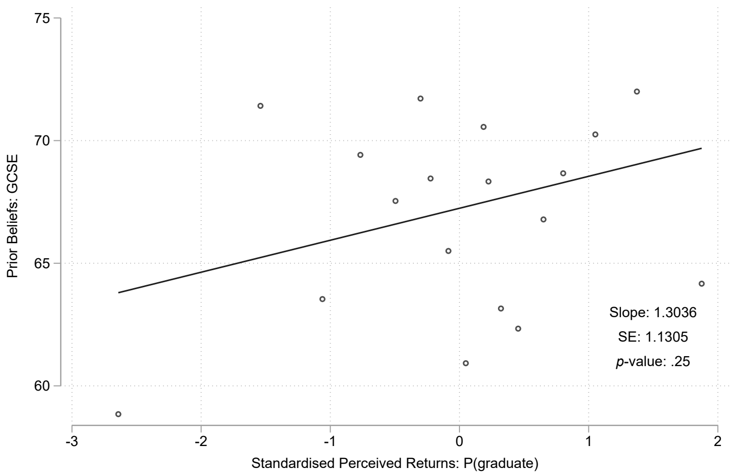


Notes: Panel A (Subfigure 3.1a) shows the cumulative distribution function for beliefs about the passing rate of 5 or more GCSEs (with at least C/4) for when the mother works 35 hours or more per week. Panel B (Subfigure 3.1b) reports the same for beliefs about behavioral problems using only control group respondents. In both panels, the short-dashed lines respectively indicate the true levels.

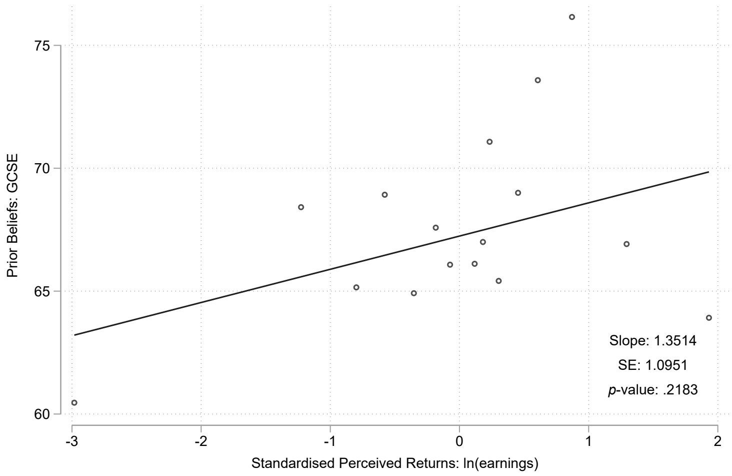
shifted to the left with a greater density over a much lower GCSE pass-rate. In Panel B, where we split by $\theta_{earnings}$ we see a similar, though less stark, pattern. Thus, overall, those with more positive perceptions based on the hypothetically elicited beliefs have on average more accurate incentivized beliefs about the GCSE pass-rates.

Figure 3.2. Hypothetically Elicited Beliefs and Beliefs about GCSE Pass Rates

(a) Panel A: Perceived Return on $\mathbb{P}$(graduate) and GCSE pass rates



(b) Panel B: Perceived Return on ln(earnings) and GCSE pass rates

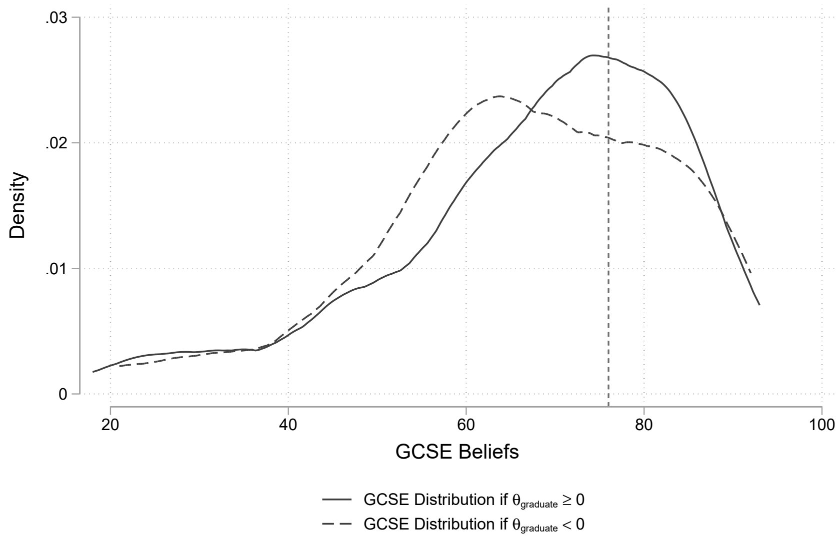


Notes: $N = 249$. Both dimensions of perceived return are standardised, to be read as z-scores, with a mean equal to 0 and a standard deviation (SD) equal to 1. A one SD shift in the perceived returns for the probability to graduate is associated with higher beliefs on GCSE pass rates, for when the mother works longer, by about 1.30%. A one SD shift in the perceived returns for the child's log of expected earnings at 30 is associated with higher beliefs on GCSE pass rates, for when the mother works longer, by about 1.35%.

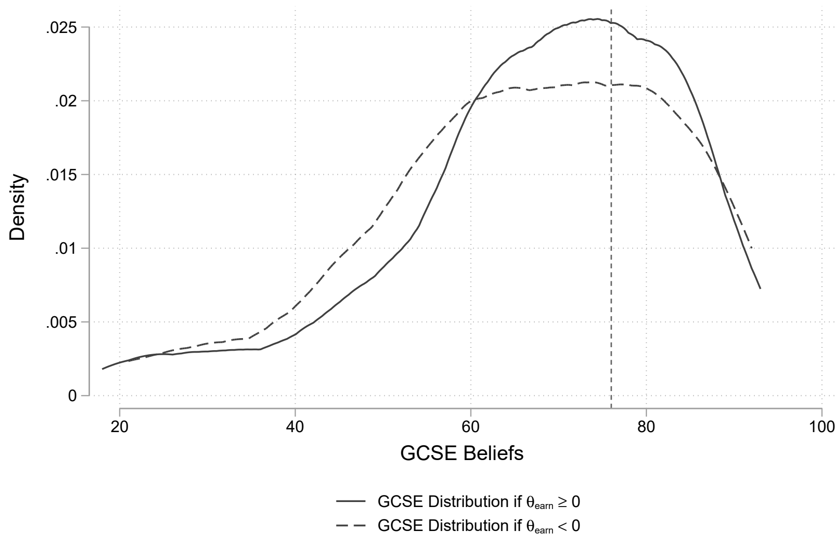
Predictors of beliefs and norms. Our aim here is to understand how both the perceived returns drawn from the hypothetical scenarios and the incentivized belief

Figure 3.3. Distribution of GCSE Pass Rate Beliefs by Hypothetically Elicited Beliefs

(a) Panel A: Split by $\theta_{graduate}$



(b) Panel B: Split by $\theta_{earnings}$



Notes: $N = 249$. We report density plots for the GCSE pass-rate expectation splitting by positive or negative values of $\theta_{graduate}$ and $\theta_{earnings}$.

on GCSE pass rates associate with a range of self-reported beliefs related to gender norms. To do so, we regress each self-reported gender norms in Table 3.4 and the scale based on Table 3.5 separately on each of the perceived return measures as well as the incentivized belief about GCSE pass rates. We also explore how the GCSE pass-rates correlate with a range of participant characteristics. We report these results in the Appendix, Table B.12.

These associations are limited due to our sample size for the pilot. However, all measures of initial beliefs appear to positively predict the gender norms index (score), i.e., more positive beliefs about women working longer hours predict more liberal self-reported gender norms. Also, there is suggestive evidence that the initial beliefs are positively linked with expecting that when mothers control the allocation of the budget they will allocate more to the child. At the same time, more positive initial beliefs appear to also correlate with believing the father to be more likely to control the budget. Thus, participants with more positive values of the $\theta's$ (believed returns to mothers working longer hours) seems to expect better resource allocation from mothers but fathers to be the one more likely exercising control over the allocation. However, note that we ask about who is more likely to make resource allocation decisions when both the mother and father earn the same amount. This may not match well with our hypothetical design where in a given scenario the parents never have equal income. Nevertheless, beliefs about resource allocation may play a role in the heterogeneity of the initial beliefs that we observe as well as overall traditional versus liberal gender norms.

3.4.3 Information Treatment Effects

We go through the following set of information treatment effect results: (i) belief updating (expected behavioral problems), (ii) beliefs on gender norms, and (iii) policy views from an obfuscated one-week follow-up survey.

Belief Updating

Information – average effect. We begin with the incentivized belief on the share of children out of 100 with more behavioral problems than the median when mothers work full-time. We initially estimate an average treatment effect for exposure to information about the GCSE pass rate when mothers work full-time or more hours given by the following:

$$y_i = \beta_0 + \gamma D_i + \sum_{j=1}^{J} \beta_j X_{ij} + \epsilon_i. \tag{3.4}$$

The outcome y_i is the behavior belief for each individual (i) rescaled to lie between 0 and 1.[24] Exposure to the information treatment is captured by $D_i = 1$ and otherwise it is 0. The X_{ij}'s are individual and predetermined demographic variables.

Information – effects by prior belief splits. We expect, though, that this treatment effect may be heterogeneous based on initial beliefs. We look at this first by disaggregating the treatment effect across under- and over-estimators of the GCSE initial belief. We define a binary indicator, U_i for under-estimators as those who report a value strictly less than the actual pass rate of five or more GCSE's when mothers work full-time or more. We then estimate the following:

$$\begin{aligned} y_i = \beta_0 + \gamma_1 D_i \times \mathbb{1}\{U_i = 1\} + \gamma_2 D_i \times \mathbb{1}\{U_i = 0\} \\ + \gamma_3 U_i + \sum_{j=1}^{J} \beta_j X_{ij} + \epsilon_i. \end{aligned} \tag{3.5}$$

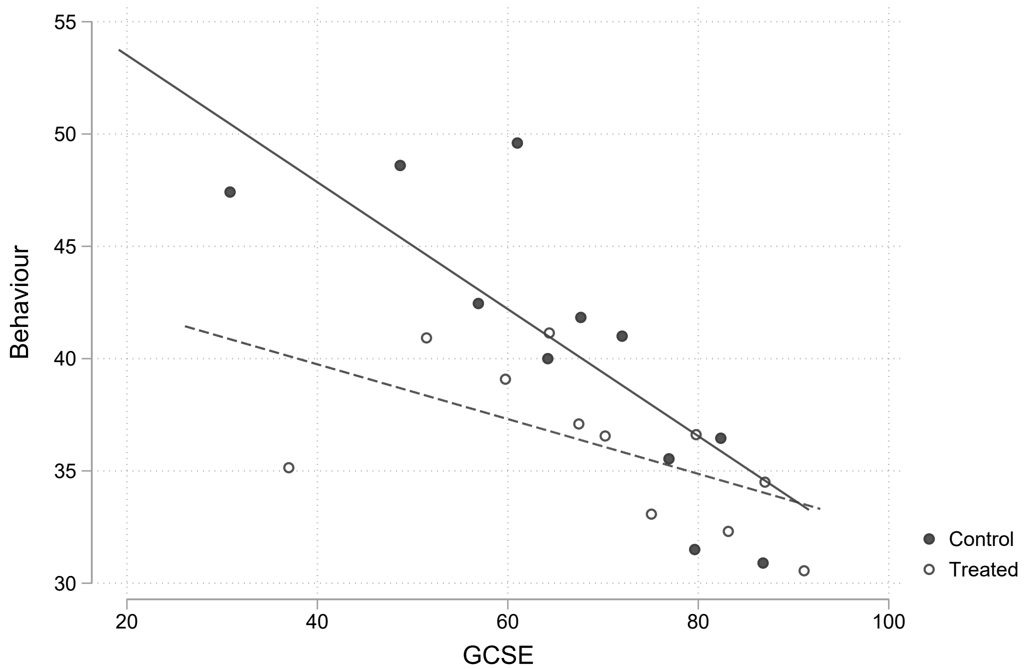
Second, we repeat this exercise but use the hypothetical scenario-based perceived returns estimates for when the mother works longer hours. We do this replacing U_i with an indicator for whether a person has a strictly negative perceived return or a positive or null perceived return looking separately at both return dimensions, i.e., by $\theta_{graduate,i}$ and then by $\theta_{earnings,i}$.

Information Results – average effects and by prior beliefs. We start in Figure 3.4 by plotting beliefs on behavior (the posterior) against the initial GCSE pass-rate expectation split by treatment status. For control participants, the prior belief (GCSE) predicts the posterior belief (behavior). Control participants yield a negative relationship between the two beliefs suggesting those who expect low GCSE pass-rates also

[24] We asked participants to report the share of children with more behavioral problems than the median for families where the mothers worked at least full-time hours in whole numbers ranging from 0 to 100. Thus, we rescale this dividing by 100.

expect high behavioral problems when mothers work full-time relative to working part-time or less in families with similar education and income levels. Treated participants, however, show a weaker relationship between the prior GCSE belief and the posterior. It appears the treatment partially breaks the link between the GCSE initial belief and the behavioral belief.

Figure 3.4. Belief Updating in Response to the Information Treatment



Notes: This figure displays a binscatter plot of the behavioral belief against the initial GCSE pass-rate belief split by treatment status.

Next, we report results for the specifications in equations 3.4 and 3.5 in Table 3.6. The average effect from the information treatment is a reduction (or improvement) in beliefs about behavioral problems for children when mothers work full-time or more by about 4.4 percentage points (*pp*). In column (2), we show that this information treatment effect seems to be driven by participants who under-estimated GCSE pass rates. Under-estimators have a significant and negative treatment effect estimate of about a 6.3*pp* decrease, while over-estimators have a insignificant and close to zero estimate.

In column (3), we look at results split by participants who hold a positive versus negative perceived return to University graduation when mothers work longer hours. We find a significant and negative information effect ($\approx -6.1pp$ in column (3)) for those with a more positive perception of mothers working longer hours relative to fathers ($\theta_{graduate,i} \geq 0$). Those who hold a negative perceived return ($\theta_{graduate,i} < 0$) on average have a weaker point estimate ($\approx -2.8pp$) that is insignificant in our pilot data. The picture is less clear, however, when we split by more positive versus negative perceived returns to earnings ($\theta_{earnings,i}$) in columns (4). Here, we find significant, negative effects for both groups and the effects appear even stronger for those with more negative views.

Table 3.6. Belief Updating and Information Effects by Sub-Groups of Initial Beliefs

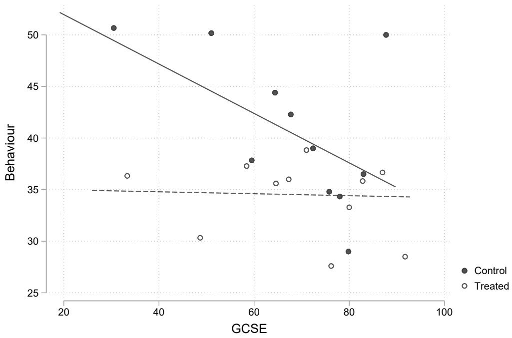
	(1) All Participants	(2) By GCSE Beliefs	(3) By $\theta_{\text{graduate,i}}$	(4) By $\theta_{\text{earnings,i}}$
ATE: γ	-0.044*** (0.016)			
GCSEs under-estimators: γ_1		-0.063*** (0.019)		
GCSEs over-estimators: γ_2		0.001 (0.025)		
($\theta_{\text{graduate,i}} \geq 0$)× Treat: γ_1			-0.061*** (0.021)	
($\theta_{\text{graduate,i}} < 0$)× Treat: γ_2			-0.028 (0.025)	
($\theta_{\text{earnings,i}} \geq 0$)× Treat: γ_1				-0.035* (0.019)
($\theta_{\text{earnings,i}} < 0$)× Treat: γ_2				-0.058** (0.028)
Individuals	249	249	249	249
($\gamma_1 - \gamma_2$) Test: p-value		.0394	.3087	.5019

Notes: * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors in parentheses. This Table presents regressions of equation 3.5 for all participants (column 1), by prior beliefs (column 2), and by average perceived returns for the probability of the child to graduate from University (column 3-4) and for the log of expected earnings (column 5-6). Control variables include participant's gender, age, a dummy variable for ethnicity (white vs. non-white), and a dummy variable for having a degree or less.

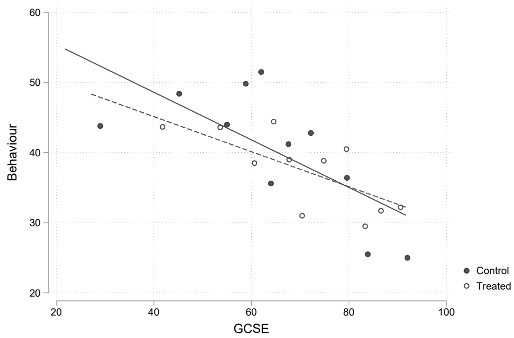
To explore these results further, we repeat analysis in Figure 3.5 with binscatter plots based on equally spaced bins of the GCSE pass-rate belief. We now, however, condition on sub-samples based on splits of the θ's around positive or negative values. In Panel A, we report this analysis for $\theta_{graduate} \geq 0$ and see a stark pattern. Those who received the treatment in this group completely break the link between the initial GCSE pass-rate belief and behavior. In Panel B, for participants with $\theta_{earnings} < 0$ we see the complete opposite, with a strong negative relationship between behavior and the GCSE pass-rate regardless of treatment status. Put differently, the information responders were those with more positive views based on $\theta_{graduate}$, while those with more negative views did not respond at all.

Figure 3.5. Beliefs Updating in Response to the Information Treatment

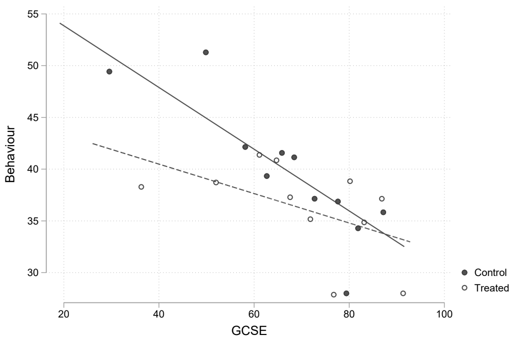
(a) Panel A: When $\theta_{graduate} \geq 0$



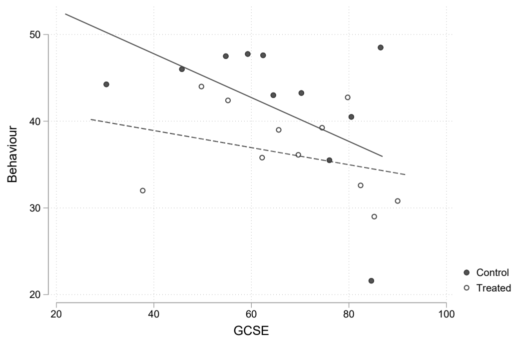
(b) Panel B: When $\theta_{graduate} < 0$



(c) Panel C: When $\theta_{earnings} \geq 0$



(d) Panel D: When $\theta_{earnings} < 0$



Notes: Binscatter plots are reported for behavior beliefs against the GCSE pass-rate beliefs each split by treatment status.

The patterns in Panels C and D are less clear. Here we split by positive (Panel C) and negative (Panel D) values of $\theta_{earnings}$ and see evidence that both groups have some response to treatment. The response here appears nevertheless much weaker that what we see in Panel A when looking at those with positive views of mothers working longer hours based on the likelihood a child graduates university. This continues the trend in our results of a clearer picture with $\theta_{graduate}$ but less so with the earnings dimensions. It may be that participants are less clear or more unsure how to answer for the earnings dimension leading to more noise and less clarity. This is an issue we need to think about beyond the pilot.

Information effects and the degree of learning. Finally, we replace the binary splits around U_i in equation 3.5 with the perception gap (PG_i). This is the difference between the prior GCSE beliefs and the signal participants receive (i.e., 74% the actual pass rate of 5 GCSEs with at least C/4 when the mother works more than 35 hours). We use this, as in Haaland and Roth (2023), to assess the degree of learning in response to the information treatment. The specification is now

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 D_i \times PG_i + \beta_3 PG_i + \sum_{j=1}^{J} \beta_j X_{ij} + \varepsilon_i. \tag{3.6}$$

Participants with a positive perception gap are over-estimators on GCSE pass-rates and those with a more negative gap are under-estimators. In Table 3.7, we see that among all participants there is stronger updating when the perception gap is more negative. We illustrate this in Figure 3.6 by plotting the linear predictions of the behavioral belief across the perception gap split by treatment status. The information treatment leads to less biased and more accurate beliefs among those who had negative perceptions and this is particularly strong for more negative perceptions.

In Table 3.7, we then split the sample by positive or negative initial beliefs based on $\theta_{graduate,i}$. We see that among those a positive view of women working – based on $\theta_{graduate,i}$ – the treatment leads to better expectations about behavior when mothers work at least full-time, particularly for those who initially misperceived the GCSE pass-rate. Again, we find no effects for those with initially more negative views. Turning to the same analysis split by the $\theta_{earnings,i}$ dimension we continue to see a similar pattern across either positive or negative values.

Table 3.7. Beliefs Updating and the Perception Gap

	All articipants	By $\theta_{\text{graduate,i}}$		By $\theta_{\text{earnings,i}}$	
		≥ 0	< 0	≥ 0	< 0
Perception Gap × Treat	-0.002*	-0.002*	-0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Perception Gap	0.003***	0.002**	0.003**	0.003***	0.003
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Treatment	-0.025	-0.048**	-0.006	-0.021	-0.038
	(0.017)	(0.021)	(0.029)	(0.020)	(0.034)
Individuals	249	127	122	150	99
Controls	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var: Control Group	.4063				

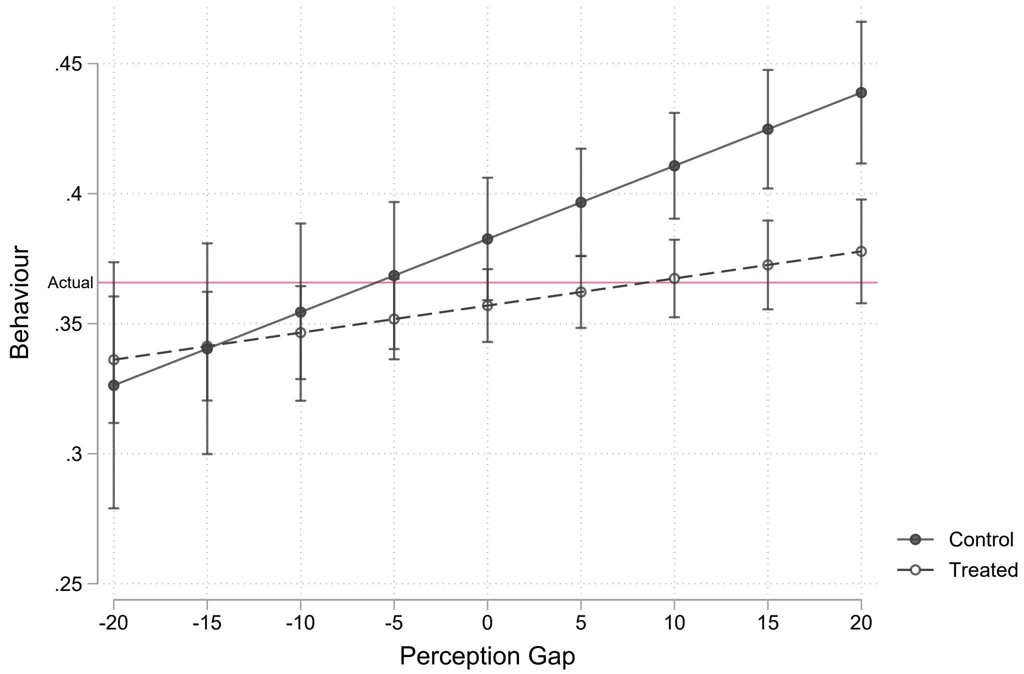
Notes: * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors in parentheses. This table shows OLS regression results of equation 3.6 where the dependent variable is the posterior belief, i.e., the reported belief, for when the mother works 35+ hours a week, on the share of children having more behavioral problems than the median. The estimates for all participants are then split in Columns 2 and 3, by values of $\theta_{i,k}$. For the ease of interpretation, we divide this by 100 for a final variable ranging from 0 to 1.

Summary. We now summarize our results on belief updating.

Result 3. *Information about GCSE pass-rates when mothers work full-time yields more positive and accurate beliefs on children's behavior. This appears to be driven by a strong response to information by those who already hold more positive views about children graduating from university when women work longer hours who then fully break any link between the initial beliefs on GCSE pass-rates and expectations about behavior.*

We do not draw a strong conclusion here due to our small, pilot sample, but our key result in Result 3 bears some discussion. Those who respond to information appear to be those who were unsure about GSCE pass-rates when mothers work full-time relative to the rate when mothers work part-time or less but at the same time hold weaker beliefs about the overall harm to children's future when mothers work more than fathers. This is suggestive that responders to information are those who while unclear about how well the children will do in terms of GCSEs nevertheless have more positive expectations overall when women work. We stress, though, that the results when split

Figure 3.6. Belief Updating: Perception Gap



Notes: This Figure displays the linear predictions of the behavioral belief across the perception gap by treatment status. Linear predictions are based on the regression results reported in Table 3.7, column 1 for all participants corresponding to the specification in equation 3.6.

by $\theta_{earnings,i}$ do not allow us to conclude that beliefs only respond when intial norms are more postive and we must leave further investigation to beyond the pilot. Additionally, it will be useful to allow further heterogeneity in the information treatment effect by initial beliefs. With more data, we will disaggregate the effect at quintiles of each initial belief thereby allowing for treatment effects at a strongly negative, mildly negative, equal, mildly positive, and strongly positive beliefs on mothers at work and returns to children's future outcomes.

Heterogeneity by participant's characteristics. Additionally, we look at whether there are heterogeneous treatment effects on incentivized beliefs, by individual characteristics. We regress equation 3.5 for each of the participant's main characteristic, i.e., for when the participant is a man, a woman, has a degree, etc., and we present the estimated $\hat{\gamma}$, corresponding to the average treatment effect. Our heterogeneity results are presented in Table B.15, in the Appendix, Section B.5. Overall, our pilot results suggest that women on average respond more strongly to the information, as do those with more education, suggesting information on the performance of children when mothers work longer hours may be more useful to women and those with more education.

Additional Results

We aim to look at channels to understand the effects of information under (i) self-reported beliefs about gender norms and (ii) policy views from the obfuscated follow-up. We have included a range of measures discussed previously to assess these channels. However, we lack the variation in our pilot data to adequately assess them. We include these in the Appendix Section B.5 for reference but omit them from the discussion here.

3.5 Conclusion

We contribute to the literature on gender gaps in the labor market by examining beliefs about the effects of mothers working long hours relative to fathers on children's skill development. The belief that mothers' and fathers' time has different returns in terms of children's skill development and future success in the labor market can drive different decisions by gender in terms of labor supply and ultimately be responsible for the gender earnings gap that we observe in most countries.

In order to estimate gendered beliefs, we develop a novel survey design linked to an experiment on a sample of parents in England. First, we elicit initial beliefs using hypothetical scenarios, presenting participants with a hypothetical family of a mother and father with a primary school aged child. We then extract their beliefs about the child's future outcomes (graduation from university and earnings) across whether the mother or father works longer hours in the labor market while we hold constant the family's income. Second, we incentivize beliefs presenting participants with the pass-rate for the share of five or more GCSE tests when mothers work part-time or fewer hours. We then ask them what they believe this pass-rate to be when mothers work full-time or more hours where we have compared families with similar income and education levels. Our findings point to beliefs that mothers who work longer hours relative to fathers are harmful to their child's skill development and future labor market success. There is also some evidence that these beliefs are driven by men and those who voted for more conservative parties in past elections.

Finally, we provide an information treatment about children's school performance when mothers work full-time. We find that information provision leads on average to more positive beliefs about children's behavior when mothers work full-time. The group most responsive to the information treatment are those who hold more positive views about women working longer hours based on hypothetical initial beliefs. We then argue that providing information in this context leads to belief updating among those with pre-existing more positive perceptions but who nevertheless have some uncertainty about performance, while those with pre-existing negative perceptions are less likely to react to the information.

Chapter 4

One-Child Policy in China and the Intergenerational Effects on Health

4.1 Introduction

The One-Child Policy is a family planning policy implemented nationwide in China from late 1979 to 2016 to control the country's rapid population growth. The policy strictly limited the number of children in each family to a minimum number, except those from ethnic minorities or living in rural areas. More than three decades of implementing the policy significantly changed the social dynamics and family structures in China (Settles et al., 2012; Zhang, 2017). Low fertility rates and reduced family size have led to population aging and increasing pressures on elderly care (Bai and Lei, 2020; Chen and Liu, 2009; Nie and Zhao, 2023). However, findings indicate that reducing sibling size would prompt increased investment in children, subsequently leading to improved education and health outcomes (e.g. Cáceres-Delpiano, 2006; Lee, 2008; Rosenzweig and Zhang, 2009; Zhong, 2017). Additionally, if the parental generation receives significant investment and wealth as a result of family planning policies, we expect them to allocate more resources to their offspring. Furthermore, improved health outcomes of parents can also be passed on to the next generation, resulting in better health observed among their children (Emanuel et al., 1992; Eriksson et al., 2005; Strauss and Thomas, 2007).

At present, there are only a limited number of studies addressing the effects of sibling size on health, and no research has explored the intergenerational effects of family planning policies and family size on health across successive generations. We fill this gap in literature by utilizing the one-child policy in China and examining the policy's spillover effects on the health outcomes of the next generations whose parents are single kids. Since the policy was strictly enforced in urban areas and only applied to Han Chinese[1] (Huang et al., 2016b), our study focuses on Han Chinese in urban areas with strict adherence to the policy.

We leverage data from the China Family Panel Studies (CFPS), a nationally representative biennial longitudinal survey funded by the Chinese government and conducted by the Institute of Social Science Survey (ISSS) at Peking University since 2010. The CFPS covers both economic and non-economic aspects of the Chinese population, providing rich data on economic activities, education, family dynamics, and health (Xie and Hu, 2014). We focus on three physical health measures, including the likelihood of being sick, self-rated health, interviewer-rated health and one mental health indicator – distress level, which we construct based on the Kessler Psychological Distress Scale (K6) and the Center for Epidemiologic Studies Depression Scale (CES-D).[2] We employ a regression discontinuity design (RDD), exploiting the policy cut-off in 1980 that creates a discontinuity in the number of single-kid families, thereby precisely isolating the policy's local average treatment effect (LATE) on the next generation's health outcomes. We show that our design passes multiple checks to ensure the validity of core RDD assumptions. Our results suggest that children born to policy-affected parents, especially policy-affected mothers, demonstrate better physical and mental health. Finally, we provide empirical evidence that increased investment in children's health and

[1]This group is the major ethnicity group in China, which accounts for 92% of the population (Xu et al., 2009).

[2]Detailed description of the data and our selection of outcome variables are presented in Section 4.3.

improved health outcomes in the parents' generations are mechanisms driving our results. Additionally, these parents are less demanding and more responsive toward their children, which explains the lower levels of distress observed among their children.

Our baseline results, in section 4.5, show that children born to policy-affected mothers exhibit significant improvements in both physical and mental health. Specifically, the likelihood of them being sick decreases by 1.8 percentage points. Their self-rated health improves by 8.2 percentage points, representing an increase of 20% over the mean. Interviewers also observe better health among these children, with observed health improved by 1.8 percentage points. In addition, they experience lower levels of mental distress, with the probability of having distress reduced by roughly two-thirds. We find similar but less statistically efficient results when examining children whose fathers were born after 1980 and affected by the policy. Children born to policy-affected fathers are less likely to be sick and have better interviewer-observed health. They also have lower levels of distress, although the estimate is not statistically significant. Given the significantly smaller sample size for father data and numerous evidence showing that urban Han mothers benefited from the demographic pattern created by the one-child policy (Fong, 2002; Veeck et al., 2003; Zhang, 2019), we focus primarily on the effects of the policy from the mothers' side.[3] We conduct a wide range of sensitivity checks to show that our baseline estimates from the mothers' data remain robust.

In Section 4.6, we empirically investigate several mechanisms to explain our results. First, higher investment in children's health, when family size becomes smaller, can explain better health in children. This is consistent with the quantity-quality trade-off, formulated by Becker (1960), illustrating a negative correlation between family size and the resources allocated to each child. Second, the intergenerational transmission of health and household characteristics is another mechanism that elucidates our results. A lower fertility rate, which is transmitted from grandparents to parents (Kolk, 2014; Murphy and Knudsen, 2002), leads to increased human capital investment per child, supporting our narratives on child health investment. Additionally, mothers af-

[3] Although analysis using fathers' data yields similar results, given the smaller sample size and narrower bandwidth, we interpret these results with caution. See Section 4.5.1 for more details.

fected by the policy exhibit better health outcomes, which can pass on to their children (Emanuel et al., 1992; Eriksson et al., 2005). Third, parenting practices and parent-child interactions show that policy-affected parents demonstrate high responsiveness but low parental demand toward their children, resulting in their children becoming more relaxed and exhibiting lower levels of distress. This aligns with literature on parental demands (Lo et al., 2020; Soysa and Weiss, 2014; Wong et al., 2019) and parental responsiveness (Davidov and Grusec, 2006; Miller-Slough et al., 2018), especially in the Chinese context where children are the only child (Liu et al., 2010; Lu and Chang, 2013).

Related Literature. Our research question is relevant to several strands of literature. The theoretical basis for the quantity and quality of children is commonly referred to as the quantity-quality trade-off. This framework was first theorized in the work of Becker (1960), who considered children a consumption good, requiring a family to decide not only on the number of children but also on the corresponding expenditure allocated to them. This theory was further developed in the work of Becker and Lewis (1973), Becker and Tomes (1976), and Willis (1973), which emphasized the negative correlation between the quantity and quality of children due to both the "price effect" and the "income effect", given the limited resources and time a family has to invest in its children – budget and time constraints. In essence, this trade-off arises from the fact that parents have to spread their time and resources more thinly as the number of children increases (Hanushek, 1992). This model is consistent with the resource dilution model in sociology, which demonstrates that as the number of children in a family increases, resources are divided among them, resulting in each child having fewer resources and, consequently, lower quality of life (Blake, 1981, 2022).

The literature presents mixed empirical evidence on the quantity-quality trade-off. Numerous studies have found a negative association between family size and investment in children (Cáceres-Delpiano, 2006; Chen, 2020; Lee, 2008; Li et al., 2008; Ponzo and Scoppa, 2022; Rosenzweig and Zhang, 2009). However, several others have observed no evidence of the quantity-quality trade-off (Angrist et al., 2010; Black et al., 2005; Diaz and Fiel, 2021) or even a positive relationship (Gomes, 1984; Lao and Lin, 2022; Qian, 2009). The majority of these studies focus on educational attainment and schooling as a quality indicator, with only limited literature examining health outcomes as a determinant of child quality.

Several studies provide evidence of a negative association between family size and health, consistent with the quantity-quality trade-off. Liu (2014) and Zhong (2017) investigate the one-child policy in China as an exogenous shock and find negative impacts of family size on child height. A similar study conducted by Liang and Gibson (2018) considers nutrient intake as a measure of parental investment in children and discovers that an additional sibling reduces nutrient intake by between one-tenth and one-fifth of the recommended level. Chen (2021) exploits the two-kid policy in Vietnam and shows that having another sibling worsens the health of children in terms of height and weight. Moreover, using twin data, Glick et al. (2007) indicate that unplanned births in Romania have negative effects on children's human capital, measured through nutrition and schooling and that these effects extend significantly to later-born siblings of first-born twins. Similarly, Rosenzweig and Zhang (2009) examine the effects of twinning in China and provide evidence that an additional child leads to a significant decline in self-assessed health of all children within the family.

However, some studies have found positive impacts of having siblings on children's health. Lordan and Frijters (2013) utilizing data from the Young Lives Project (YLP) in Peru find that the association between family size and health outcomes such as height is negative specifically for unplanned pregnancies, while it becomes positive for planned childbirths. Datar (2017) investigates the relationship between family size and obesity in the US and provides evidence that children with siblings have lower BMI and are less likely to be obese because they have healthier diets. Meanwhile, Millimet and Wang (2011) employing data from the Indonesia Family Life Survey observe only modest evidence. Zhong (2014) even finds null evidence of the trade-off in terms of the height and BMI using China's one-child policy, while Zhang et al. (2020) find being raised in a one-child family increases the probability of being overweight or obese.

In terms of subjective wellbeing, several studies have found being an only child negatively impacts self-reported psychological health (Wu, 2014; Zeng et al., 2020). Cameron et al. (2013), leveraging the one-child policy in China, provide evidence that these "little emperors" are less trustworthy and more pessimistic. However, Liu et al. (2010) and Rao et al. (2024) compare single-kid and multiple-kid families and show that single kids reported lower levels of psychological distress and mental health problems, as a result of higher parental responsiveness.

Our paper also relates to the literature on the intergenerational transmission of health and household characteristics across generations. Studies consistently demonstrate modest yet persistent effects in the transmission of parents' fertility patterns to their children (Kolk, 2014). Several papers have found that parents' fertility or family size preferences influence the preferences of their offspring (Anderton et al., 1987; Johnson and Stokes, 1976; Murphy and Knudsen, 2002). Another body of work focuses on investigating intergenerational transmission of health from parents to their off-spring. These studies are commonly rooted in the "fetal origin hypothesis" formulated by Barker (1990), which asserts that early-life (in-utero) health and circumstances play a crucial role in shaping health and economic conditions in later stages of life. A different framework is summarized by Strauss and Thomas (2007) who emphasize the transformation of health inputs into health outputs, given technological and biological constraints, as the mechanism explaining health transmission within a family. In particular, in addition to genetic endowments that would be transmitted across generations, non-genetic aspects of parental health also influence their ability to manage inputs into the health production function of their children. These frameworks have been validated by the work of Emanuel et al. (1992) and Eriksson et al. (2005), which show a robust correlation between the health of parents and that of their offspring. There is also a large and growing body of literature exploring the impact of external health shocks on the health of subsequent generations (e.g. Camacho, 2008; Islam et al., 2017; Moyano, 2017).

Contribution. Our primary contribution is to provide new causal evidence about the intergenerational effects of family size and family planning policies on subsequent generations. First, we leverage data from the China Family Panel Studies (CFPS), a nationally representative and one of the most comprehensive social panel surveys conducted in China (Xie and Hu, 2014), ensuring a high level of reliability and coverage. Second, we examine the spillover effects on both the physical and mental health of children whose parents were affected by the policy. Third, we also explore several mechanisms from the CFPS data to explain our results, contributing to the literature on quantity-quality trade-offs and intergenerational effects across generations. In addition, we further address the question of parenting behaviors and parent-child relationships in modern Chinese families.

The paper is structured as follows: Section 4.2 provides historical background and the introduction of the one-child policy in China; Section 4.3 describes the data and sample; Section 4.4 illustrates empirical strategy and identification assumptions; Section 4.5 presents estimation results and robustness tests; Section 4.6 investigates possible mechanisms that can explain our results and Section 4.7 concludes the paper.

4.2 Background

Since 1949, China started its industrialization process, experiencing substantial population growth, with the belief that this growth would contribute to the national effort (Zhu, 2012). However, consistent poverty and high fertility rates caused fears of overpopulation (De Silva and Tenreyro, 2017). From 1970 onwards, citizens were encouraged to marry at a later age because of the large population, and in the early 1970s, the state introduced a series of birth planning policies. In 1978, the authorities began to encourage one-child families, and in early 1979 they announced their intention to advocate for one-child families, which later became a national policy.

The one-child policy was introduced in China in 1979 as a strict family planning policy to curb the country's rapid population growth (Wang et al., 2016), and it was formally written into the country's constitution in 1982. The policy focuses on the Han Chinese, which makes up 92% of the population (Huang et al., 2016b). In principle, a couple was only allowed to have one child from late 1979, except in some rural ethnic minority areas such as Xinjiang, Yunnan, Ningxia and Qinghai.

The evolution of the one-child policy had several phases (Greenhalgh, 2008; Scharping, 2013a). It was first announced as "Best is one, at most two; eliminate third births" in the second half of 1978. In December 1979, the National Population and Family Planning Commission announced the policy as "Best is one". From February 1980, it quickly changed to "One for all" policy. Over time, however, various exceptions were made and the policy was further revised in early 1989 to "One child with exceptions for rural couples with only a daughter".

In 2016, China officially relaxed its one-child policy, marking a significant shift in its approach to population control. The policy change, which allowed families to have two children from 2015 with some modifications, reflected growing concerns about the policy's negative demographic and socio-economic impacts. The relaxation aimed to address issues such as the rapidly aging population, shrinking workforce, and gender imbalances. However, more urban families choose to have only one child spontaneously because of the financial and social pressures, even after the relaxation (Qian and Jin, 2024). The long-term effects of this policy change on family dynamics, economic stability and population health remain areas of policy debate.

Social consequences of the one-child policy (OCP). The one-child policy in China has led to several social consequences, notably a significant decline in the fertility rate, which had already been decreasing due to earlier family planning campaigns in the 1970s (Feng et al., 2014). The average family size reduced from 4.8 in the early 1970s to 3.1 in 2010 (Aird, 1983; Census Office of the State Council, 2020). Single-child families became prevalent, especially in urban areas, where about 80 percent of families consisted of three members by the end of the 20th century (Tu, 2016). Other direct outcomes included a skewed male-to-female ratio and higher fertility rates in rural areas compared to urban ones, disadvantaging rural families economically (Ebenstein, 2010; Hannum, 2003). Despite criticisms, urban daughters often benefited from the policy, receiving more family resources and achieving higher educational attainment and empowerment (Fong, 2002; Huang et al., 2016a, 2021). Long-term impacts include accelerated population aging, increased pressure on elderly care, and the rise of ”empty nest” families in urban areas (Bai and Lei, 2020; Chen and Liu, 2009; Nie and Zhao, 2023; Yuesheng, 2014; Zhu and Walker, 2021).

Policy effects on the first generation. The first generation subjected to the one-child policy experienced notable benefits, particularly in urban areas. Families were able to concentrate their resources on their single child's education and health, leading to substantial investments in these areas (Zhang, 2019). This focus resulted in higher educational attainment for females and improved overall health outcomes (Fong, 2002; Huang et al., 2016a; Rao et al., 2024). Women born under the one-child policy achieved higher educational levels (Huang et al., 2016a), and stricter early-life fertility restrictions increased female empowerment, as evidenced by a rise in female-headed households (Huang et al., 2021).

The policy also brought qualitative changes in family dynamics, including simplified family structures, reduced patriarchy in daughter-only families, and greater individual choice regarding family living arrangements and childbearing (Fong, 2002; Shi, 2017). Furthermore, greater parental involvement in childcare led to improved parent-child interactions (Short et al., 2001). As a result, this generation enjoys higher income levels and reduced overall stress. The persistence of intergenerational income in urban China also highlights the lasting economic impacts of the policy (Yi, 2016).

Policy cut-off in this paper. We examine the introduction of the policy during the 1979-1980 period to identify the exact policy cut-off for our study. In China, before the birth of the one-child policy, the government had imposed restrictions on the number of children a couple could have. In 1977 and 1978, both urban and rural couples were required to limit their family size to only two children (Hardee-Cleaveland and Banister, 1988). In early 1979, several intentions for the universal one-child policy were introduced, and the policy “Best is One” was officially announced in December 1979. “One for all” policy followed quickly in February 1980, clearly stating that every couple is only allowed to have one kid. Later on, the Chinese Communist Party's Central Committee issued a public letter urging all party members and the Communist

Youth League to adhere to the one-child policy on 25th September 1980, a date often mentioned as the policy's "official" start date (Scharping, 2013b). The revised 1980 Marriage Law, ratified during the Third Session of the National People's Congress on September 10, 1980, also explicitly mandated that all couples must practice birth control (Hardee-Cleaveland and Banister, 1988; Hare-Mustin, 1982; Santana Cooney et al., 1991). Before that, however, strict fines for violating the one-child policy already started to be imposed nationwide in January 1980 (Santana Cooney et al., 1991). During this period, abortions were required in several provinces in China, even in the second and third trimesters of pregnancies. The number of induced abortions increased sharply in 1979 and rose even higher in subsequent years (Hardee-Cleaveland and Banister, 1988).

We expect that there was a sharp increase in the proportion of single-child births in 1980 (the first quarter of 1980 according to our data structure we will mention later). Although official announcements and legislation related to the one-child policy and birth control were issued in September 1980, the policy started to be strictly enforced in 1980 with a wide range of rigorous measures like required abortions and birth control practices. Additionally, beginning in 1980, penalties were imposed on women who had a second child without official permission (Hardee-Cleaveland and Banister, 1988), confirming that it was difficult for Han mothers living in urban areas to have another child. We later verify this cut-off date by illustrating the discontinuity in the ratio of individuals with no siblings at our predicted cut-off point (the first quarter of 1980) within our dataset.

4.3 Data and Descriptive Statistics

4.3.1 Data

We use the China Family Panel Studies (CFPS) from the Institute of Social Science Survey (ISSS) at Peking University, China.[4] CFPS is a nationally representative, bien-

[4]The data are from the China Family Panel Studies (CFPS), funded by the 985 Program of Peking University and carried out by the Institute of Social Science Survey of Peking University.

nial longitudinal survey of Chinese families and individuals, starting in 2010. For our analysis, the CFPS dataset has several key features. First, it allows us to identify the intra-household relationships, which we need for our empirical strategy using parents' birth information. Second, adults were asked about their parents' information and their siblings in the first wave – 2010, providing valuable data for studying sibling size and policy effects. Third, with six waves up to 2020, the CFPS provides national-level information on family dynamics and health outcomes for our study.

Children's health status. We use the following outcomes as measures of children's health status:

1. Whether a child was ever sick in the last month: CFPS interprets sickness as a situation in which the child experiences physical discomfort and needs to take treatment (medicines or others).
2. Children's self-rated health: The survey asked respondents to rate their health status on a scale from 1 to 5 indicates healthy, fair, relatively unhealthy, unhealthy, and very unhealthy respectively. Only those aged 10 and above answered this question. We recode this into a binary variable with 1 for healthy and 0 for all other ratings.
3. Interviewer-observed child health: The interviewer from the ISSS recorded their assessment of the health of the presented respondent, choosing from 1 (worst) to 7 (best). We also recode this variable into a binary format, with 1 for observed health rating greater than or equal to 4, otherwise 0.
4. Distress indicator based on K6 and CES-D: We incorporate two psychological scales – Kessler Psychological Distress Scale (K6) (Kessler et al., 2002) and the Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977b) – to measure children's mental health because these scales are available in different waves. A detailed description of the mental health scales can be found in the Appendix C.1. We construct a consistent variable called "distress", which is coded 1 if a child shows signs of mental distress, i.e. K6 is greater than or equal to 5 (Prochaska et al., 2012), CES-D8 is greater than or equal to 7, or CES-D20

is greater than or equal to 16 (Bi et al., 2023).

Mechanisms. We will later examine several potential mechanisms through which the effects of the policy on the first generation, now parents, could be passed on to the second generation, their children. First, we look at family income and expenditure, focusing on expenditure directly on the child. This may influence the family resources available for the child's well-being. Second, we examine the policy's impact on parents' health status, the number of siblings they have, their fertility choices – the number of children they decide to have, and their education level. Together, these factors shape the environment in which children grow up and can affect their upbringing. Lastly, we look at the interactions between parents and children, which can affect children's mental health. The quality and nature of these interactions are crucial in determining the emotional and psychological well-being of children. By studying these mechanisms, we aim to understand how policy effects are transmitted across generations.

4.3.2 Sample Selection and Summary Statistics

Table 4.1 presents summary statistics for our sample. We compile data from the six waves of CFPS based on information from the child questionnaires to create a repeated cross-sectional dataset and merge it with data on parents and family from the corresponding questionnaires. We first focus on the sample with parents born before and after 5 years of the policy cut-off in 1980. We then drop observations missing key demographic characteristics such as age, gender, and rural or urban residence, birth information of both parents, and family size (around 1% dropped at this stage). Next, we exclude observations from provinces with fewer than 50 observations and restrict the sample to Han ethnicity (87.11%) and urban parents (51.56% after keeping only Han ethnicity) due to the policy focus.

These steps ensure complete information on the birth years and months of urban Han parents. Due to the specific focus on urban Han parents, which limits the number of observations, we construct our data based on quarterly birth information for these parents to ensure statistical power. We then exclude those with missing information on children's health outcomes at each wave, resulting in different sample sizes for various outcomes. Finally, we create separate datasets using either mothers' or fathers' information along with their children's data. Overall, our final sample size is approximately 4,000 observations.

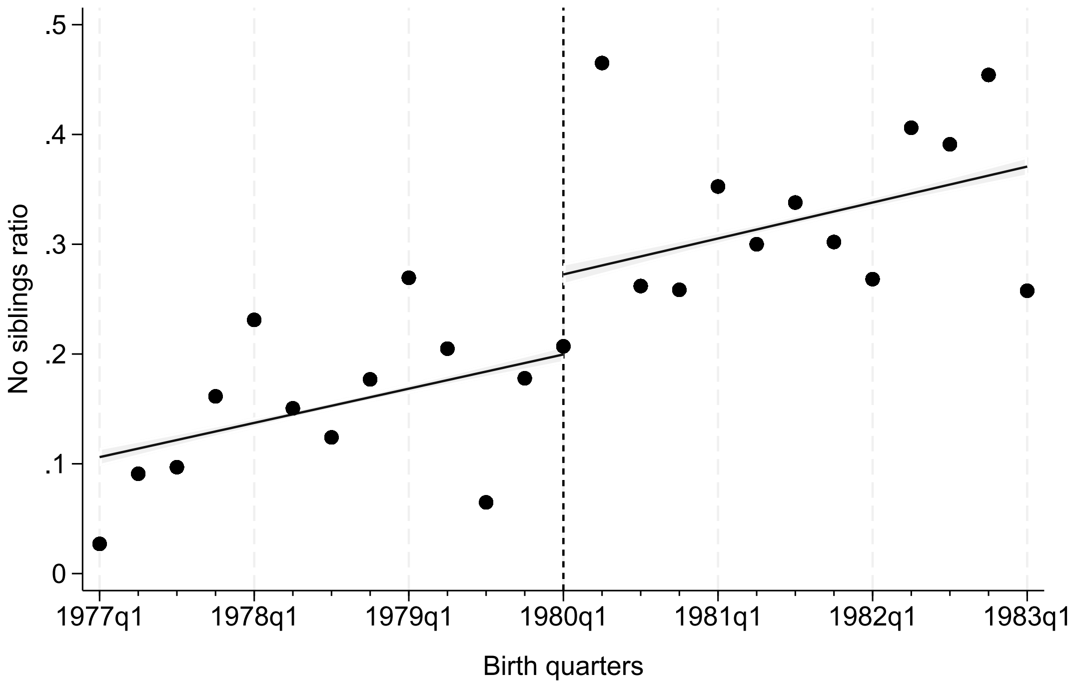
Table 4.1. Descriptive statistics

	Mothers				Fathers			
	All	Control	Treated	Diff	All	Control	Treated	Diff
A. Health Outcomes								
Child was ever sick last month	0.26	0.25	0.27	-0.02	0.28	0.28	0.28	-0.00
	(0.44)	(0.43)	(0.45)	(-1.70)	(0.45)	(0.45)	(0.45)	(-0.02)
Self-rated health (healthy = 1)	0.38	0.38	0.38	-0.01	0.39	0.37	0.41	-0.04
	(0.49)	(0.48)	(0.49)	(-0.24)	(0.49)	(0.48)	(0.49)	(-1.39)
Interviewer-observed health ($\geq$ 4 on 1-7 scale)	0.98	0.98	0.98	-0.00	0.98	0.98	0.97	0.00
	(0.14)	(0.15)	(0.13)	(-0.56)	(0.15)	(0.15)	(0.16)	(0.47)
Distress indicator based on K6 and CESD	0.12	0.14	0.10	0.04*	0.11	0.11	0.11	0.00
	(0.33)	(0.34)	(0.30)	(2.42)	(0.31)	(0.31)	(0.31)	(0.11)
B. Demographics								
Child's age	8.29	9.23	7.48	1.76***	7.74	8.36	7.15	1.21***
	(4.26)	(4.18)	(4.16)	(13.96)	(4.19)	(4.30)	(3.99)	(9.42)
Child's gender (female = 1)	0.48	0.47	0.50	-0.03	0.49	0.47	0.50	-0.03*
	(0.50)	(0.50)	(0.50)	(-1.74)	(0.50)	(0.50)	(0.50)	(-2.23)
Child's birthyear	2006	2005	2007	-2.36***	2007	2006	2008	-1.62***
	(3.94)	(3.89)	(3.64)	(-20.68)	(3.75)	(3.89)	(3.43)	(-14.31)
Mother's age	34.75	36.16	33.51	2.65***	33.63	34.79	32.51	2.28***
	(3.64)	(3.36)	(3.42)	(25.46)	(4.32)	(4.25)	(4.09)	(16.69)
Mother's birthyear	1980	1978	1981	-3.21***	1981	1980	1983	-2.70***
	(1.84)	(0.81)	(1.00)	(-117.88)	(3.18)	(3.03)	(2.73)	(-29.49)
Father's age	36.93	38.08	35.89	2.19***	35.26	36.56	33.99	2.57***
	(4.50)	(4.48)	(4.26)	(15.13)	(3.68)	(3.47)	(3.41)	(23.27)
Father's birthyear	1978	1976	1979	-2.95***	1980	1978	1981	-3.01***
	(3.47)	(3.29)	(3.01)	(-29.77)	(1.75)	(0.80)	(0.96)	(-110.72)
Family size	4.82	4.73	4.90	-0.17**	5.17	4.84	5.48	-0.64***
	(1.75)	(1.74)	(1.76)	(-3.16)	(1.90)	(1.59)	(2.12)	(-11.13)
C. Grandparent's characteristics								
Grandfather's age	59.03	61.37	56.87	4.50***	58.61	60.71	56.46	4.25***
	(6.14)	(6.18)	(5.26)	(24.62)	(5.76)	(6.00)	(4.60)	(23.16)
Grandmother's age	60.95	62.72	59.35	3.37***	61.09	63.00	59.13	3.87***
	(6.03)	(6.16)	(5.45)	(18.07)	(6.12)	(6.00)	(5.60)	(19.23)
Literacy (grandfather)	0.83	0.78	0.87	-0.09***	0.83	0.85	0.80	0.05***
	(0.38)	(0.41)	(0.33)	(-7.46)	(0.38)	(0.36)	(0.40)	(3.98)
Literacy (grandmother)	0.63	0.57	0.69	-0.12***	0.59	0.59	0.58	0.01
	(0.48)	(0.50)	(0.46)	(-7.67)	(0.49)	(0.49)	(0.49)	(0.81)
Unemployment (grandfather)	0.14	0.12	0.16	-0.04***	0.14	0.11	0.18	-0.07***
	(0.35)	(0.32)	(0.36)	(-3.39)	(0.35)	(0.32)	(0.38)	(-5.43)
Unemployment (grandmother)	0.25	0.27	0.23	0.04**	0.24	0.23	0.25	-0.02
	(0.43)	(0.44)	(0.42)	(2.98)	(0.43)	(0.42)	(0.43)	(-1.61)
Either of grandparents is communist	0.16	0.16	0.16	-0.00	0.15	0.13	0.17	-0.04***
	(0.37)	(0.36)	(0.37)	(-0.20)	(0.36)	(0.34)	(0.38)	(-3.83)
Observations	4418	2054	2364	4418	4206	2058	2148	4206

Note: The table provides the mean/standard deviation of the corresponding variables within 3 years (12 quarters) around the policy cut-off date (January 1980). "All" means the whole sample, "Treated" means mothers/ fathers have no siblings, and "Control" means mothers/fathers have siblings. "Diff" shows the mean difference between treated and control groups. The scale for interviewer-observed child health ranges from 1 (worst) to 7 (best).

We further demonstrate the validity of the policy cut-off date in our data. Figure 4.1 shows the proportion of Han urban adults born within 12 quarters of the 1980Q1 policy cut-off who have no siblings. As noted above, adult sibling information is only available in the first wave – 2010. We use weighted 2010 adult data using the sampling weights provided by the CFPS to ensure national representativeness. The graph simply shows the national average in each quarter and the solid lines show the linear fit estimated separately on each side of the cut-off. The graph presents an upward trend in the proportion of adults without siblings immediately following the policy, and suggests a jump in this ratio in 1980Q1 among Han urban residents.

Figure 4.1. Proportion of adults without siblings born within 12 quarters of policy cut-off



Notes: This figure shows the proportion of Han urban adults nationally (using sampling weights provided by CFPS) born within 12 quarters of the policy cut-off with no siblings. The policy cut-off is 1980Q1. We use the 2010 CFPS adult data as this is the only wave that provides information on adults' siblings.

4.4 Empirical Strategy

4.4.1 Main Specifications

To explore the spillovers of the policy, we employ a regression discontinuity design (RDD) with the number of quarters between mothers' date of birth and policy date as the running variable. In an RDD, the running variable plays a crucial role in determining the treatment status when there is a discontinuity in the treatment at a specific cut-off point, such as the first quarter of 1980 in our context. However, all other covariates should exhibit smoothness at the cut-off.

In this paper, we rely on a non-parametric approach to estimate the effects of the policy on the health outcomes of children born to parents exposed to the policy. In all regressions, we employ a triangular kernel weighting function where the weight assigned to each observation decreases as the distance from the cut-off increases. In addition, we estimate the effects from mothers and fathers separately.

For bandwidth selection, Table 4.2 shows the data-driven optimal bandwidths for different health outcomes following Calonico et al. (2020). We select the bandwidth equals to the average of optimal bandwidths generated for these outcomes within the parent's gender. In particular, we choose a bandwidth of 11 quarters when examining the effects from the mothers' side (maternal effects) and a bandwidth of 9 quarters when examining the effects from the fathers' side (paternal effects). All tables in our paper will show the estimates using 11-quarter bandwidth for maternal effects and 9-quarter bandwidth for paternal effects unless it has been specified differently. Specifications with other bandwidth choices will be considered in our robustness checks.

The regression measuring the direct effect of the policy on the first-generation or parents' outcomes takes the following form:

$$\begin{aligned} Y_i^P = \alpha_i^P + \beta_i^P Policy_i + f(quarter_i^P) + \gamma^P \boldsymbol{X}_i^{GP} + \theta^P \boldsymbol{Age}_i^P \\ + \lambda^P Province_i^P \times Birthyear_i^P + \tau_t^P + \nu_i^P \end{aligned} \tag{4.1}$$

where Y_i^P denotes a parent's outcomes; $Policy_i$ is a dummy variable equal to 1 if

Table 4.2. Optimal bandwidths for RDD

	Mother's information	Father's information
	(1)	(2)
Sick	9.742	10.788
Self-rated health status	9.815	6.529
Observed health status	11.911	7.673
Distress	12.210	9.774

Notes: The table shows the mean square error optimal bandwidths of main outcomes following Calonico et al. (2020) (CCT bandwidths with a linear polynomial). Standard errors are clustered at the mothers' or fathers' birth years. Column (1) presents the optimal bandwidth for each outcome using mothers' information while column (2) using father's information. The average optimal bandwidth using mother's information is around 11 quarters, and the average optimal bandwidth using father's information is around 9 quarters.

the mother's/father's date of birth is from 1980Q1. $f(quarter_i^P)$ is RD polynomials controlling for the distance from the cut-off (1980Q1) in quarters. $\boldsymbol{X}_i^{GP}$ contains pre-determined demographic and social characteristics of grandparents, including their age, literacy, employment status and whether either of them is a member of the communist party.[5] $\boldsymbol{Age}_i^P$ is a non-linear control for the parent's age, including age and age^2. We also control for the parent's province-birthyear fixed effects. τ_t^P is interview year fixed effects. Standard errors are clustered by parents' year of birth. The causal effect of the policy on the outcomes of parents is β_i^P. However, our main focus is to examine the spill-over effects of the policy on the second generation: children. The regression estimating the intergenerational effects on children's health outcomes takes the form of:

$$\begin{aligned} Y_i^C = \alpha_i^C + \beta_i^C Policy_i + f(quarter_i^P) + \gamma^C \boldsymbol{X}_i^{GP} + \theta^C \boldsymbol{Age}_i^P + \delta^C \boldsymbol{X}_i^C \\ + \lambda^C Province_i^P \times Birthyear_i^P + \zeta^C Province_i^P \times Birthyear_i^C + \tau_t^C + \nu_i^C \end{aligned} \tag{4.2}$$

where Y_i^C denotes a child's health outcomes. In addition to the pre-determined characteristics of the grandparents $\boldsymbol{X}_i^{GP}$ and the parent's non-linear age control $\boldsymbol{Age}_i^P$, we

[5]Party members were urged to "take the lead" in the one-child policy campaign. In September 1980, the Central Committee of the Chinese Communist Party issued an "Open Letter" to all Party and Youth League members, asking them to lead the way in implementing the policy (Committee, 1984; White, 1990).

also control for the characteristics of children $\boldsymbol{X}_i^C$, in particular, age and gender, and province-by-children's birthyear fixed effects.

In our RD design, $f(quarter_i^P)$ are RD polynomials controlling for the distance from the cut-off in quarters. We use a linear RD polynomial in the baseline specifications (Gelman and Imbens, 2019), and higher orders of RD polynomials in our robustness checks. Additionally, in robustness testing, we will also examine specifications that include the interaction between the treatment variable and the running variable $Policy_i \times f(quarter_i^P)$. This interaction term allows for different functions on either side of the cut-off.

4.4.2 Identifying Assumptions

Continuity assumption. The first assumption to make the RD design valid is the smoothness of the covariates at the cut-off point. We expect that the changes in our potential outcomes are solely due to the treatment initiated at the cut-off point. No other changes or discontinuities occur at the policy cut-off. Since our treatment is that the parents were born after the cut-off policy date, this assumption is only satisfied when all other relevant covariates related to the parent's birth date, in this case, grandparents' characteristics are continuous in 1980Q1.

We conduct a balance check on a list of pre-determined characteristics of both maternal and parental grandparents using the main specifications. These characteristics include birth year, literacy, employment status and whether either of the grandparents is a member of the communist party. Table 4.3 and Table 4.4 show evidence that our design satisfies the continuity assumption. Across all specifications, we do not observe discontinuities of grandparents' pre-determined characteristics at the cut-off quarter, except the parental grandmother's age. However, the coefficient is small in magnitude, only 0.8% compared to the mean.

No manipulation. The second assumption ensures that participants are unable to sort themselves on either side of the cut-off point. In our context, parents' birth quarters must not be manipulated around the policy date. Even though some families were pre-

Table 4.3. Pre-determined characteristics of mothers' parents

	Dependent variable is:						
	Maternal grandfather			Maternal grandmother			Either
	(1) Age	(2) Literacy	(3) Unemployed	(4) Age	(5) Literacy	(6) Unemployed	(7) Communist
Policy	-0.006	-0.087	0.082	-0.007	0.020	-0.019	-0.014
	(0.027)	(0.062)	(0.045)	(0.017)	(0.029)	(0.091)	(0.017)
Mean	59.234	0.822	0.141	61.093	0.622	0.256	0.154
Observations	3,284	3,205	3,284	3,258	3,239	3,258	3,238
R^2	0.998	0.200	0.163	0.999	0.226	0.131	0.173

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The policy cut-off is the first quarter of 1980. Regressions include mothers born within 11 quarters around the policy cut-off. Columns (1) (2) (3) show the characteristics of mothers' fathers, and columns (4) (5) (6) show the characteristics of mothers' mothers.

Table 4.4. Pre-determined characteristics of fathers' parents

	Dependent variable is:						
	Paternal grandfather			Paternal grandmother			Either
	(1) Age	(2) Literacy	(3) Unemployed	(4) Age	(5) Literacy	(6) Unemployed	(7) Communist
Policy	0.021	-0.020	0.003	-0.500**	0.056	-0.030	0.014
	(0.134)	(0.056)	(0.026)	(0.155)	(0.044)	(0.046)	(0.035)
Mean	58.558	0.838	0.141	61.243	0.581	0.222	0.155
Observations	2,257	2,215	2,257	2,233	2,205	2,233	2,226
R^2	0.965	0.304	0.196	0.667	0.299	0.180	0.205

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the fathers' birth years. The policy cut-off is the first quarter of 1980. Regressions include fathers born within 9 quarters around the policy cut-off. Columns (1) (2) (3) show the characteristics of fathers' fathers, and columns (4) (5) (6) show the characteristics of fathers' mothers.

aware of the policy in early 1979, because grandmothers need a gestation period of ten months, there is minimal to no opportunity for manipulation at the cut-off quarter (1980Q1).

4.5 Results

4.5.1 Baseline Results

We show our baseline results using mothers' birth information in Table 4.5. Sickness outcomes, which have the most observations, are available for all children in the CFPS data because they include responses from both adult proxies and the children themselves. In contrast, the other three outcomes require self-reporting or children's presence at the interview, resulting in smaller sample sizes.

The children of mothers born after the policy came into effect are less likely to be sick, rate their health status better, and show better overall health at the time of the interview. Specifically, the policy is associated with a 1.8 percentage point (pp) decrease in the likelihood of being sick in the last month, an increase of 8.2pp of children rating themselves as healthy, which represents around a 20% increase compared to the mean, and a 1.8pp increase in the probability that the interviewer rating the child as being in good health. Moreover, these children are also reported to have better mental health, being 8.2pp less likely to be distressed. The results suggest significant improvements in the physical and mental health of children born to mothers after the policy came into effect. In Figure 4.2, we also present a visual representation of the policy's impact on various child health outcomes using maternal birth information.[6] These figures complement the regression results and show a clear improvement in children's health and mental well-being associated with the policy-taking effects.

In Table 4.6, we show the results for children using fathers' birth information. The results are consistent with those based on mothers' birth information but are less statistically efficient. Children of fathers born after the policy cut-off are less likely to be sick (by 7.2pp) and have better overall observed health (by 4.6pp). They also tend to be less distressed, although this result is not statistically significant. Unlike the findings using birth information from mothers, there is no observed improvement in the children's self-rated health. Due to a much smaller sample size and narrower bandwidth of the fathers' data[7], these results tend to be less statistically powerful and

[6]See Figure C.4 for quadratic polynomial regressions.

[7]The sex-specific marital status distribution of CFPS is characterized by a higher proportion of

Table 4.5. Children's results using mothers' birth information

	(1) Sick	(2) Self-rated health	(3) Observed health	(4) Distress
Policy	-0.018**	0.082***	0.018***	-0.082**
	(0.005)	(0.018)	(0.004)	(0.024)
Mean	0.263	0.379	0.980	0.123
Observations	3,056	1,176	1,564	1,175
R^2	0.103	0.122	0.069	0.254

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The policy cut-off is 1980Q1. Regressions include children of mothers born within 11 quarters around the policy cut-off. Only those aged 10 and over responded to the self-rated health question, and self-rated health is equal to 1 if children rate themselves as healthy. The scale for interviewer-observed child health ranges from 1 (worst) to 7 (best). The observed health variable equals 1 if the rating is greater than or equal to 4. The interviewers only assessed the health of those children who were present at the interview. Distress in column (4) equals to 1 if Kessler Psychological Distress Scale (K6) is larger than or equal to 5 or CES-D8 is larger than or equal to 7 or CES-D20 is larger than or equal to 16.

Table 4.6. Children's results using fathers' birth information

	(1) Sick	(2) Self-rated health	(3) Observed health	(4) Distress
Policy	-0.072**	-0.084	0.046**	-0.042
	(0.025)	(0.045)	(0.014)	(0.036)
Mean	0.283	0.384	0.978	0.102
Observations	2,100	666	957	666
R^2	0.100	0.136	0.102	0.257

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the fathers' birth years. The policy cut-off is 1980Q1. Regressions include children of fathers born within 9 quarters around the policy cut-off.

should be interpreted with caution.

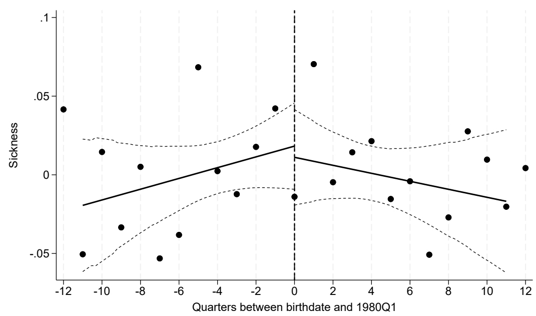
Our analysis will later on focus mothers' data for several reasons. First, datasets using father information have smaller sample sizes and narrower bandwidth, which reduces our study's statistical power. Second, research indicates that urban Han daughters have particularly benefited from the demographic patterns created by the one-child policy. In particular, many studies show that daughters in urban Han families received

unmarried males. The average of the optimal bandwidths for fathers' data is 9 quarters, compared to 11 quarters for mothers' data.

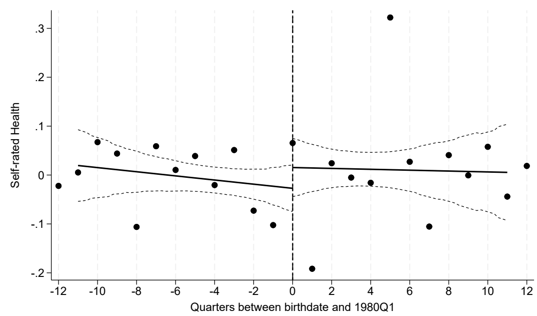
more resources and opportunities, leading to improved outcomes in education and overall well-being (Fong, 2002; Huang et al., 2016a, 2021; Veeck et al., 2003; Zhang, 2019). Therefore, we will primarily use the mothers' dataset for our main identification strategy in the following analysis.

Figure 4.2. RD plots for all outcomes

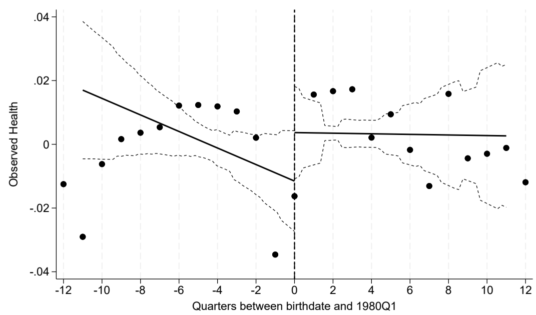
(a) Sick



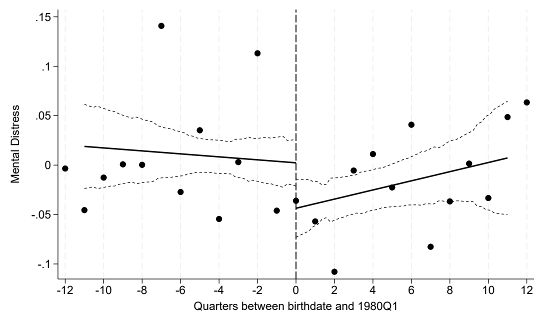
(b) Self-rated health



(c) Observed health



(d) Distress



Notes: The points depict binned residuals from a main regression of the outcome variable on a linear polynomial in birth quarter, along with other control variables. Solid lines display local linear regressions, separately estimated on each side of the cut-off, with dashed lines indicating 90% confidence intervals. Figure C.4 in Appendix C.4 displays RD plots for quadratic polynomial regressions.

4.5.2 Robustness Checks

Bandwidth sensitivity. We show bandwidth sensitivity in Appendix Figure C.1. Each sub-graph reports coefficient estimates and 90% confidence intervals of our main results for bandwidths ranging from 5 to 15 quarters. Our baseline results show robustness to changes in bandwidth.

For the sickness outcome, the policy effect remains negative and stabilizes at a small, significant level as the bandwidth increases from 10 quarters onwards. For self-rated health status, the point estimates are quite stable and positive around 0.8pp and are statistically significant when the bandwidth is greater than 8 quarters. The estimates for observed health show slight fluctuations but generally remain positive and stable. Lastly, for children's mental health, the policy effects become more pronounced with larger bandwidths, showing a consistent reduction in distress. This may be due to increased statistical power with a larger number of observations and the smaller bandwidth chosen for the main analysis, as shown in Table 4.2.

Choice of polynomial orders. In our main analysis, we use a linear polynomial of our running variable – mothers' birth quarters, which is the most common choice in RD designs. We show the sensitivity of our results to RD polynomials up to the fourth order in Appendix Figure C.2. The results suggest that our findings are not sensitive to the choice of polynomial order, reinforcing the robustness of our main results. The only exception is the sickness outcome in Appendix Figure C.2a, where policy effects disappear when we move to a third order or fourth order polynomial, although the point estimates still suggest a slightly negative effect. While higher order polynomials may obscure the significance of the sickness results, the general trend remains consistent with our main analysis.

Different specifications. We test the robustness of our main results using various specifications, as shown in Appendix Table C.2. These include models where the treatment is interacted with the running variable and with the quadratic of the running variable, changing the triangular kernel weight to no weights or panel weights provided by the data, and conducting a donut exercise that excludes observations near the cut-off (within 1 quarter from the cut-off). In general, the results are consistent with our main findings although we do lose efficiency in some results.

We first include the linear interaction of the treatment with the running variable in column (1) to allow for differences in slopes on either side of the cut-off when making extrapolation. We see some negative effects on the interaction term for the sickness outcome and the results for other outcomes are consistent with the main. We further add the interaction with the quadratic term of the running variable in column (2) to account for non-linearities in the treatment effects. The point estimates generally get larger when we do so and maintain the consistency of the results.

In columns (3) and (4), we remove the kernel weights and replace them with the panel weights to see if our results are sensitive to the weighting schemes. Here, we lose the significance on observed health status, but the point estimates suggest the same direction. Other outcomes remain stable. In column (5), we conduct a donut exercise by removing all observations close to the policy cut-off – 1 quarter, and keeping the rest of the sample to fit our main specification. The consistent results suggest that the policy effects are not driven only by observations close to the cut-off. In general, we argue that these results support the reliability of our main results across different model specifications.

Placebo tests. Our identification strategy assumes that there is a discontinuity in the treatment at the policy implementation date. One way to test against failures of this assumption is with placebo tests of different policy cut-offs. In Appendix Figure C.3, we reproduce our main specification results with alternative policy cut-offs up to 4 quarters prior and post the policy cut-off employed – 1980 Q1. We find no significant effects from the policy at the placebo cut-offs for sickness and self-rated health. However, for observed health, the effects become stronger as we move the policy cut-off to later quarters. This trend is not concerning because it suggests that the effects seen from 1980 Q1 onwards are driven more strongly by children born to mothers who were born later in the year, as suggested in Figure 4.2c. For children's mental distress, we observe a significant effect at the 1979 Q4 placebo cut-off. While this effect is noteworthy, it does not persist across other outcomes, suggesting it may be due to random variation rather than a systematic issue with our identification strategy.

4.5.3 Heterogeneous Effects

Next, we examine the heterogeneous effects of the policy by children's gender and age groups.

Table 4.7 presents the effects of the policy passed onto boys and girls separately. We run our main specification on subsamples of boys and girls separately. Boys born to a mother born after the policy have significantly higher self-rated health and lower distress, while the effects on sickness and observed health are not significant. For girls, there is a significant improvement in observed health but no significant effects on others. The effects on self-rated health and distress for girls are modest and less pronounced. This difference may be due to the tendency of adolescent girls to rate themselves more conservatively in self-assessments, whether in self-rated health or distress levels (Boerma et al., 2016; Van Droogenbroeck et al., 2018). Research has shown that girls often report lower self-rated health and higher levels of psychological distress than boys, which may explain the observed discrepancies between self-rated and observed health outcomes (Breidablik et al., 2009; Jerdén et al., 2011).

Table 4.7. Heterogeneous effects by gender

	Sick		Self-rated health		Observed health		Distress	
Gender groups	Boy (1)	Girl (2)	Boy (3)	Girl (4)	Boy (5)	Girl (6)	Boy (7)	Girl (8)
Policy	-0.036	-0.002	0.178**	0.039	0.000	0.031**	-0.108**	-0.032
	(0.034)	(0.031)	(0.057)	(0.073)	(0.010)	(0.011)	(0.039)	(0.044)
Mean	0.26	0.25	0.46	0.43	0.98	0.98	0.14	0.12
Observations	1,601	1,455	599	577	811	753	598	577
R^2	0.127	0.108	0.184	0.250	0.163	0.116	0.303	0.300

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The policy cut-off is 1980Q1. Regressions include children of mothers born within 11 quarters around the policy cut-off.

Table 4.8 presents the impact of the policy on mothers passed on to children across different age groups (0-10 and 11-15). We group children under 10 together because of the relatively small sample size in our data. The majority of significant effects are seen in the 11-15 age group, which may be because older children are more capable of self-reporting their health status and mental health. Specifically, for children aged 11-15, the effects are consistent with the main aggregated results, except for observed health, which is no longer significant but still indicates a similar trend. For the younger age group (0-10), there is a significant improvement in observed health and a reduction in distress. These two health status outcomes complement each other, indicating a generally better health status for both age groups. However, due to the constraints of the sample size, the results should be interpreted with caution.

Table 4.8. Heterogeneous effects by age groups

	Sick		Self-rated health		Observed health		Distress	
Age groups	0-10	11-15	0-10	11-15	0-10	11-15	0-10	11-15
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy	-0.003	-0.069**	0.004	0.121***	0.021*	0.009	-0.221*	-0.061**
	(0.016)	(0.018)	(0.124)	(0.021)	(0.008)	(0.011)	(0.087)	(0.019)
Mean	0.32	0.16	0.47	0.44	0.98	0.98	0.12	0.13
Observations	2,122	934	223	953	837	727	222	953
R^2	0.107	0.119	0.358	0.131	0.191	0.104	0.450	0.275

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The policy cut-off is 1980Q1. Regressions include children of mothers born within 11 quarters around the policy cut-off.

4.6 Mechanisms

In this section, we discuss potential mechanisms that we can investigate empirically from the CFPS data. First, better health in children can be explained by higher investment in child health when family sizes become smaller, which is consistent with the quantity-quality trade-off. Second, we provide evidence that parents affected by the policy show improvements in their health status, which can subsequently benefit their offspring. Finally, parenting practices and parent-child interactions reveal that these parents are highly responsive and often put less pressure on their children, explaining the lower levels of psychological distress observed among their children.

Higher investment in child health. The first mechanism is increased investment in children's health, as suggested by the quantity-quality trade-off. We expect children born to parents affected by the policy will receive higher investments from their parents, given the fact that their family size is reduced. Table 4.9 presents the estimated effects of the policy on family income, overall expenditure, and expenditure specifically allocated towards children's health. We observe a statistically significant increase in total family income, although the magnitude is negligible compared to the mean (column 1).[8] Higher family income can be attributed to either greater inherited wealth or higher personal income thanks to better education. In the next analysis, we present evidence showing that policy-affected mothers are more likely to have no siblings, potentially receiving more wealth from their parents, and they also tend to be more educated than those not affected by the policy.

In terms of expenditure, there is also no significant difference in both total expenditure (column 2) and direct medical expenses for children (column 3). Similarly, the likelihood of children having public insurance is the same for those born to policy-affected parents and those not (column 4). However, children born to mothers affected by the policy are 28.2% more likely to have commercial or private health insurance (column 5). Additionally, column 6 shows a considerable increase in parents' spending on children's commercial insurance (0.467, a 33% increase compared to the mean). This empirical evidence suggests that policy-affected mothers are more concerned about their children's health. They allocate more resources towards child healthcare and invest in preventive measures such as health insurance. However, our analysis of fathers' data reveals no increase in investment in children's health (Table C.5 in Appendix), which emphasizes that the improvements in child health primarily come from their mothers' concerns and investments.

Intergenerational transmission effects. The second possible mechanism for improved child health is the intergenerational transmission of health and household characteristics across generations. Existing literature has shown fertility patterns can be transmitted from parents to their offspring (Kolk, 2014). Reduced fertility would lead to increased human capital investment per child, which aligns with our narrative on investing in children's health. Moreover, we anticipate that the policy will have positive causal effects on mothers' health, which can, in turn, be passed on to their children (Emanuel et al., 1992; Eriksson et al., 2005).

Table 4.10 displays the directly intended effects of the policy. Those mothers af-

[8]Total family income comprises five components: wage income, total/net business income, property income, transfer income, and other income.

Table 4.9. Family income and expenditure

	(1) Total Income	(2) Total Exp.	(3) Med. Exp.	(4) Public Ins.	(5) Commercial Ins.	(6) Commercial Ins. Spending
Policy	0.221**	0.084*	-0.215	0.019	0.062***	0.467***
	(0.070)	(0.036)	(0.143)	(0.029)	(0.009)	(0.094)
Mean	10.702	10.939	5.370	0.714	0.220	1.412
Observations	3,006	3,004	1,463	3,048	3,035	3,031
R^2	0.277	0.320	0.234	0.164	0.079	0.088

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The first two are taken from family level expenditure and the rest are directly on children. Total family income comprises five components: wage income, total/net business income, property income, transfer income, and other income. We take natural logs of total income and expenditure (columns 1 and 2), medical expenditure (column 5), and commercial insurance spending (column 6). Public and commercial insurance in columns (3) and (4) are binary variables.

Table 4.10. Policy effects on mothers' demographic characteristics

	(1) No siblings	(2) Number of children	(3) College+
Policy	0.104**	-0.168***	0.137***
	(0.029)	(0.041)	(0.019)
Mean	0.103	1.661	0.228
Observations	3,135	3,146	3,146
R^2	0.292	0.246	0.233

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. No siblings is a dummy variable and College+ is column (3) takes 1 if a mother has a college degree or higher.

fected by the policy are more likely to have no siblings (column 1) and tend to have fewer children (column 2). Their fertility rate decreases by more than 10% compared to the average.[9] Furthermore, they receive better education, with their likelihood of attending college increasing by 13.7 percentage points, which is over 60% above the mean (column 3). This result is consistent with other studies examining the impacts of the one-child policy on women's education (Huang et al., 2016a; Qin et al., 2017). This

[9]We believe the policy does not influence their decision on the number of children they have. Mothers within our selected bandwidths were born between 1977 and 1982, making them 34 to 39 years old when the policy was eliminated in 2016, which means they could still have another child. Additionally, when the policy was still effective, richer couples in urban areas were willing to pay fines to have another child (Burgess and Zhuang, 2002; Li et al., 2008).

positive effect on education can also explain higher family incomes we observe above.

Table C.6 presents the policy effects on fathers' demographic characteristics. We observe contrary effects on fathers: their sibling sizes do not change and they tend to have more children. Additionally, we do not see any difference in the rate of attending college, suggesting that the policy does not improve fathers' education achievement.

Table 4.11. Policy effects on mothers' health status

	(1) Discomfort	(2) Chronic Disease	(3) Self-rated health	(4) Unhealthy	(5) Observed health	(6) Distress
Policy	-0.082**	-0.062**	-0.112**	-0.066**	0.004	-0.009
	(0.030)	(0.016)	(0.037)	(0.019)	(0.004)	(0.020)
Mean	0.245	0.074	0.225	0.186	0.981	0.175
Observations	2,477	2,477	3,141	3,141	2,632	2,476
R^2	0.052	0.046	0.219	0.129	0.040	0.150

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. Discomfort takes 1 if a mother reported physical discomfort in the last two weeks. Chronic disease is a dummy variable indicating whether a mother was diagnosed with a chronic disease in the past six months. Self-rated health is a binary variable, with 1 indicating good health, while Unhealthy takes 1 if they rated themselves as very unhealthy. Interviewer-observed health is a binary variable, with 1 indicating good health. Distress is a binary variable where 1 indicates psychological distress.

In terms of their health status, generally, we can see mothers who were born after the policy date have better health outcomes (Table 4.11). Particularly, for mothers born after the policy date, the likelihood of experiencing physical discomfort in the last two weeks decreases by 8.2 percentage points (column 1) and the probability of being diagnosed with a chronic disease in the past six months also decreases by 6.2 percentage points (column 2). These effects are not only statistically significant but also substantial in magnitude. Columns 3 and 4 present the policy effects on their self-rated health. Although they are more likely to rate their health as inferior (column 3), only 22.96% of the sample think they have good health. Therefore, we constructed a variable called "unhealthy", which was coded 1 if they perceived their body as very unhealthy. A negative and significant estimate demonstrates that the likelihood of mothers being very unhealthy is lower for those affected by the policy (column 4). With regard to interviewer-observed health and mental distress, however, we do not see any significant effects of the policy (column 5). We found similar results when examining the policy effects on fathers' health, with substantial improvements in physical health but insignificant improvements in mental health (Table C.7). These results support our main analysis that children born to policy-affected fathers have better physical health but no improvement in mental health. However, we still interpret these results with caution due to smaller sample sizes and a narrower bandwidth associated with fathers' dataset.

Parenting and Family interactions. Finally, we examine the parenting practices and interactions between parents and children to explain the lower level of distress among children whose parents are affected by the policy. Previous research has extensively examined the relationship between child-rearing practice and children's anxiety. Children may lose their chances to advocate for their interests under parental psychological control, which triggers higher levels of mental distress (Chyung et al., 2022; Luebbe et al., 2014; Rapee, 1997). McCoby (1983), building on the work of Baumrind (1971), identifies four parenting styles characterized by levels of demandingness and responsiveness. Parental demandingness or control significantly influences children's anxiety levels (Pinquart, 2017). High demands from parents cause worry and anxiety, especially for those with executive functioning deficits to manage these concerns. Conversely, low parental demands reduce anxiety among children because they may not be worried about meeting parents' expectations (Lo et al., 2020; Soysa and Weiss, 2014; Wong et al., 2019). Meanwhile, parental responsiveness is another important element that decides the level of anxiety among children. High responsiveness from parents strengthens the family bond and fosters children's social and emotional development, whereas children with less responsive parents are more prone to mental disorders and struggle with social functioning (Davidov and Grusec, 2006; McCoby, 1983; Miller-Slough et al., 2018).

We constructed several variables from the CFPS surveys to explore this mechanism. Table 4.12 reports the policy effects on the interactions between parents and children. We can see positive and significant impacts on overall home environment, as parents are more actively involved in communicating with their children (column 1). However,

Table 4.12. Interaction between parents and children

	Interviewers' observation		Children's Response		Parents' Response			
	(1) Active Communication	(2) Care about Education	(3) Quarrel	(4) Heart-to-heart Talk	(5) Give up watching TV	(6) Discuss	(7) Homework Check	(8) TV Restriction
Policy	0.020**	0.010	0.788***	0.422	0.063***	-0.064	-0.063**	-0.052**
	(0.005)	(0.009)	(0.187)	(0.450)	(0.015)	(0.037)	(0.024)	(0.015)
Mean	0.869	0.855	1.335	2.527	0.567	0.452	0.705	0.596
Observations	2,530	2,573	1,090	1,022	2,014	2,019	1,991	2,019
R^2	0.572	0.396	0.127	0.127	0.288	0.248	0.526	0.151

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The first two variables are dummy variables, showing interviewers' observations on whether parents communicate with their child actively and on whether home environment indicates parents care about their child's education. The next two variables are only reported by those aged 9-15. Quarrel refers to the number of times children quarrelled with their parents last month (column 3). Heart-to-heart talk refers to the number of times children had a heart-to-heart talk with parents last month (column 4). The last four variables are all dummy variables constructed based on parents' responses: whether parents often give up watching TV to avoid disturbing their child (column 5), whether parents often discuss happenings at school with their child this semester (column 6), whether parents often ask their child to finish homework or check their child's homework (column 7) and whether parents restrict their child from watching TV or restrict the type of TV programs their child could watch (column 8).

there is no significant difference in their concern for children's education (column 2). From children's responses[10], we see children more frequently communicate with their parents, either talking or arguing (columns 3-4). For those early teenagers who are beginning to seek independence, the increase in arguments may be a natural part of enhanced communication and may not necessarily indicate a worse environment.[11] Overall, our results suggest that parents are more responsive, and as a result, children also have more opportunities to speak up for themselves.

Meanwhile, based on parents' responses, it is evident that parents born after the policy cut-off tend to put less educational pressure on their children. They are 6.3 pp more likely to forgo watching TV to avoid disturbing their children (column 5). Although they are less likely to engage in discussions about school activities with them, this estimate is statistically insignificant (column 6). They tend to exert less pressure on their children, as evidenced by their reduced likelihood of checking children's homework or requiring their children to complete homework (column 7). Additionally, these parents are also less likely to impose restrictions on their children's TV watching

[10] Only children aged 9-15 provide answers to these questions.

[11] Table C.9 displays the results when we restrict our sample to those aged 9 or above. The results stay consistent, suggesting more communication between parents and children.

(column 8).[12]

Literature on parenting styles suggests that the ideal parenting style is "authoritative", associated with high responsiveness and an appropriate level of parental control that can promote child autonomy (Baumrind et al., 2010; Doepke and Zilibotti, 2017b). We lack sufficient evidence to determine the exact level of control these parents exert on their children; however, our results indicate that policy-affected parents exhibit high responsiveness and put less pressure on their children. Children benefit from increased parents' warmth and support, enjoy a more relaxed environment, and have greater freedom in their actions. Our findings align with existing research on Chinese parents, which indicates that increased parental responsiveness towards their children results in lower psychological distress, especially among single kids (Liu et al., 2010; Lu and Chang, 2013).

4.7 Conclusion

China's one-child policy, although formally abolished and relaxed in 2016, has had long-lasting and profound effects on the entire population. For 35 years, the policy restricted most Chinese families to one child, directly affecting at least two generations and leading to many unexpected consequences for family structure. In this paper, we empirically examine the intergenerational effects of the one-child policy on the health outcomes of children whose parents were directly affected by it. Using data from the China Family Panel Studies (CFPS), we provide causal evidence on how the policy, which significantly altered family size and dynamics, affected the physical and mental health of the next generation.

Our results show that children of mothers born after the policy was implemented show significant improvements in both physical and mental health. These children are less likely to be ill, rate their health more positively, and are observed by interviewers to be healthier. They also show lower levels of psychological distress, suggesting a positive

[12] We do not see any effects on parental practices and parent-child interaction from the fathers' side (Table C.8). This supports our main finding using fathers' dataset that there is no discontinuity in children's level of distress at the policy cut-off.

impact of the policy on their mental well-being.

We emphasize the importance of focusing on mothers' side because of the unique benefits observed for urban Han daughters under the one-child policy. These findings are consistent with the theoretical framework of the quantity-quality trade-off, where reduced family size leads to increased investment in the health and education of each child. Furthermore, our results suggest that the intergenerational transmission of health and improved parental health are key mechanisms driving these results. Parents who benefit from the policy tend to be more responsive and less demanding, contributing to the lower levels of distress observed in their children.

Our study contributes to the literature on quantity-quality trade-offs and intergenerational effects by providing comprehensive evidence from a nationally representative dataset. We highlight the broader impact of family planning policies on child health and well-being, and emphasize the need to consider both direct and spillover effects in policy evaluations.

In conclusion, despite its controversial nature and significant social consequences, the one-child policy in China has led to remarkable improvements in the health of the next generation. These findings provide valuable insights for policymakers and researchers interested in the long-term effects of family planning policies and their role in shaping population health and family dynamics.

Chapter 5

Conclusion

5.1 Summary

This thesis aims to provide evidence on the unintended consequences of income inequality, gendered labor market beliefs, and family planning policies. In three distinct essays, it provides comprehensive evidence on how peers, beliefs, and policies interact with individual and group behaviors, ultimately shaping long-term outcomes in education, labor markets, and health.

Chapter 2 focuses on the effects of income inequality among school peers on students' educational outcomes. The results underscore that exposure to income inequality within schools has heterogeneous effects across the income distribution. Low-income students benefit from an increase in low-income peers, which increases their probability of university completion, while high-income students experience a decrease in motivation and completion rates. The proposed model of social comparison helps to rationalize these findings, highlighting how relative positioning within the income distribution can generate either motivation or frustration. The descriptive evidence further shows that social integration within schools can moderate these effects, suggesting a potential mitigating solution of increasing cross-income group links and social cohesion.

Chapter 3 examines gendered beliefs about parental labor supply and its perceived impact on children's development. The novel survey design and experiment reveal that beliefs about mothers' and fathers' working hours significantly influence labor supply decisions and their children's human capital development. The study finds that beliefs about the negative impact of mothers working long hours on children's outcomes are particularly strong among men and conservative voters. However, the provision of information showing positive outcomes for children when mothers work full-time can lead to an updating of beliefs, particularly among those with initially more positive perceptions. This suggests that targeted information campaigns could be effective in addressing gendered labor market disparities.

Chapter 4 studies the intergenerational effects of China's one-child policy on health outcomes. The results show that, despite its controversial nature, the policy has had significant positive effects on the physical and mental health of the second generation in urban areas. Children born to urban Han mothers affected by the policy report better health outcomes, consistent with the quantity-quality trade-off theory. The study highlights the role of improved parental health and reduced family size in driving these results, providing important insights into the long-term and spillover effects of family planning policies on population health.

In conclusion, these chapters contribute to a deeper understanding of how income inequality, gendered beliefs, and family planning policies interact with social and individual behaviors to produce complex and often unintended consequences. The findings have important implications for policymakers aiming to address educational and income inequalities, gender gaps in the labor market, and the long-term effects of population control policies.

5.2 Further Research

For the extension of Chapter 3, we are conducting a full-scale experiment designed to address some limitations of the pilot study. This refined experiment will have more participants, thereby increasing the statistical power and robustness of the results.

Bibliography

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti (2020). “It takes a village: The economics of parenting with neighborhood and peer effects”. *National Bureau of Economic Research* Working Paper Series 27050.

Aguiar, Mark and Erik Hurst (2007). “Measuring Trends in Leisure: The Allocation of Time over Five Decades”. *The Quarterly Journal of Economics* 122.3, pp. 969–1006.

Aird, John S. (1983). “The Preliminary Results of China’s 1982 Census”. *The China Quarterly* 96, pp. 613–640.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay (2021a). “Building social cohesion in ethnically mixed schools: An intervention on perspective taking”. *The Quarterly Journal of Economics* 136.4, pp. 2147–2194.

Alan, Sule, Elif Kubilay, Elif Bodur, and Ipek Mumcu (2021b). “Social Status in Student Networks and Implications for Perceived Social Climate in Schools”. *SSRN Electronic Journal.*

Anderton, Douglas L, Noriko O Tsuya, Lee L Bean, and Geraldine P Mineau (1987). “Intergenerational transmission of relative fertility and life course patterns”. *Demography*, pp. 467–480.

Andresen, Martin Eckhoff and Emily Nix (2022). “What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples”. *Journal of Labor Economics* 40.4, pp. 971–1004.

Anelli, Massimo and Giovanni Peri (2019). “The effects of high school peers’ gender on college major, college performance and income”. *The Economic Journal* 129.618, pp. 553–602.

Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016). “Parenthood and the Gender Gap in Pay”. *Journal of Labor Economics* 34.3, pp. 545–579.

Angrist, Joshua, Victor Lavy, and Analia Schlosser (2010). “Multiple experiments for the causal link between the quantity and quality of children”. *Journal of Labor Economics* 28.4, pp. 773–824.

Angrist, Joshua D and Victor Lavy (1999). “Using Maimonides’ rule to estimate the effect of class size on scholastic achievement”. *The Quarterly journal of economics* 114.2, pp. 533–575.

Angrist, Joshua D, Victor Lavy, Jetson Leder-Luis, and Adi Shany (2019). “Maimonides’ rule redux”. *American Economic Review: Insights* 1.3, pp. 309–24.

Athey, Susan and Guido Imbens (2016). “Recursive partitioning for heterogeneous causal effects”. *Proceedings of the National Academy of Sciences* 113.27, pp. 7353–7360.

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019). “Generalized random forests”. *The Annals of Statistics* 47.2, pp. 1148–1178.

Athey, Susan and Stefan Wager (2019). “Estimating treatment effects with causal forests: An application”. *Observational Studies* 5.2, pp. 37–51.

Attanasio, Orazio, Teodora Boneva, and Christopher Rauh (2020). “Parental Beliefs about Returns to Different Types of Investments in School Children”. *Journal of Human Resources.*

Aucejo, Esteban, Patrick Coate, Jane Cooley Fruehwirth, Sean Kelly, and Zachary Mozenter (July 2022). “Teacher Effectiveness and Classroom Composition: Understanding Match Effects in the Classroom*”. *The Economic Journal.*

Bai, Chen and Xiaoyan Lei (2020). "New trends in population aging and challenges for China's sustainable development". *China Economic Journal* 13.1, pp. 3–23.

Balsa, Ana I, Michael T French, and Tracy L Regan (2014). "Relative deprivation and risky behaviors". *Journal of Human Resources* 49.2, pp. 446–471.

Barker, David J (1990). "The fetal and infant origins of adult disease." *BMJ: British Medical Journal* 301.6761, p. 1111.

Baumrind, Diana (1971). "Current patterns of parental authority." *Developmental psychology* 4.1p2, p. 1.

Baumrind, Diana, Robert E Larzelere, and Elizabeth B Owens (2010). "Effects of preschool parents' power assertive patterns and practices on adolescent development". *Parenting: Science and practice* 10.3, pp. 157–201.

Becker, Gary S (1960). "An economic analysis of fertility". In: *Demographic and economic change in developed countries.* Columbia University Press, pp. 209–240.

Becker, Gary S and H Gregg Lewis (1973). "On the interaction between the quantity and quality of children". *Journal of Political Economy* 81.2, Part 2, S279–S288.

Becker, Gary S and Nigel Tomes (1976). "Child endowments and the quantity and quality of children". *Journal of Political Economy* 84.4, Part 2, S143–S162.

Bertoni, Marco and Roberto Nisticò (2019). "Ordinal rank and peer composition: Two sides of the same coin?" *IZA Discussion Paper No. 12789.*

Bertrand, Marianne (2011). "New perspectives on gender". In: *Handbook of labor economics.* Vol. 4. Elsevier, pp. 1543–1590.

— (2020). "Gender in the twenty-first century". In: *AEA Papers and proceedings.* Vol. 110. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 1–24.

Bi, Kaiwen, Peiyu Chen, and Shuquan Chen (2023). "Validating the 8-item Center for Epidemiological Studies Depression Scale-Chinese (CESD-Chinese): Data from the China Family Panel Studies (CFPS)".

Bietenbeck, Jan (2020). "The long-term impacts of low-achieving childhood peers: evidence from Project STAR". *Journal of the European Economic Association* 18.1, pp. 392–426.

Bifulco, Robert, Jason M Fletcher, Sun Jung Oh, and Stephen L Ross (2014). "Do high school peers have persistent effects on college attainment and other life outcomes?" *Labour economics* 29, pp. 83–90.

Bifulco, Robert, Jason M Fletcher, and Stephen L Ross (2011). "The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health". *American Economic Journal: Economic Policy* 3.1, pp. 25–53.

Billings, Stephen B. and Mark Hoekstra (2023). "The Effect of School and Neighborhood Peers on Achievement, Misbehavior, and Adult Crime". *Journal of Labor Economics* 41.3, pp. 643–685.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes (2005). "The more the merrier? The effect of family size and birth order on children's education". *The Quarterly Journal of Economics* 120.2, pp. 669–700.

— (2013). "Under pressure? The effect of peers on outcomes of young adults". *Journal of Labor Economics* 31.1, pp. 119–153.

Black, Sandra E, Erik Grönqvist, and Björn Öckert (2018). "Born to lead? The effect of birth order on noncognitive abilities". *Review of Economics and Statistics* 100.2, pp. 274–286.

Blake, Judith (1981). "Family size and the quality of children". *Demography* 18, pp. 421–442.

— (2022). "Family size and achievement". *Univ of California Press* 3.

Blau, Francine D. and Lawrence M. Kahn (Sept. 2017). "The Gender Wage Gap: Extent, Trends, and Explanations". *Journal of Economic Literature* 55.3, pp. 789–865.

Boerma, Ties, Ahmad Reza Hosseinpoor, Emese Verdes, and Somnath Chatterji (2016). "A global assessment of the gender gap in self-reported health with survey data from 59 countries". *BMC public health* 16, pp. 1–9.

Bonacich, Phillip (1987). "Power and centrality: A family of measures". *American Journal of Sociology* 92.5, pp. 1170–1182.

Boneva, Teodora, Marta Golin, Katja Kaufmann, and Christopher Rauh (2022). "Beliefs about maternal labor supply". *IZA Working Paper 15788.*

Boneva, Teodora and Christopher Rauh (Mar. 2018). "Parental Beliefs about Returns to Educational Investments—The Later the Better?" *Journal of the European Economic Association* 16.6, pp. 1669–1711.

Booij, Adam S, Edwin Leuven, and Hessel Oosterbeek (2017). "Ability peer effects in university: Evidence from a randomized experiment". *The Review of Economic Studies* 84.2, pp. 547–578.

Borbely, Daniel, Jonathan Norris, and Agnese Romiti (2023). "Peer gender and schooling: Evidence from Ethiopia". *Journal of Human Capital* 17.2, pp. 207–249.

Borra, Cristina and Almudena Sevilla (2019). "COMPETITION FOR UNIVERSITY PLACES AND PARENTAL TIME INVESTMENTS: EVIDENCE FROM THE UNITED KINGDOM". *Economic Inquiry* 57.3, pp. 1460–1479.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou (2024). "Toward a general theory of peer effects". *Econometrica* 92.2, pp. 543–565.

Breidablik, Hans-Johan, Eivind Meland, and Stian Lydersen (2009). "Self-rated health during adolescence: stability and predictors of change (Young-HUNT study, Norway)". *The European Journal of Public Health* 19.1, pp. 73–78.

Burgess, Robin and Juzhong Zhuang (2002). “Modernization and son preference in People’s Republic of China”.

Bursztyn, Leonardo, Georgy Egorov, Ingar K Haaland, Aakaash Rao, and Christopher Roth (Feb. 2022). *Justifying Dissent*. Working Paper 29730. National Bureau of Economic Research.

Cáceres-Delpiano, Julio (2006). “The impacts of family size on investment in child quality”. *Journal of Human Resources* 41.4, pp. 738–754.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell (2020). “Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs”. *The Econometrics Journal* 23.2, pp. 192–210.

Camacho, Adriana (2008). “Stress and birth weight: evidence from terrorist attacks”. *American Economic Review* 98.2, pp. 511–515.

Cameron, Lisa, Nisvan Erkal, Lata Gangadharan, and Xin Meng (2013). “Little emperors: behavioral impacts of China’s One-Child Policy”. *Science* 339.6122, pp. 953–957.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez (2012). “Inequality at Work: The Effect of Peer Salaries on Job Satisfaction”. *American Economic Review* 102.6, pp. 2981–3003. ISSN: 00028282.

Carlana, Michela (2019). “Implicit stereotypes: Evidence from teachers’ gender bias”. *The Quarterly Journal of Economics* 134.3, pp. 1163–1224.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti (2022a). “Goals and gaps: Educational careers of immigrant children”. *Econometrica* 90.1, pp. 1–29.

— (2022b). “Implicit Stereotypes in Teachers’ Track Recommendations”. In: *AEA Papers and Proceedings*. Vol. 112, pp. 409–14.

Carrell, Scott E, Mark Hoekstra, and Elira Kuka (2018). “The long-run effects of disruptive peers”. *American Economic Review* 108.11, pp. 3377–3415.

Carrell, Scott E. and Mark L. Hoekstra (2010). "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids". *American Economic Journal: Applied Economics* 2.1, pp. 211–28.

Cattan, Sarah, Kjell G Salvanes, and Emma Tominey (2023). "First generation elite: the role of school networks". *23/18. London: Institute for Fiscal Studies.*

Census Office of the State Council, Population Statistics Bureau of National Bureau of Statistics (2020). "National Population Census Data." *Beijing, China: China Statistics Press.*

Chen, Feinian and Guangya Liu (2009). "Population aging in China". In: *International Handbook of Population Aging*. Springer, pp. 157–172.

Chen, Qihui (2021). "Population policy, family size and child malnutrition in Vietnam–Testing the trade-off between child quantity and quality from a child nutrition perspective". *Economics & Human Biology* 41, p. 100983.

Chen, Shuang (2020). "Parental investment after the birth of a sibling: the effect of family size in low-fertility China". *Demography* 57.6, pp. 2085–2111.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR". *The Quarterly Journal of Economics* 126.4, pp. 1593–1660.

Chetty, Raj, John N Friedman, and Jonah E Rockoff (2014). "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood". *American Economic Review* 104.9, pp. 2633–79.

Chetty, Raj and Nathaniel Hendren (2018a). "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects". *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.

— (2018b). "The impacts of neighborhoods on intergenerational mobility II: County-level estimates". *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz (2016). "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment". *American Economic Review* 106.4, pp. 855–902.

Chyung, Yun Joo, Young Ae Lee, Seon Jeong Ahn, and Hee Sung Bang (2022). "Associations of perceived parental psychological control with depression, anxiety in children and adolescents: A meta-analysis". *Marriage & Family Review* 58.2, pp. 158–197.

Clark, Andrew E, David Masclet, and Marie Claire Villeval (2010). "Effort and comparison income: Experimental and survey evidence". *ILR Review* 63.3, pp. 407–426.

Clarke, Damian, Joseph P Romano, and Michael Wolf (2020). "The Romano–Wolf multiple-hypothesis correction in Stata". *The Stata Journal* 20.4, pp. 812–843.

Committee, China Communist Party Central (1984). "Chinese Communist Party Central Committee calls on Party and Youth League members to take the lead in having only one child (per couple),[open letter, September 25, 1980]". *Chinese sociology and anthropology* 16.3-4, pp. 83–89.

Conti, Gabriella, James Heckman, and Sergio Urzua (2010). "The Education-Health Gradient". *American Economic Review* 100.2, pp. 234–38.

Cortés, Patricia, Gizem Koşar, Jessica Pan, and Basit Zafar (Oct. 2022). *Should Mothers Work? How Perceptions of the Social Norm Affect Individual Attitudes Toward Work in the U.S.* Working Paper 30606. National Bureau of Economic Research.

Cortés, Patricia and Jessica Pan (2023). "Children and the Remaining Gender Gaps in the Labor Market". *Journal of Economic Literature.*

Cutler, David M and Adriana Lleras-Muney (2010). "Understanding differences in health behaviors by education". *Journal of Health Economics* 29.1, pp. 1–28.

Datar, Ashlesha (2017). "The more the heavier? Family size and childhood obesity in the US". *Social Science & Medicine* 180, pp. 143–151.

Davidov, Maayan and Joan E Grusec (2006). "Untangling the links of parental responsiveness to distress and warmth to child outcomes". *Child development* 77.1, pp. 44–58.

De Quidt, Jonathan and Johannes Haushofer (2019). "Depression through the Lens of Economics". *The Economics of Poverty Traps*, pp. 127–152.

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth (2018). "Measuring and bounding experimenter demand". *American Economic Review* 108.11, pp. 3266–3302.

De Silva, Tiloka and Silvana Tenreyro (2017). "Population control policies and fertility convergence". *Journal of Economic Perspectives* 31.4, pp. 205–228.

Denning, Jeffrey T., Richard Murphy, and Felix Weinhardt (Oct. 2021). "Class Rank and Long-Run Outcomes". *The Review of Economics and Statistics*, pp. 1–45.

Diaz, Christina J and Jeremy E Fiel (2021). "When size matters: IV estimates of sibship size on educational attainment in the US". *Population Research and Policy Review* 40, pp. 1195–1220.

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti (2019). "The Economics of Parenting". *Annual Review of Economics* 11.1, pp. 55–84.

Doepke, Matthias and Fabrizio Zilibotti (2017a). "Parenting with style: Altruism and paternalism in intergenerational preference transmission". *Econometrica* 85.5, pp. 1331–1371.

— (2017b). "Parenting with style: Altruism and paternalism in intergenerational preference transmission". *Econometrica* 85.5, pp. 1331–1371.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2011). "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya". *American Economic Review* 101.5, pp. 1739–74.

Dynarski, Susan M (2003). "Does aid matter? Measuring the effect of student aid on college attendance and completion". *American Economic Review* 93.1, pp. 279–288.

Ebenstein, Avraham (2010). "The "missing girls" of China and the unintended consequences of the one child policy". *Journal of Human Resources* 45.1, pp. 87–115.

Elsner, Benjamin and Ingo E Isphording (2017). "A big fish in a small pond: Ability rank and human capital investment". *Journal of Labor Economics* 35.3, pp. 787–828.

— (2018). "Rank, sex, drugs, and crime". *Journal of Human Resources* 53.2, pp. 356–381.

Emanuel, Irvin, Haroulla Filakti, Eva Alberman, and Stephen JW Evans (1992). "Intergenerational studies of human birthweight from the 1958 birth cohort. 1. Evidence for a multigenerational effect". *BJOG: An International Journal of Obstetrics & Gynaecology* 99.1, pp. 67–74.

Eriksson, Tor, Bernt Bratsberg, Oddbjorn Raaum, et al. (2005). "Earnings persistence across generations: Transmission through health?" In: *Unpublished paper presented at the EALE/SOLE meeting.*

Fabregas, Raissa (2022). "Trade-offs of Attending Better Schools: Achievement, Self-Perceptions and Educational Trajectories". *Working Paper.*

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers (2021). "Socioeconomic status and inequalities in children's IQ and economic preferences". *Journal of Political Economy* 129.9, pp. 2504–2545.

Feld, Jan and Ulf Zölitz (2017). "Understanding peer effects: On the nature, estimation, and channels of peer effects". *Journal of Labor Economics* 35.2, pp. 387–428.

Feng, Xiao-Tian, Dudley L Poston Jr, and Xiao-Tao Wang (2014). "China's one-child policy and the changing family". *Journal of comparative family studies* 45.1, pp. 17–29.

Fong, Vanessa L (2002). “China’s one-child policy and the empowerment of urban daughters”. *American Anthropologist* 104.4, pp. 1098–1109.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek (2016). “Parental responses to public investments in children: Evidence from a maximum class size rule”. *Journal of Human Resources* 51.4, pp. 832–868.

Gagete-Miranda, Jessica (2020). “An aspiring friend is a friend indeed: school peers and college aspirations in Brazil”. *Manuscript*.

Gelman, Andrew and Guido Imbens (2019). “Why high-order polynomials should not be used in regression discontinuity designs”. *Journal of Business & Economic Statistics* 37.3, pp. 447–456.

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, pp. 715–746.

Glick, Peter J, Alessandra Marini, and David E Sahn (2007). “Estimating the consequences of unintended fertility for child health and education in Romania: An analysis using twins data”. *Oxford Bulletin of Economics and Statistics* 69.5, pp. 667–691.

Goldin, Claudia (2006). “The quiet revolution that transformed women’s employment, education, and family”. *American Economic Review* 96.2, pp. 1–21.

Golsteyn, Bart HH, Arjan Non, and Ulf Zölitz (2021). “The impact of peer personality on academic achievement”. *Journal of Political Economy* 129.4, pp. 1052–1099.

Gomes, Melba (1984). “Family size and educational attainment in Kenya”. *Population and Development Review*, pp. 647–660.

Gong, Jie, Yi Lu, and Hong Song (2021). “Gender peer effects on students’ academic and noncognitive outcomes evidence and mechanisms”. *Journal of Human Resources* 56.3, pp. 686–710.

Greaves, Ellen, Iftikhar Hussain, Birgitta Rabe, and Imran Rasul (2023). “Parental Responses to Information about School Quality: Evidence from Linked Survey and Administrative Data”. *The Economic Journal* 133.654, pp. 2334–2402.

Greenbaum, Lisa (2015). *This American Life.* https://www.thisamericanlife.org/550/three-miles. Episode 550: Three Miles.

Greenhalgh, Susan (2008). *Just one child: Science and policy in Deng’s China.* Univ of California Press.

Guyon, Nina, Eric Maurin, and Sandra McNally (2012). “The effect of tracking students by ability into different schools a natural experiment”. *Journal of Human resources* 47.3, pp. 684–721.

Haaland, Ingar and Christopher Roth (2023). “Beliefs about racial discrimination and support for pro-black policies”. *Review of Economics and Statistics* 105.1, pp. 40–53.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2023). “Designing information provision experiments”. *Journal of Economic Literature* 61.1, pp. 3–40.

Hannum, Emily (2003). “Poverty and basic education in rural China: Villages, households, and girls’ and boys’ enrollment”. *Comparative Education Review* 47.2, pp. 141–159.

Hanushek, Eric A (1992). “The trade-off between child quantity and quality”. *Journal of Political Economy* 100.1, pp. 84–117.

Hao, Zhuang and Benjamin W Cowan (2019). “The effects of graduation requirements on risky health behaviors of high school students”. *American Journal of Health Economics* 5.1, pp. 97–125.

Hardee-Cleaveland, Karen and Judith Banister (1988). "Fertility policy and implementation in China, 1986-88". *Population and development review*, pp. 245–286.

Hare-Mustin, Rachel T (1982). "China's marriage law: A model for family responsibilities and relationships". *Family process* 21.4, pp. 477–481.

Haushofer, Johannes and Ernst Fehr (2014). "On the psychology of poverty". *Science* 344.6186, pp. 862–867.

Heckman, James J. and Stefano Mosso (2014). "The Economics of Human Development and Social Mobility". *Annual Review of Economics* 6.1, pp. 689–733.

Hotz, V Joseph, Per Johansson, and Arizo Karimi (2018). *Parenthood, family friendly workplaces, and the gender gaps in early work careers.* Tech. rep. National Bureau of Economic Research.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow (2019). "The Allocation of Talent and U.S. Economic Growth". *Econometrica* 87.5, pp. 1439–1474.

Huang, Wei, Xiaoyan Lei, and Ang Sun (2016a). "When fewer means more: impact of one-child policy on education of girls". *Cambridge: Harvard University.*

— (2021). "Fertility restrictions and life cycle outcomes: Evidence from the One-Child Policy in China". *Review of Economics and Statistics* 103.4, pp. 694–710.

Huang, Wei, Xiaoyan Lei, and Yaohui Zhao (2016b). "One-child policy and the rise of man-made twins". *Review of Economics and Statistics* 98.3, pp. 467–476.

Islam, Asadul, Chandarany Ouch, Russell Smyth, and Liang Choon Wang (2017). "The intergenerational effect of Cambodia's genocide on children's education and health". *Population and Development Review* 43.2, pp. 331–353.

Islam, Asadul and Russell Smyth (2015). "Do fertility control policies affect health in old age? Evidence from China's one-child experiment". *Health Economics* 24.5, pp. 601–616.

Jackson, C Kirabo (2016). "The effect of single-sex education on test scores, school completion, arrests, and teen motherhood: Evidence from school transitions".

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico (Oct. 2015). "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms *". *The Quarterly Journal of Economics* 131.1, pp. 157–218.

Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, and Sebastián Kiguel (2022). "Who Benefits From Attending Effective High Schools?" *National Bureau of Economic Research* Working Paper Series 28194.

Jackson, Matthew O (2021). "Inequality's economic and social roots: the role of social networks and homophily". *Working Paper SSRN 3795626*.

Jerdén, Lars, Gunilla Burell, Hans Stenlund, Lars Weinehall, and Erik Bergström (2011). "Gender differences and predictors of self-rated health development among Swedish adolescents". *Journal of Adolescent Health* 48.2, pp. 143–150.

Johnson, Nan E and C Shannon Stokes (1976). "Family size in successive generations: The effects of birth order, intergenerational change in lifestyle, and familial satisfaction". *Demography* 13, pp. 175–187.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47.2, pp. 263–292.

Kenkel, Donald, Dean Lillard, and Alan Mathios (2006). "The Roles of High School Completion and GED Receipt in Smoking and Obesity". *Journal of Labor Economics* 24.3, pp. 635–660.

Kessler, Ronald C, Gavin Andrews, Lisa J Colpe, Eva Hiripi, Daniel K Mroczek, S-LT Normand, Ellen E Walters, and Alan M Zaslavsky (2002). "Short screening scales to monitor population prevalences and trends in non-specific psychological distress". *Psychological medicine* 32.6, pp. 959–976.

Kessler, Ronald C, Peggy R Barker, Lisa J Colpe, Joan F Epstein, Joseph C Gfroerer, Eva Hiripi, Mary J Howes, Sharon-Lise T Normand, Ronald W Manderscheid, Ellen

E Walters, et al. (2003). "Screening for serious mental illness in the general population". *Archives of general psychiatry* 60.2, pp. 184–189.

Kiessling, Lukas (2021). "How do parents perceive the returns to parenting styles and neighborhoods?" *European Economic Review* 139, p. 103906.

Kiessling, Lukas and Jonathan Norris (2022). "The Long-Run Effects of Peers on Mental Health". *The Economic Journal*, ueac039.

Kim, Giyeon, Jamie DeCoster, Ami N Bryant, and Katy L Ford (2016). "Measurement equivalence of the K6 scale: The effects of race/ethnicity and language". *Assessment* 23.6, pp. 758–768.

Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante (Aug. 2023). *The Child Penalty Atlas*. Working Paper 31649. National Bureau of Economic Research.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). "Children and gender inequality: Evidence from Denmark". *American Economic Journal: Applied Economics* 11.4, pp. 181–209.

Kolk, Martin (2014). "Multigenerational transmission of family size in contemporary Sweden". *Population Studies* 68.1, pp. 111–129.

Kristoffersen, Jannie Helene Grøne, Morten Visby Krægpøth, Helena Skyt Nielsen, and Marianne Simonsen (2015). "Disruptive school peers and student outcomes". *Economics of Education Review* 45, pp. 1–13.

Krueger, Alan B and Diane M Whitmore (2001). "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR". *The Economic Journal* 111.468, pp. 1–28.

Lao, Yehui and Suxu Lin (2022). "Comparison of educational performance between the only children and children in two-child families". *Scientific Reports* 12.1, p. 15355.

Lavy, Victor and Analia Schlosser (2011). "Mechanisms and impacts of gender peer effects at school". *American Economic Journal: Applied Economics* 3.2, pp. 1–33.

Lee, Jungmin (2008). “Sibling size and investment in children’s education: An Asian instrument”. *Journal of Population Economics* 21, pp. 855–875.

Lee, Soohyung, Lesley J Turner, Seokjin Woo, and Kyunghee Kim (2014). “All or nothing? The impact of school and classroom gender composition on effort and academic achievement”.

Li, Hongbin, Junsen Zhang, and Yi Zhu (2008). “The quantity-quality trade-off of children in a developing country: Identification using Chinese twins”. *Demography* 45, pp. 223–243.

Liang, Yun and John Gibson (2018). “Do siblings take your food away? Using China’s One-Child Policy to test for child quantity-quality trade-offs”. *China Economic Review* 48, pp. 14–26.

Lichand, Guilherme and Anandi Mani (2020). “Cognitive droughts”. *University of Zurich, Department of Economics, Working Paper* 341.

Liu, Haoming (2014). “The quality–quantity trade-off: evidence from the relaxation of China’s one-child policy”. *Journal of Population Economics* 27, pp. 565–602.

Liu, Ruth X, Wei Lin, and Zeng-yin Chen (2010). “The effect of parental responsiveness on differences in psychological distress and delinquency between singleton and non-singleton Chinese adolescents”. *Journal of child and family studies* 19, pp. 547–558.

Lo, Barbara Chuen Yee, Sin Kan Chan, Ting Kin Ng, and Anna Wai Man Choi (2020). “Parental demandingness and executive functioning in predicting anxiety among children in a longitudinal community study”. *Journal of Youth and Adolescence* 49, pp. 299–310.

Lordan, Grace and Paul Frijters (2013). “Unplanned pregnancy and the impact on sibling health outcomes”. *Health Economics* 22.8, pp. 903–914.

Lu, Hui Jing and Lei Chang (2013). "Parenting and socialization of only children in urban China: An example of authoritative parenting". *The Journal of genetic psychology* 174.3, pp. 335–343.

Luebbe, Aaron M, Kari A Bump, Lauren M Fussner, and Kathryn J Rulon (2014). "Perceived maternal and paternal psychological control: Relations to adolescent anxiety through deficits in emotion regulation". *Child Psychiatry & Human Development* 45, pp. 565–576.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger (July 2023). "The Value of Working Conditions in the United States and the Implications for the Structure of Wages". *American Economic Review* 113.7, pp. 2007–47.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2013). "Poverty impedes cognitive function". *science* 341.6149, pp. 976–980.

Manski, Charles F (2004). "Measuring expectations". *Econometrica* 72.5, pp. 1329–1376.

McCoby, EE (1983). "Socialization in the context of the family: Parent-child interaction". *Handbook of child psychology* 4, pp. 1–101.

Miller-Slough, Rachel L, Julie C Dunsmore, Janice L Zeman, Wesley M Sanders, and Jennifer A Poon (2018). "Maternal and paternal reactions to child sadness predict children's psychosocial outcomes: A family-centered approach". *Social Development* 27.3, pp. 495–509.

Millimet, Daniel L and Le Wang (2011). "Is the quantity-quality trade-off a trade-off for all, none, or some?" *Economic Development and Cultural Change* 60.1, pp. 155–195.

Moyano, Paloma (2017). "The Intergenerational Health Effects of the US Bombing Campaign in Cambodia". *Working Paper*.

Murphy, Michael and Lisbeth B Knudsen (2002). "The intergenerational transmission of fertility in contemporary Denmark: The effects of number of siblings (full and half), birth order, and whether male or female". *Population Studies* 56.3, pp. 235–248.

Murphy, Richard and Felix Weinhardt (2020). "Top of the class: The importance of ordinal rank". *The Review of Economic Studies* 87.6, pp. 2777–2826.

Nie, Peng and Yaohui Zhao (2023). "Ageing In China". In: *The Routledge Handbook of the Economics of Ageing*. Routledge, pp. 691–712.

Norris, Jonathan (2020). "Peers, parents, and attitudes about school". *Journal of Human Capital* 14.2, pp. 290–342.

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.

Papageorge, Nicholas W, Seth Gershenson, and Kyung Min Kang (2020). "Teacher expectations matter". *Review of Economics and Statistics* 102.2, pp. 234–251.

Payne, B Keith, Jazmin L Brown-Iannuzzi, and Jason W Hannay (2017). "Economic inequality increases risk taking". *Proceedings of the National Academy of Sciences* 114.18, pp. 4643–4648.

Pinquart, Martin (2017). "Associations of parenting dimensions and styles with internalizing symptoms in children and adolescents: A meta-analysis". *Marriage & Family Review* 53.7, pp. 613–640.

Ponzo, Michela and Vincenzo Scoppa (2022). "Human Capital Investments and Family Size in Italy: IV Estimates Using Twin Births as an Instrument". *IZA Discussion Paper*.

Prochaska, Judith J, Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong (2012). "Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization". *International journal of methods in psychiatric research* 21.2, pp. 88–97.

Qian, Nancy (2009). "Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China". *National Bureau of Economic Research.*

Qian, Yue and Yongai Jin (2024). "Women's fertility autonomy in urban China: The role of couple dynamics under the universal two-child policy". In: *Fertility and Childcare in East Asia.* Routledge, pp. 195–225.

Qin, Xuezheng, Castiel Chen Zhuang, and Rudai Yang (2017). "Does the one-child policy improve children's human capital in urban China? A regression discontinuity design". *Journal of Comparative Economics* 45.2, pp. 287–303.

Radloff, Lenore Sawyer (1977a). "The CES-D scale: A self-report depression scale for research in the general population". *Applied psychological measurement* 1.3, pp. 385–401.

— (1977b). "The CES-D scale: A self-report depression scale for research in the general population". *Applied psychological measurement* 1.3, pp. 385–401.

Ramey, Garey and Valerie A Ramey (2009). *The rug rat race.* Tech. rep. National Bureau of Economic Research.

Rao, Gautam (2019). "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools". *American Economic Review* 109.3, pp. 774–809.

Rao, Wen-Wang, Fan He, Yanjie Qi, Grace Ka In LOK, Todd Jackson, Yi Zheng, and Yu-Tao Xiang (2024). "Mental health among school children and adolescents in China: A comparison of one-child and multiple-children families from a nationwide survey". *Asian Journal of Psychiatry*, p. 104130.

Rapee, Ronald M (1997). "Potential role of childrearing practices in the development of anxiety and depression". *Clinical psychology review* 17.1, pp. 47–67.

Romano, Joseph P and Michael Wolf (2005). "Exact and approximate stepdown methods for multiple hypothesis testing". *Journal of the American Statistical Association* 100.469, pp. 94–108.

Rosenberg, Morris (1989). *Society and the adolescent self-image (Rev. ed.)* Wesleyan University Press.

Rosenzweig, Mark R and Junsen Zhang (2009). "Do population control policies induce more human capital investment? Twins, birth weight and China's "one-child" policy". *The Review of Economic Studies* 76.3, pp. 1149–1174.

Rothstein, Jesse (2017). "Measuring the impacts of teachers: Comment". *American Economic Review* 107.6, pp. 1656–84.

Sacerdote, Bruce (2014). "Experimental and quasi-experimental analysis of peer effects: two steps forward?" *Annu. Rev. Econ.* 6.1, pp. 253–272.

Santana Cooney, Rosemary, Jin Wei, and Mary G Powers (1991). "The one child certificate in Hebei province, China: Acceptance and consequence, 1979–1988". *Population Research and Policy Review* 10, pp. 137–155.

Scharping, Thomas (2013a). "Birth Control in China 1949-2000: Population policy and demographic development". *Routledge.*

— (2013b). *Birth Control in China 1949-2000: Population policy and demographic development.* Routledge.

Settles, Barbara H, Xuewen Sheng, Yuan Zang, and Jia Zhao (2012). "The one-child policy and its impact on Chinese families". In: *International Handbook of Chinese Families.* Springer, pp. 627–646.

Shi, Lihong (2017). *Choosing daughters: Family change in rural China.* Stanford University Press.

Short, Susan E, Zhai Fengying, Xu Siyuan, and Yang Mingliang (2001). "China's one-child policy and the care of children: An analysis of qualitative and quantitative data". *Social Forces* 79.3, pp. 913–943.

Soysa, Champika K and Andrea Weiss (2014). "Mediating perceived parenting styles–test anxiety relationships: Academic procrastination and maladaptive perfectionism". *Learning and Individual Differences* 34, pp. 77–85.

Strauss, John and Duncan Thomas (2007). "Health over the life course". *Handbook of Development Economics* 4, pp. 3375–3474.

Thapar, Anita, Stephan Collishaw, Daniel S Pine, and Ajay K Thapar (2012). "Depression in adolescence". *The lancet* 379.9820, pp. 1056–1067.

Tincani, Michela M (2018). "Heterogeneous peer effects in the classroom". *Manuscript, Dept. Econ., University College London.*

Tu, Mengwei (2016). "Chinese one-child families in the age of migration: middle-class transnational mobility, ageing parents, and the changing role of filial piety". *The Journal of Chinese Sociology* 3.1, p. 15.

Tversky, Amos and Daniel Kahneman (1991). "Loss Aversion in Riskless Choice: a Reference Dependent Model". *Quarterly Journal of Economics* 106.4, pp. 1039–1061.

U.S. Census Bureau (2022). *CPS Historical Time Series Tables.* https://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html.

Van Droogenbroeck, Filip, Bram Spruyt, and Gil Keppens (2018). "Gender differences in mental health problems among adolescents and the role of social support: results from the Belgian health interview surveys 2008 and 2013". *BMC psychiatry* 18, pp. 1–9.

Veeck, Ann, Laura Flurry, and Naihua Jiang (2003). "Equal dreams: The one child policy and the consumption of education in urban China". *Consumption, Markets and Culture* 6.1, pp. 81–94.

Wager, Stefan and Susan Athey (2018). “Estimation and inference of heterogeneous treatment effects using random forests”. *Journal of the American Statistical Association* 113.523, pp. 1228–1242.

Wang, Zhihe, Ming Yang, Jiaming Zhang, and Jiang Chang (2016). “Ending an era of population control in China: was the one-child policy ever needed?” *American Journal of Economics and Sociology* 75.4, pp. 929–979.

Wasserman, Melanie (July 2022). “Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents”. *The Review of Economic Studies* 90.3, pp. 1535–1568.

Weissman, Myrna M, Diane Sholomskas, Margaret Pottenger, Brigitte A Prusoff, and Ben Z Locke (1977). “Assessing depressive symptoms in five psychiatric populations: a validation study”. *American journal of epidemiology* 106.3, pp. 203–214.

White, Tyrene (1990). “Postrevolutionary mobilization in China: The one-child policy reconsidered”. *World Politics* 43.1, pp. 53–76.

Willis, Robert J (1973). “A new approach to the economic theory of fertility behavior”. *Journal of Political Economy* 81.2, Part 2, S14–S64.

Wiswall, Matthew and Basit Zafar (Aug. 2017). “Preference for the Workplace, Investment in Human Capital, and Gender*”. *The Quarterly Journal of Economics* 133.1, pp. 457–507.

— (2021). “Human capital investments and expectations about career and family”. *Journal of Political Economy* 129.5, pp. 1361–1424.

Wong, Daniel Fu Keung, Xiao Yu Zhuang, and Ting Kin Ng (2019). “Is parental control beneficial or harmful to the development of young children in Hong Kong?” *Journal of Child and Family Studies* 28, pp. 831–838.

Wu, Lingwei (2014). “Are only children worse off on subjective well-being?: Evidence from China’s One-Child Policy”. PhD thesis. Hong Kong University of Science and Technology Hong Kong.

Wu, Xiaoyu and Lixing Li (2012). “Family size and maternal health: evidence from the One-Child policy in China”. *Journal of Population Economics* 25, pp. 1341–1364.

Xie, Yu and Jingwei Hu (2014). “An introduction to the China family panel studies (CFPS)”. *Chinese sociological review* 47.1, pp. 3–29.

Xu, Shuhua, Xianyong Yin, Shilin Li, Wenfei Jin, Haiyi Lou, Ling Yang, Xiaohong Gong, Hongyan Wang, Yiping Shen, Xuedong Pan, et al. (2009). “Genomic dissection of population substructure of Han Chinese and its implication in association studies”. *The American Journal of Human Genetics* 85.6, pp. 762–774.

Yi, FAN (2016). “Intergenerational income persistence and transmission mechanism: Evidence from urban China”. *China Economic Review* 41, pp. 299–314.

Yuesheng, Wang (2014). “An Analysis of changes in the Chinese family structure between urban and rural areas: on the basis of the 2010 National Census Data”. *Social Sciences in China* 35.4, pp. 100–116.

Zeng, Shuxi, Fan Li, and Peng Ding (2020). “Is being an only child harmful to psychological health?: evidence from an instrumental variable analysis of China’s one-child policy”. *Journal of the Royal Statistical Society Series A: Statistics in Society* 183.4, pp. 1615–1635.

Zhang, Jie, Peng Xu, and Feng Liu (2020). “One-child policy and childhood obesity”. *China Economic Review* 59, p. 100938.

Zhang, Junsen (2017). “The evolution of China’s one-child policy and its effects on family outcomes”. *Journal of Economic Perspectives* 31.1, pp. 141–160.

Zhang, Y Jane (2019). “Culture, institutions and the gender gap in competitive inclination: Evidence from the communist experiment in China”. *The Economic Journal* 129.617, pp. 509–552.

Zhao, Liqiu and Zhong Zhao (2021). “Disruptive Peers in the Classroom and Students’ Academic Outcomes: Evidence and Mechanisms”. *Labour Economics* 68, p. 101954.

Zhong, Hai (2014). “The effect of sibling size on children’s health: a regression discontinuity design approach based on China’s one-child policy”. *China Economic Review* 31, pp. 156–165.

— (2017). “The effect of sibling size on children’s health and education: Is there a quantity-quality trade-off?” *The Journal of Development Studies* 53.8, pp. 1194–1206.

Zhu, Huoyun and Alan Walker (2021). “Population ageing and social policies in China: Challenges and opportunities”. *The Routledge Handbook of Chinese Studies*, pp. 191–204.

Zhu, Xiaodong (2012). “Understanding China’s growth: Past, present, and future”. *Journal of Economic Perspectives* 26.4, pp. 103–124.

Appendix A

Chapter 2 Appendix

A.1 Mathematical Proofs

Proof of Proposition 1. To begin with, it is useful to summarize the properties of the functional forms adopted in the model of Section 2.5, that is, $b(0) = 0$, $b'(y) > 0$, $b''(y) < 0$, and $\lim_{y\to\infty} b'(y) = 0$, $\lim_{y\to 0} b'(y) = \infty$; $c(0) = 0$, $c'(e) > 0$, $c''(e) > 0$, and $\lim_{e\to\infty} c'(e) = \infty$; and $\mu(0) = 0$, $\mu'(y-r) > 0$, $\mu''(y-r) < 0$ if $y > r$ (concavity over gains) and $\mu''(y-r) > 0$ if $y < r$ (convexity over losses), and $\lim_{y\to r} \mu'(y-r) = \infty$. All functions are continuous, and continuously differentiable, the only exception being μ which is not differentiable at $y = r$. Next, we proceed by analyzing the properties of the solution for the case in which $y > a$, denoted by $\tilde{e}(\theta, a)^+$ and then for the case in which $y < a$, denoted by $\tilde{e}(\theta, a)^-$.

Case of $y > r$. By definition, $\tilde{e}(\theta, r)^+$ is the level of effort at which the first-order condition given by (2.5) is satisfied. Since

$$u_{ee} = [b''(y) + \mu''(y-r)]\theta^2 - c'' < 0$$

where $\mu''(y-r) < 0$ when $y > r$, we conclude that u is strictly concave in e. This, together with the fact that as e gets smaller (so that y approaches r from above), $\lim_{y\to r} u_e = \infty$ due to the fact that $\lim_{y\to r} \mu'(y-r) = \infty$, and as e gets larger, $\lim_{e\to\infty} u_e = -\infty$ due to the fact that $\lim_{e\to\infty} c'(e) = \infty$, enables us to conclude that

$\tilde{e}(\theta, r)^{+}$ always exists, it is unique, and strictly positive. Moreover, note that

$$u_{er} = -\mu''(y - r)\theta > 0,$$

where $\mu''(y - r) < 0$ when $y > r$, enabling us to deduce by implicit differentiation that $\tilde{e}_r^{+} = -u_{er}/u_{ee} > 0$, implying that $\tilde{e}(\theta, r)^{+}$ is increasing in r.

Case of $y < a$. By definition $\tilde{e}(\theta, r)^{-}$ would be the level of effort at which the first-order condition given by (2.6) is satisfied. However,

$$u_{ee} = [b''(y) + \mu''(y - r)]\theta^2 - c'',$$

the sign of which remains ambiguous, since $\mu''(y-r) > 0$ when $y < r$. Hence, we cannot conclude whether u is concave or convex in e in the domain of losses. Nevertheless, we can deduce that the marginal benefit of effort have a U-shape form, since $\lim_{y\to 0} b' = \infty$, $\lim_{y\to a} \mu' = \infty$ and

$$u_{eee} = \Big\{[\alpha - 2][\alpha - 1]\alpha[\theta e]^{\alpha-3} + [\beta - 2][\beta - 1]\beta[r - \theta e]^{\beta-3}\Big\}\,\theta^3 > 0.$$

This imply that we cannot be sure that a solution exists, or that if it does, that it is unique. To proceed, we denote the value of r at which the slope of the left-hand side of (2.6) is equal to the slope of the right-hand side, by $\hat{r}$. Formally, this is the value of r at which $u_{ee} = 0$. This value exists, and it is unique since

$$u_{er} = -\mu''(y - r)\theta < 0$$

when $y < r$. That is, as we increase r, the left-hand side of (2.6) will cross the right-hand side, and $\hat{r}$ is the value of r at which these are tangent. This imply that if $r < \hat{r}$ then there is no solution for the case of $y < r$ and the solution is $\tilde{e}(\theta, r)^{+}$; while if $r > \hat{r}$ then there are two solutions, one at which $u_{ee} > 0$ and one at which $u_{ee} < 0$. If we had modelled variable marginal cost (e.g. $c'(e) = \phi e$, $\phi > 0$, rather than ϕ being normalised to one as in the model section), $u_{ee} > 0$ would imply that $\tilde{e}(\theta, r)^{-}$ is increasing in its marginal cost, while $u_{ee} < 0$ would imply that $\tilde{e}(\theta, r)^-$ is decreasing in its marginal cost, which is the one we consider. Hence, if $r < \hat{r}$ there is no solution in the loss domain, and the student will choose $\tilde{e}(\theta, r)^+$ (which always exists); while if $r \geq \hat{r}$ there always exist two local solutions: one such that $y > r$ and one such that $y < r$. In this case, we assume the student will choose the one that yields the higher utility, in line with the principle of utility maximization. To prove that $\tilde{e}(\theta, r)^-$ is decreasing in r we use the fact that $u_{er} = -\mu''(y-r)\theta < 0$ which follows from the fact that $\mu''(y-r) > 0$ when $y < r$. Hence, implicit differentiation yields $\tilde{e}_r^- = -u_{er}/u_{ee} < 0$, implying $\tilde{e}(\theta, r)^-$ is decreasing in r.

Next, we prove that r^* exists and that it is unique. For this, it is sufficient to consider a situation in which $r \in [0, \underline{r}]$ where $\underline{r} > \hat{r}$ and suppose that $y > r$ such that the solution is $\tilde{e}(\theta, r)^+$ and that $u(\tilde{e}(\theta, r)^+, \theta, r) > u(\tilde{e}(\theta, r)^-, \theta, r)$. Application of the envelope theorem implies that both $u(\tilde{e}(\theta, r)^+, \theta, r)$ and $u(\tilde{e}(\theta, r)^-, \theta, r)$ are decreasing in r, where (from the first order conditions (2.5) and (2.6))

$$\begin{aligned}\frac{du(\tilde{e}(\theta, r)^+, \theta, r)}{dr} &= -\beta[\theta\tilde{e}^+ - r]^{\beta-1} = \alpha[\theta\tilde{e}^+]^{\alpha-1} - \frac{\tilde{e}^+}{\theta} \\ &< \alpha[\theta\tilde{e}^-]^{\alpha-1} - \frac{\tilde{e}^-}{\theta} = -\beta[r - \theta\tilde{e}^-]^{\beta-1} = \frac{du(\tilde{e}(\theta, r)^-, \theta, r)}{dr},\end{aligned}$$

and where the inequality follows from the concavity of b and the fact that $\tilde{e}^+ - \tilde{e}^- > 0$ at a given r. This implies that as we increase r, $u(\tilde{e}(\theta, r)^+, \theta, r)$ decreases faster than $u(\tilde{e}(\theta, r)^-, \theta, r)$, and that there exists a value of r, denoted by $r^* \equiv r^*(\theta)$, at which

$$u(\tilde{e}(\theta, r)^+, \theta, r) - u(\tilde{e}(\theta, r)^-, \theta, r) = 0,$$

(and for which we assume the solution to be given by $\tilde{e}(\theta, r)^+$). Further note that if $r > r^*$ then it must be that $u(\tilde{e}(\theta, r)^+, \theta, r) < u(\tilde{e}(\theta, r)^-, \theta, r)$, implying that the solution switches from being $\tilde{e}(\theta, r)^+$ to $\tilde{e}(\theta, r)^-$ and $y < r$. Next, since $\tilde{e}(\theta, r)^-$ is decreasing in r it implies that as we increase r further beyond r^* then y remains below r. The same logic applies for all $r \in [0, r^*]$, since as we increase r, and $\tilde{e}(\theta, r)^+$ is increasing in r, then y remains above r. This implies that r^* is unique.

To conclude, we prove that r^* is increasing in θ. From the definition of r^* above, implicit differentiation yields:

$$\frac{dr^*(\theta)}{d\theta} = -\frac{\dfrac{du(\tilde{e}(\theta,r)^+,\theta,r)}{d\theta} - \dfrac{du(\tilde{e}(\theta,r)^-,\theta,r)}{d\theta}}{\dfrac{du(\tilde{e}(\theta,r)^+,\theta,r)}{dr} - \dfrac{du(\tilde{e}(\theta,r)^-,\theta,r)}{dr}} > 0,$$

since the results above imply that the denominator is negative, while application of the envelope theorem implies that

$$\begin{aligned}
\frac{du(\tilde{e}(\theta,r)^+,\theta,r)}{d\theta} &= \left\{\alpha[\theta\tilde{e}^+]^{\alpha-1} + \beta[\theta\tilde{e}^+ - r]^{\beta-1}\right\}\tilde{e}^+ \\
&= \frac{\tilde{e}^+}{\theta}\tilde{e}^+ \\
&> \frac{\tilde{e}^-}{\theta}\tilde{e}^- \\
&= \left\{\alpha[\theta\tilde{e}^-]^{\alpha-1} + \beta[r - \theta\tilde{e}^-]^{\beta-1}\right\}\tilde{e}^- = \frac{du(\tilde{e}(\theta,r)^-,\theta,r)}{d\theta}.
\end{aligned}$$

Hence, the numerator is positive, implying r^* is increasing in θ. ■

Proof of Proposition 2. This proof proceeds in two steps. First we show that the level of effort at which $y = r$, defined by $\bar{e}(\theta, r(\theta, F^\theta)) \equiv \frac{r(\theta,F^\theta)}{\theta}$, is decreasing in θ. Then we prove the existence and uniqueness of θ^*.

Consider $\bar{e}(\theta, r(\theta, F^\theta))$ for given capacity θ and distribution F^θ, our assumptions on r imply that, for $\theta_2 = \lambda\theta_1$ with $\lambda > 1$:

$$\begin{aligned}
\bar{e}(\theta_2, r(\theta_2, F^\theta)) &= \frac{r(\theta_2, F^\theta)}{\theta_2} \\
&< \frac{r(\theta_2, \lambda F^\theta)}{\theta_2} = \frac{r(\lambda\theta_1, \lambda F^\theta)}{\lambda\theta_1} = \frac{\lambda r(\theta_1, F^\theta)}{\lambda\theta_1} = \bar{e}(\theta_1, r(\theta_1, F^\theta)).
\end{aligned}$$

Hence $\bar{e}(\theta, r(\theta, F^\theta))$ decreasing in θ (and increasing in r). Next, from Proposition 1 we know that if $r < r^*$ then $\tilde{e}^+ > \bar{e}$, and if $r > r^*$ then $\tilde{e}^- < \bar{e}$. Hence, there exists a unique $\bar{e}^*(\theta, r^*(\theta))$ such that for all (θ, r) with $\bar{e}(\theta, r) > \bar{e}^*(\theta, r^*(\theta))$ then $r > r^*$ and $y < r$, while for all (θ, r) with $\bar{e}(\theta, r) < \bar{e}^*(\theta, r^*(\theta))$ then $r < r^*$ and $y > r$. This, along with the result established above that $\bar{e}(\theta, r(\theta, F^\theta))$ is decreasing in θ, can be used to deduce that if $\bar{e}(\theta, r(\theta, F^{\theta})) > \bar{e}^{*}(\theta, r^{*}(\theta))$ for some $\theta_1 \leq \theta^{*}$ so that $r(\theta_1, F^{\theta}) > r^{*}(\theta_1)$, then this will be the case for all $\theta < \theta_1$; while if $\bar{e}(\theta, r(\theta, F^{\theta})) < \bar{e}^{*}(\theta, r^{*}(\theta))$ for some $\theta_2 > \theta^{*}$ so that $r(\theta_1, F^{\theta}) < r^{*}(\theta_1)$, then this will be the case for all $\theta > \theta_2$. ■

Proof of Proposition 3. First note that if $G^I > F^I$, then $r(I, G^I) < r(I, F^I)$ for any given I, hence, the reference outcome decreases for all students. Next, consider low income students, for which initially $r(I, F^I) > r^{*}(I)$. If $r(I, G^I) < r(I, F^I)$ then $\tilde{e}(I, r(I, G^I))^{-} > \tilde{e}(I, r(I, F^I))^{-}$ by the results established in Proposition 1, since our assumption that G^I is such that $r(I_F^l, G^I) > r^{*}(I)$ ensures that this is true for any $r(I, G^I)$ with $I \leq I_F^l$: even the richest of the low income students remains frustrated. Then consider high income students, for which initially $r(I, F^I) < r^{*}(I)$. Since $r(I, G^I) < r(I, F^I)$ then $r(I, F^I) < r^{*}(I)$ and $\tilde{e}(I, r(I, G^I))^{+} < \tilde{e}(I, r(I, F^I))^{+}$ by the results established in Proposition 1. Finally, consider middle income students. There is a fraction of these students endowed with $I \in (I_F^{*}, I_F^{h})$ for which $r(I, G^I) < r^{*}(I)$, which implies they behave the same as high income students. However, there is also a fraction of these students endowed with $I \in (I_F^l, I_F^{*})$ whom will increase their effort only as long as the decrease in r is such that $r(I, G^I) > r^{*}(I)$; while they will decrease their effort if the decrease in r is such that $r(I, G^I) < r^{*}(I)$. ■

A.2 Additional Tables and Figures

Table A.1. Summary statistics

	Analytic Sample = 11,165				Full Sample		
	Mean	SD	Min	Max	Mean	Mean diff.	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Outcome and Treatment							
College Graduate in wave IV	0.33	0.47	0	1	0.32	0.01	0.01
Share of Low Income Peers (SLP_{-ics})	0.34	0.20	0	1	0.35	-0.01	0.00
B. Student Characteristics							
Logged Household Income	3.56	0.84	0	7	3.52	0.04	0.00
Female	0.52	0.50	0	1	0.51	0.01	0.00
Age	15.47	1.68	11	19	15.66	-0.19	0.00
Hispanic	0.15	0.35	0	1	0.17	-0.02	0.00
White	0.59	0.49	0	1	0.52	0.07	0.00
Black	0.20	0.40	0	1	0.22	-0.02	0.00
Asian	0.05	0.21	0	1	0.07	-0.02	0.00
Other Races	0.02	0.13	0	1	0.02	0.00	0.08
Family Size	3.79	1.21	2	12	3.77	0.02	0.19
Child of and Immigrant	0.17	0.38	0	1	0.22	-0.05	0.00
Less than HS Parents	0.10	0.30	0	1	0.13	-0.03	0.00
HS or GED Parents	0.29	0.46	0	1	0.30	-0.01	0.16
Some College Parents	0.22	0.42	0	1	0.21	0.01	0.02
College Parents	0.25	0.43	0	1	0.23	0.02	0.00
Postgraduate Parents	0.13	0.34	0	1	0.12	0.01	0.07
Single Parent Household	0.30	0.46	0	1	0.32	-0.02	0.00
Grade 7	0.14	0.35	0	1	0.13	0.01	0.00
Grade 8	0.14	0.35	0	1	0.13	0.01	0.00
Grade 9	0.19	0.39	0	1	0.17	0.02	0.00
Grade 10	0.20	0.40	0	1	0.19	0.01	0.21
Grade 11	0.18	0.38	0	1	0.18	0.00	0.46
Grade 12	0.15	0.35	0	1	0.16	-0.01	0.00

Notes: Column (1) - (4) in this table present summary statistics for the sample in wave I of AddHealth after restricting to our analytic sample but before imputing the sample, which has 11, 165 observations left. Column (5) presents the mean of full sample available from the dataset. Each variable has around 20,000 observations in the full sample. Column (6) shows the difference in means and column (7) presents the p-values from the mean-comparison tests.

Table A.2. Additional summary statistics

	Mean	SD	Min	Max
A. GPA and Advanced Courses Taking				
Self-reported GPA at wave I	2.80	0.77	1	4
Transcript average GPA after treatment	2.44	0.89	0	4
Advanced Math courses taking	0.41	0.49	0	1
Advanced Science courses taking	0.46	0.50	0	1
Advanced English couses taking	0.24	0.43	0	1
Taking more than one advanced couses	0.60	0.49	0	1
B. Frustration and Motivation				
Self esteem	28.56	4.14	7	35
Intelligent feelings compared to others	3.90	1.08	1	6
CES-D mental health scale	11.02	7.46	0	54
Motivation	3.78	0.91	1	5
Observations	11165			

Notes: This table presents additional summary statistics on GPA and advanced courses taking in Table 2.3 and frustration and motivation measures in Table 2.4 after restricting to our analytic sample.

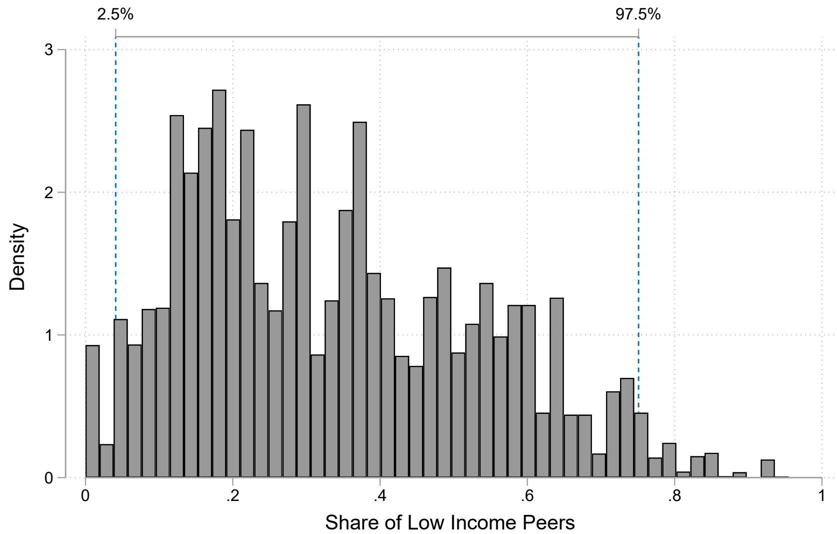
Table A.3. Frustration and motivation variables

	Original questions	Answer modalities	Final indicator
Self-Esteem	1. You have a lot of good qualities. 2. You are physically fit. 3. You have a lot to be proud of. 4. You like yourself just the way you are. 5. You feel like you are doing everything just about right. 6. You feel socially accepted. 7. You feel loved and wanted.	1. strongly agree 2. agree 3. neither agree nor disagree 4. disagree 5. strongly disagree	We reverse code the raw variables, then aggregate those 7 variables to get the self-esteem variable. Higher values imply higher self-esteem.
Intelligent Feeling	Compared with other people your age, how Intelligent are you?	1. moderately below average 2. slightly below average 3. about average 4. slightly above average 5. moderately above average 6. extremely above average	
CES-D scale	1. You were bothered by things that don't usually bother you. 2. You didn't feel like eating, your appetite was poor. 3. You felt that you could not shake off the blues, even with help from your family and your friends. 4. You felt you were just as good as other people. 5. You had trouble keeping your mind on what you were doing. 6. You felt depressed. 7. You felt that you were too tired to do things. 8. You felt hopeful about the future. 9. You thought your life had been a failure. 10. You felt fearful. 11. You were happy. 12. You talked less than usual. 13. You felt lonely. 14. People were unfriendly to you. 15. You enjoyed life. 16. You felt sad. 17. You felt that people disliked you. 18. It was hard to get started doing things. 19. You felt life was not worth living.	0. never or rarely 1. sometimes 2. a lot of the time 3. most of the time or all the time	We reverse code items 4, 8, 11, 15 and aggregate those 19 variables to get a final score ranging from 0 to 57, which higher scores indicating a higher propensity for despressive symptoms.
Motivation	1. During the 1994-1995 school year, how often did you have trouble paying attention in school? 2. During the 1994-1995 school year, how often did you have trouble getting your homework done?	0. never 1. just a few times 2. about once a week 3. almost everyday 4. everyday	We reverse code these raw variables and take the mean. Higher values imply less trouble/higher motivation.

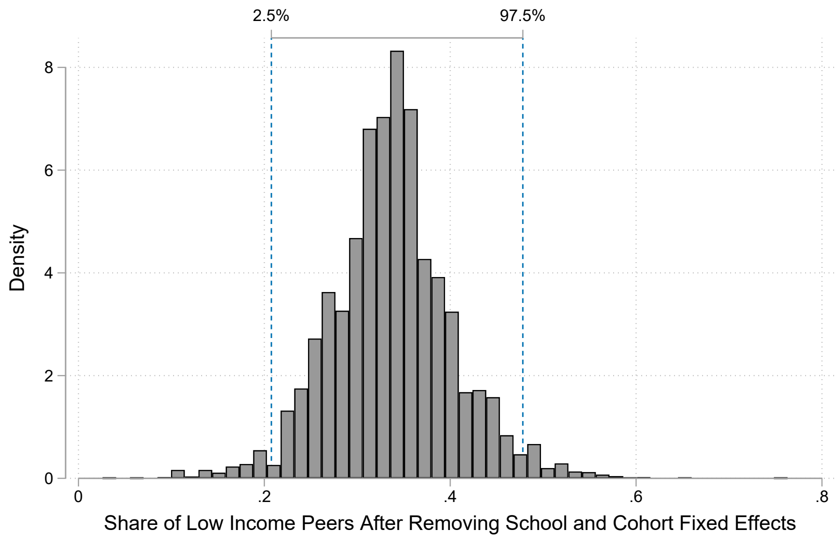
Notes: This table presents details of the construction of the frustration and motivation variables in Table 2.4.

Figure A.1. Variation in share of low-income peers

(a) Raw variation



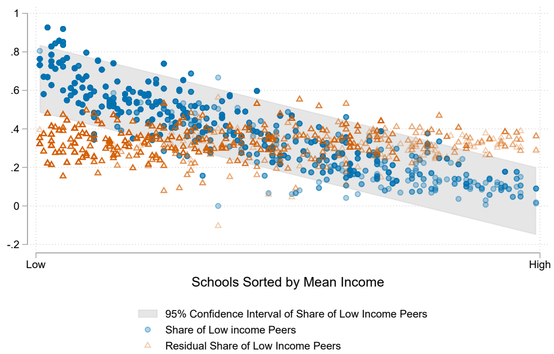
(b) Variation post removal of school and cohort fixed effects



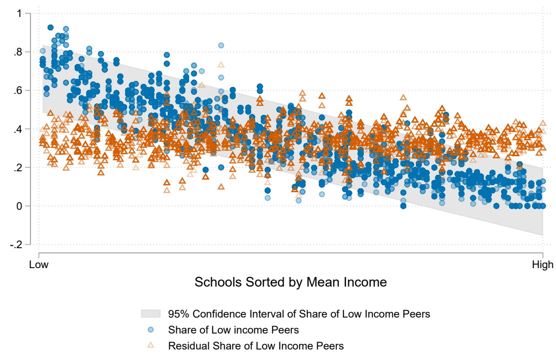
Notes: This figure presents a histogram of the share of low-income peers in our analytic sample. Panel (a) reports the variations in the sample, and panel (b) reports this variation after removal of school and cohort fixed effects with the sample mean added back to place it on the same scale as panel (a). Vertical lines denote the 2.5 and 97.5 percentiles.

Figure A.2. Variation between the share of low-income peers and school quality heterogeneous to own income groups conditional on school fixed effects

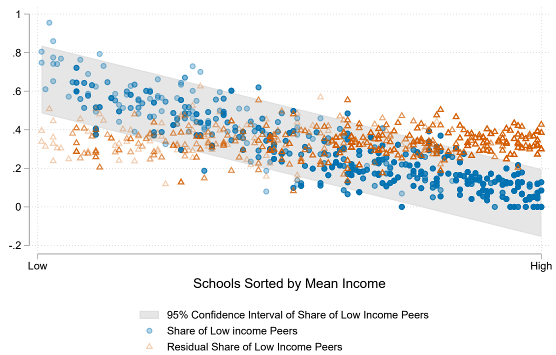
(a) Bottom 20



(b) Middle



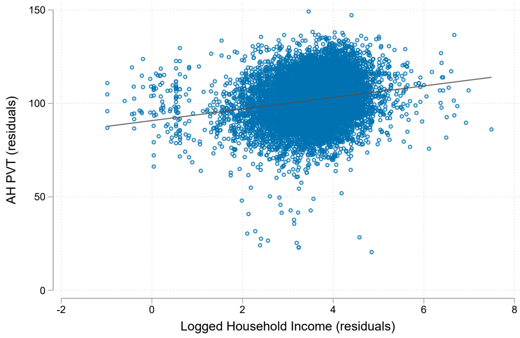
(c) Top 20



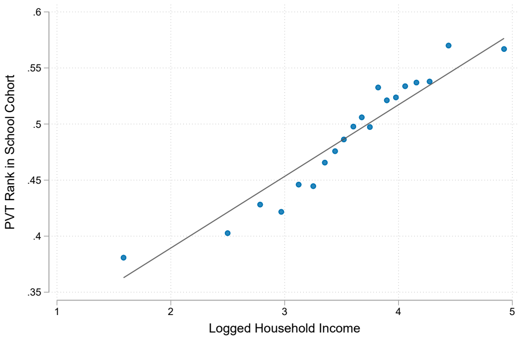
Notes: These figures present the share of low-income peers and its residual after removal of school fixed effects with the sample mean added back to it for the bottom 20th, middle, and top 20th of the household income distribution by schools. Schools are sorted based on the mean logged household income of students from the lowest to the highest.

Figure A.3. Associations: PVT scores, rank, and household income

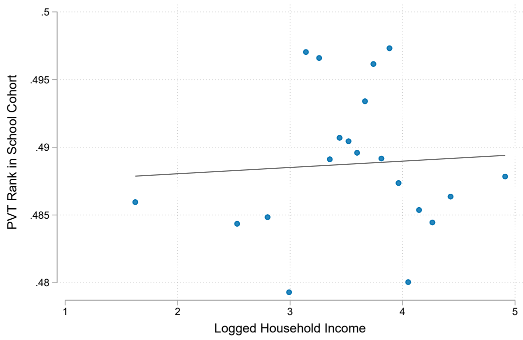
(a) PVT and ln(Income)



(b) PVT Rank and ln(Income)

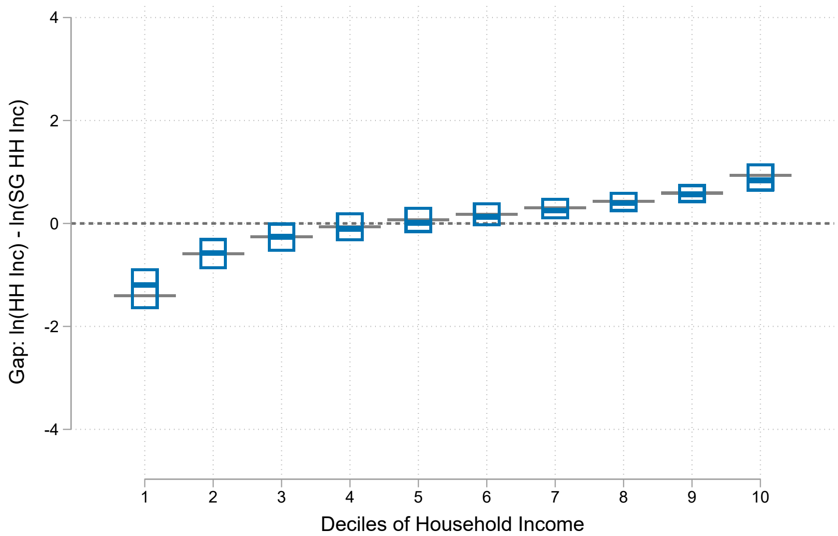


(c) PVT Rank and ln(Income): Control for PVT



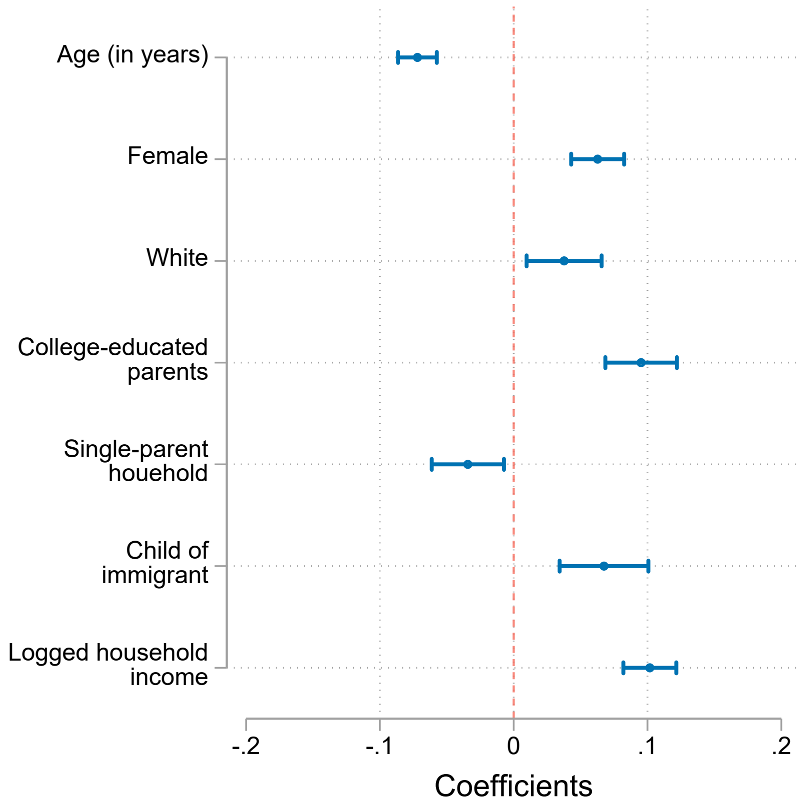
Notes: In all panels, we control for school fixed effects so associations are based on within school variation. Panel (a) reports a scatter plot and line of best fit between the residuals of the picture vocabulary test (PVT) scores and logged household income after removing school fixed effects. We add the full mean back to place the plot on the scale of the original variables. Panel (b) reports a bin scatter plot between the percentilized PVT school cohort rank based on the PVT scores and logged household income. Panel (c) reports the same as (b) but we control additionally for students' PVT scores.

Figure A.4. Gap between individual and school-cohort peer mean of logged household income



Notes: For each household income decile, this figure presents box plots of the interquartile range overlaid with lines for the mean and median.

Figure A.5. Associations of covariates with university completion



Notes: This figure presents a linear specification for logged household income and other characteristics. The base race in our specification is white, and we control for school and cohort fixed effects.

Table A.4. Long-run effects on labour market outcomes

	Wave IV Log Individual Income			
	(1)	(2)	(3)	(4)
SLP_{-ics} × Bottom 20	0.33	0.89***	0.79**	0.67*
	(0.25)	(0.29)	(0.38)	(0.39)
SLP_{-ics} × Middle	0.24	0.37*	0.33	0.30
	(0.15)	(0.21)	(0.30)	(0.19)
SLP_{-ics} × Top 20	-0.05	0.06	0.08	0.05
	(0.24)	(0.30)	(0.37)	(0.41)
School-specific Cohort Trends	No	No	Yes	No
School-specific Income Trends	No	No	No	Yes
Wave IV Sampling Weight	No	Yes	Yes	Yes
Mean Log Income	10.18	10.16	10.16	10.16
Observations	9919	9614	9614	9614
R^2	0.115	0.171	0.186	0.197

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. We trim our data to our analytic sample as in Table 2.2 and use Wave IV log household income as the long-run labor market outcome variable. We use Wave IV sampling weight to adjust the attrition in column (2) - (4). The sample weight was computed by the attrition for selecting schools and adolescents, as well as characteristics related to non-response. We further add school-specific cohort trends in column (3) and school-specific income trends in column (4). The result is consistent once we relax the sample size to the fully available sample in Table A.17.

A.3 Robustness Checks

Table A.5. Robustness to different definitions for the share of low-income peers

	SLP_{-ics} × Bottom 20	SLP_{-ics} × Middle	SLP_{-ics} × Top 20
	(1)	(2)	(3)
Original	0.18**	0.02	-0.25**
	(0.07)	(0.07)	(0.11)
Bottom 20th Percentile	0.22***	-0.01	-0.32**
	(0.08)	(0.07)	(0.16)
Below Median	0.13**	0.03	-0.09
	(0.06)	(0.05)	(0.09)
By School Region and Family Size	0.18***	-0.00	-0.19*
	(0.06)	(0.06)	(0.11)

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. The first row shows the results of our original definition of the share of low-income peers. In the second row, we define the share of low-income peers as the share of peers in the bottom 20th percentile of household size for a given family size. In the third row, we define the share of low-income peers as the share of peers below the median of household income for a given family size. In the fourth row, we define the share of low-income peers as share of peers in the bottom 3rd of the household income distribution by school region, school urbanicity, and a family size indicator (whether the family size is larger than 4). Observations are equal to 11,165 as our analytic sample size in each specification.

Table A.6. Robustness to non-linearity in household income

	Iterations of LnHHInc Polynomials				Ventiles
	(1)	(2)	(3)	(4)	(5)
SLP_{-ics} × Bottom 20	0.18**	0.17**	0.16**	0.16**	0.16**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
SLP_{-ics} × Middle	0.01	0.02	0.02	0.02	0.03
	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
SLP_{-ics} × Top 20	-0.25**	-0.25**	-0.26**	-0.26**	-0.26**
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
$(\text{LnHHInc})^3$	-0.01***	0.01	0.10**	0.02	
	(0.00)	(0.01)	(0.05)	(0.17)	
$(\text{LnHHInc})^4$		-0.00**	-0.02**	0.00	
		(0.00)	(0.01)	(0.05)	
$(\text{LnHHInc})^5$			0.00*	-0.00	
			(0.00)	(0.01)	
$(\text{LnHHInc})^6$				0.00	
				(0.00)	
H.H. Income Ventiles	No	No	No	No	Yes
Observations	11165	11165	11165	11165	11165

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. Column (5) includes household income ventiles to control for non-linearity.

Table A.7. Subsample analysis

	University Graduate					
	(1)	(2)	(3)	(4)	(5)	(6)
SLP_{-ics} × Bottom 20	0.27**			0.23*		
	(0.14)			(0.13)		
SLP_{-ics} × Middle		-0.03			-0.02	
		(0.08)			(0.08)	
SLP_{-ics} × Top 20			-0.34			-0.39*
			(0.21)			(0.21)
Own-Ability Polynomials	No	No	No	Yes	Yes	Yes
School-Cohort Ability Rank	No	No	No	Yes	Yes	Yes
Observations	2180	6920	2065	2180	6920	2065

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Columns (1) - (3) include all controls as in our preferred baseline specification in column (2) of Table 2.2. Columns (4) - (6) add additional controls as in our specification in column (4) of Table 2.2.

Table A.8. Placebo test

	Placebo treatment		Placebo outcome	
	(1)	(2)	(3)	(4)
SLP_{-ics} × Bottom 20	0.08	-0.08	-0.03	-0.11
	(0.06)	(0.10)	(0.06)	(0.12)
SLP_{-ics} × Middle	-0.04	-0.05	-0.00	0.01
	(0.05)	(0.05)	(0.05)	(0.05)
SLP_{-ics} × Top 20	-0.20**	0.03	-0.07	0.05
	(0.09)	(0.11)	(0.06)	(0.09)
School-specific Income Trends	No	Yes	No	Yes
Observations	11047	11047	11149	11149

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. Columns (1) - (2) estimate the effects of the placebo share of low-income peers on the probability of graduating from university. The placebo share of low-income peers is defined using the share of low-income peers in another cohort within the same school. Columns (3) - (4) estimate the effects of actual share of low-income peers on the placebo outcome, which is an indicator of ever repeated a cohort. Column (2) and column (4) add the school-specific income trends to the baseline specification.

Table A.9. Attrition analysis and sampling weights

	Attrited in Wave IV				University Graduate IPW Adjusted		University Graduate Weighted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of Low Income Peers	-0.05	0.07					
	(0.04)	(0.06)					
SLP_{-ics} × Bottom 20			-0.08	0.05	0.19***	0.23**	0.26***
			(0.06)	(0.07)	(0.07)	(0.10)	(0.08)
SLP_{-ics} × Middle			-0.05	0.08	0.03	-0.01	0.04
			(0.05)	(0.07)	(0.06)	(0.07)	(0.07)
SLP_{-ics} × Top 20			-0.05	0.10	-0.23**	-0.27**	-0.26*
			(0.06)	(0.09)	(0.11)	(0.13)	(0.14)
School and Grade Fixed Effects	No	Yes	No	Yes	Yes	Yes	Yes
School-specific Income Trends	No	No	No	No	No	Yes	No
Share Attrited	.22	.22	.22	.22	.22	.22	.22
Observations	14339	14339	14339	14339	11115	11115	10818
R^2	.026	.049	.027	.05	.24	.25	.27

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. The dependent variable in columns (1) - (4) is an indicator equal to one if an individual has attrited in wave IV and zero otherwise. Estimates of marginal effects are for the share of low-income peers in the bottom 20th percentile of household income, for the middle, and finally for the top 20th percentile of household income. In columns (5) - (6), we calculate treatment effects of the share of low-income peers on the probability of graduating from university using inverse probability weighting, where the weights are calculated as the predicted probability of being in wave IV follow-up sample based on the available baseline controls as in column (2) of Table 2.2. We further add the school-specific income trends to the baseline specification in column (6). We use Wave IV sampling weight designed for estimating single-level models to adjust the attrition in column (7). The sample weight was computed by the attrition for selecting schools and adolescents, as well as characteristics related to non-response.

Simulations to assess measurement errors. We present simulations to assess the role of two forms of measurement error. We assume the following data generating process (DGP):

$$Y_{is} = 0.18SLP_{is} \times B20_i - 0.25SLP_{is} \times T20_i + 0.01ln(Inc_i)$$

where Y denotes our outcome, SLP_{is} denotes the leave-one-out percentage share of peers from low-income households in their school cohort defined by the bottom third of the simulated income distribution, and Inc_i denotes a student's household income, which is randomly drawn from a log-normal distribution with the log-income mean of 3.5 and standard deviation of 0.85 ($ln(Inc) \sim N(3.5, 0.85)$), consistent with our analytical data. The indicator variables $B20_i$ and $T20_i$ flag observations in the bottom and top 20th deciles of the simulated income distribution. For the simulations, we use variation across schools abstracting away from multiple cohorts in each school. However, we model no selection effects into schools, thus variation across schools in our simulations is exogenous conditional on income. The parameters in the DGP ($\beta_1 = 0.18$, $\beta_2 = 0$, and $\beta_3 = -0.25$) are based on the specification shown in column (1) of Table 2.2.

First, we assess the consequences of observing a random subsample of students per school using Monte Carlo simulations. Based on our DGP we run 1000 repetitions with 500 schools/cohorts each and 240 students per school, re-drawing Inc_i and SLP_{is} at each repetition. We evaluate the simulated data with the following specification:

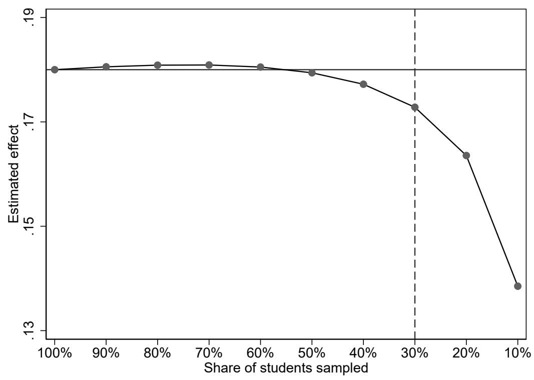
$$y = \beta_1 SLP_{ics} \times B20_i + \beta_2 SLP_{ics} \times Mid_i + \beta_3 SLP_{ics} \times T20_i + \gamma ln(Inc_i).$$

Additionally, we also evaluate it based on subgroups by income ($B20_i, Mid_i, T20_i$).

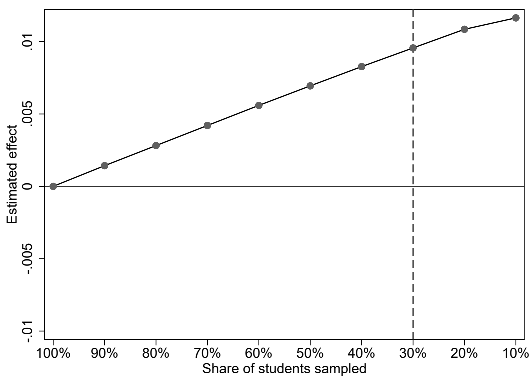
Second, we consider measurement error in our measure of income. Based on the same DGP but using mismeasured income we run the same regression as above at each repetition and combine this with the sampling error running a 100% sample and a 30% of the school sample. Results from the simulations are presented below.

Figure A.6. Simulations to assess bias due to random sampling within schools

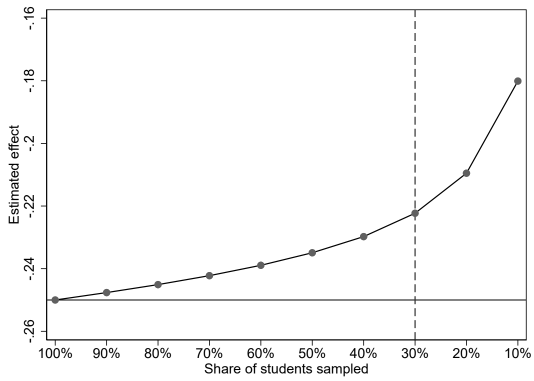
(a) Bottom 20 group



(b) Middle group



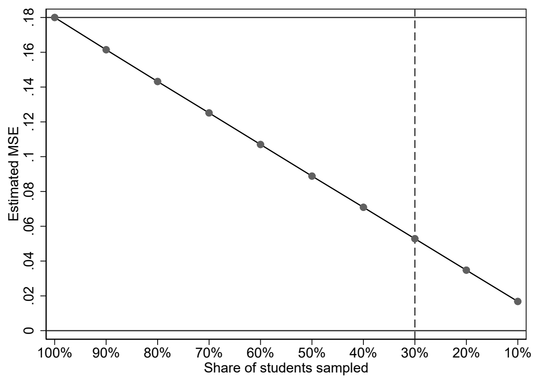
(c) Top 20 group



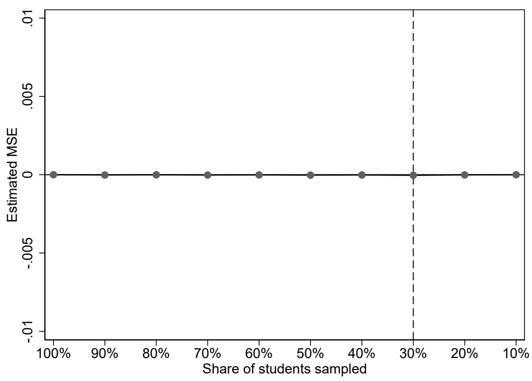
Notes: These figures present results from Monte Carlo simulations with 1000 repetitions of 500 schools for the bottom 20, middle, and top 20 groups respectively. The vertical dashed line of 30% is the average percentage of an Add Health school that was sampled.

Figure A.7. Simulations to random sampling within schools: subsample analysis

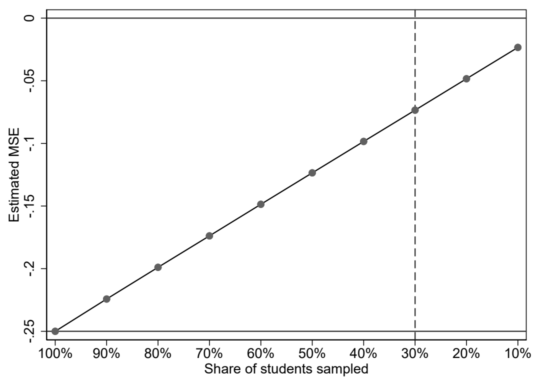
(a) Bottom 20 group



(b) Middle group



(c) Top 20 group



Notes: These figures present subsample analysis results from Monte Carlo simulations with 1000 repetitions of 500 schools for the bottom 20, middle, and top 20 groups respectively. The vertical dashed line of 30% is the average percentage of an Add Health school that was sampled.

Table A.10. Simulations to assess bias due to measurement error in income

Simulation: Measurement error in income

DGP: $Y_{is} = 0.18SLP_{-is} \times B20_i - 0.25SLP_{-is} \times T20_i + 0.01ln(Inc_i)$;

$ln(Inc_i) \sim \mathcal{N}(3.5, 0.85)$;

$\widetilde{ln(Inc_i)} \sim ln(Inc_i) + \phi \cdot v_i$; $\phi \in [0,1]$; $v_i \sim \mathcal{N}(0, 0.85)$;

Estimate: $Y_{is} = \beta_1\widetilde{SLP_{-is}} \times B20_i + \beta_2\widetilde{SLP_{-is}} \times Mid_i + \beta_3\widetilde{SLP_{-is}} \times T20_i + \gamma\widetilde{ln(Inc_i)}$

	Measurement error (ϕ)					
100% sampling	0	0.2	0.4	0.6	0.8	1.0
$SLP_{-ics} \times B20$	0.18	0.09	0.04	0.01	0.00	-0.00
	(100%)	*(52%)*	*(22%)*	*(6%)*	*(0%)*	*(-2%)*
$SLP_{-ics} \times Mid$	0.00	-0.00	-0.00	-0.00	-0.00	-0.00
$SLP_{-ics} \times T20$	-0.25	-0.15	-0.09	-0.05	-0.03	-0.02
	(100%)	*(61%)*	*(35%)*	*(20%)*	*(13%)*	*(9%)*
30% sampling						
$SLP_{-ics} \times B20$	0.17	0.09	0.04	0.02	0.00	0.00
	(94%)	*(51%)*	*(23%)*	*(9%)*	*(3%)*	*(0%)*
$SLP_{-ics} \times Mid$	0.01	0.01	0.01	0.00	0.00	0.00
$SLP_{-ics} \times T20$	-0.22	-0.13	-0.07	-0.04	-0.03	-0.02
	(88%)	*(53%)*	*(30%)*	*(17%)*	*(10%)*	*(7%)*

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process. For the middle group, the ratio is not reported because the true coefficient is 0.

A.4 Romano-Wolf p-value Adjustment

Table A.11. Romano-Wolf p-value adjustment for university graduation

	University Graduate					
	(1)	(2)	(3)	(4)	(5)	(6)
$SLP_{-ics} \times$ *Bottom 20*						
Original p-value	0.015	0.012	0.013	0.006	0.004	0.027
Romano-Wolf p-value	0.028	0.026	0.026	0.010	0.010	0.044
$SLP_{-ics} \times$ *Middle*						
Original p-value	0.854	0.810	0.922	0.986	0.396	0.783
Romano-Wolf p-value	0.948	0.926	0.948	0.982	0.521	0.926
$SLP_{-ics} \times$ *Top 20*						
Original p-value	0.030	0.028	0.017	0.014	0.139	0.028
Romano-Wolf p-value	0.052	0.052	0.028	0.028	0.190	0.052

Notes: We use Romano and Wolf's step-down adjusted p-values to conduct multiple hypothesis testing (Clarke et al., 2020; Romano and Wolf, 2005) across specifications. This table provides p-values after controlling for the family-wise error rate. The specifications match specifications in our baseline Table 2.2.

Table A.12. Romano-Wolf p-value adjustment for GPA and advanced courses

	GPA		Advanced Courses			
	Self	Transcript	Math	Science	English	More than one
	(1)	(2)	(3)	(4)	(5)	(6)
$SLP_{-ics} \times$ *Bottom 20*						
Original p-value	0.719	0.001	0.008	0.130	0.552	0.006
Romano-Wolf p-value	0.998	0.026	0.070	0.535	0.978	0.054
$SLP_{-ics} \times$ *Middle*						
Original p-value	0.553	0.018	0.522	0.928	0.821	0.337
Romano-Wolf p-value	0.978	0.122	0.978	1.000	1.000	0.884
$SLP_{-ics} \times$ *Top 20*						
Original p-value	0.304	0.891	0.494	0.089	0.356	0.994
Romano-Wolf p-value	0.858	1.000	0.968	0.413	0.892	1.000

Notes: We use Romano and Wolf's step-down adjusted p-values to conduct multiple hypothesis testing (Clarke et al., 2020; Romano and Wolf, 2005) on different outcomes. This table provides p-values after controlling for the family-wise error rate.

A.5 Mechanisms and Additional Results

A.5.1 Results Explained by Alternative Mechanisms?

We now describe our analysis testing whether our results on the share of low income peers capture dimensions related to peer ability, responses by teachers, disruptive peer behaviour, and responses by parents.

Non-linearity in Peer Ability

One possibility is that our results are explained by non-linear effects from the peer ability composition (Booij et al., 2017; Duflo et al., 2011; Feld and Zölitz, 2017). This literature suggests that non-linear peer ability effects may stem from changes in teaching practices that are more or less conducive to different ability groups. Alternatively, it also points out factors directly related to peer interactions – helping studying, inducing more effort, better information, etc. – that can generate differential responses to peer ability. Generally, the evidence suggests that students do not benefit from mixing by ability, implying that tracking by ability can be optimal. Our results on the share of low-income peers could be explained by this type of mechanism given the correlation between family income and ability. However, here we find no evidence for this.

In Table A.13, we control for nonlinear peer ability effects in several ways. We begin, in column (2), by adding to our preferred specification peer mean ability – based on PVT scores – and the standard deviation of peer ability interacted with own-income positions. Next, in column (3), we introduce peer ability heterogeneity around own-ability by adding interaction terms of peer mean ability, peer SD ability, and own-ability. Going further, in columns (4) - (5), we consider interactions of quartiles of a school's position in the school mean ability distribution and likewise for the school's position in the SD ability distribution. This is motivated by suggestions in Denning et al. (2021) aimed at capturing more effectively potential non-linear effects from reactions to the distribution of ability in the school. Across all of these specifications our estimates on the share of low-income peers remain remarkably consistent with our baseline estimates.

Finally, ability rank effects are known to exist separately from standard ability

Table A.13. Accounting for non-linearity in peer ability

	University Graduate						
	Non-linear peer ability effect					Rank effect	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SLP_{-ics} × Bottom 20	0.18**	0.18***	0.17**	0.17**	0.17**	0.22***	0.23***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
SLP_{-ics} × Middle	0.02	0.00	0.01	0.02	0.02	-0.00	-0.01
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
SLP_{-ics} × Top 20	-0.25**	-0.28**	-0.27**	-0.26**	-0.26**	-0.27**	-0.29**
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)
Peer Effects (means)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Effects (SD)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Own-Ability Polynomials	No	Yes	Yes	Yes	Yes	No	Yes
Peer Ability (means & SD) × Income Position	No	Yes	No	No	No	No	No
Peer Ability (means) × Peer Ability (SD) × Own-Ability	No	No	Yes	No	No	No	No
School Ability Quartiles (means) × Own-Ability	No	No	No	Yes	Yes	No	No
School Ability Quartiles (SD) × Own-Ability	No	No	No	No	Yes	No	No
Income Rank × Income Position	No	No	No	No	No	Yes	Yes
Ability Rank × Income Position	No	No	No	No	No	No	Yes
Observations	11165	11165	11165	11165	11165	11164	11164
R^2	0.243	0.263	0.263	0.264	0.264	0.243	0.264

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2, which is presented in column (1) of this table. Standard errors are in parentheses and clustered at the school level. School ability quartiles (means) are the quartiles of schools based on the school-level peer mean ability. School ability quartiles (SD) denote the quartiles of schools based on the school-level standard deviations of peer ability. Income rank denotes the rank of household income within school cohorts while ability rank denotes the rank of ability within school cohorts.

effects possibly from a social comparisons or a learning about ability mechanism (Elsner and Isphording, 2017; Kiessling and Norris, 2022). Thus, we expand our specification to account for ability ranks. While we have already flexibly allowed for ability rank effects in Table 2.2, we re-consider ranking concerns by allowing for both ability and income rank effects disaggregated across the income distribution. As is shown in columns (6) - (7) of the Table A.13, our results are not sensitive to ability nor income rank effects. Thus, our main results on the share of low-income peers appear distinct from, and insensitive to, both non-linear peer ability and rank effects.

Teachers

Responses by teachers that correlate with changes in the share of low-income peers could explain our results. As mentioned above, the literature on peer effects in education shows that teachers do change their behavior in response to classroom composition aimed at more effectively meeting students' needs (Aucejo et al., 2022; Duflo et al., 2011; Jackson, 2016; Lee et al., 2014; Papageorge et al., 2020). In this case, we would expect that as the share of low-income peers increases in a given school cohort, teachers may decide to devote more attention to them and also adapt their expectations and teaching practices accordingly. This will benefit low-income students, providing an explanation for our evidence on the bottom-20 students. However, the impact on middle or high-income students is somewhat ambiguous, as it will depend on whether the attention shift to low-income students comes at their expense or not. Moreover, predictions here for low-income students are not entirely clear. Alternatively, if teachers hold implicit stereotypes regarding different income groups, this may obstruct their interaction with students, acting to harm low-income students (Carlana, 2019; Carlana et al., 2022b).

Table A.14. Teachers effects: share of low-income peers

	Relationship with Teachers				University Graduation	
	Care Teachers	Close Teachers	Fair Teachers	Teacher Scale	Tracking	No Tracking
	(1)	(2)	(3)	(4)	(5)	(6)
SLP_{-ics} × Bottom 20	-0.01	-0.33	-0.06	-0.19	0.20*	0.15
	(0.22)	(0.20)	(0.18)	(0.20)	(0.11)	(0.09)
SLP_{-ics} × Middle	-0.11	-0.14	-0.07	-0.14	0.02	0.04
	(0.17)	(0.18)	(0.18)	(0.19)	(0.11)	(0.07)
SLP_{-ics} × Top 20	0.21	-0.08	0.00	0.05	-0.30*	-0.08
	(0.24)	(0.21)	(0.21)	(0.22)	(0.17)	(0.12)
Observations	11110	11164	11162	11165	6755	4265
R^2	0.068	0.074	0.055	0.066	0.227	0.254

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. Standard errors are in parentheses and clustered at the school level. The relationship with teacher variables are standardized. Columns (6) - (7) return to University graduation as the outcome but the sample is stratified by schools who report they do (or do not) to use ability tracking for English and Language Arts. Note that ability tracking was reported in the school principle's questionnaire and only asked on this dimension.

We next look at further evidence for a teacher mechanism to explain our results. In the U.S. educational context, students typically change classrooms throughout the day as they switch between classes and do not necessarily stay with the same classmates. Thereby, we would expect a teacher driven mechanism for our effects to be dominant only if all, or the significant share, of the teachers in the same school-cohort update their behavior at the same time. This may translate into average shifts in teacher-student relationships heterogeneous to the income distribution, so we look at student-reported measures of these relationships.[1] The results, in the Table A.14 columns (1) - (4), suggest there no effects here.

Finally, we look at our baseline model for University graduation split by schools who use ability tracking for English and Language Arts.[2] The effects should disappear in schools that track by ability if increases in the share of low-income peers mainly captures optimization of instruction for low-income students. Our results in columns (5) - (6) of the Table A.14 are not consistent with this for low-income students, and while the point estimates are inefficient, suggest similar results across school types. The results for high-income students, however, suggest they are mainly present in tracking schools, thus there is likely some role for the optimization story, albeit not enough to explain the overall pattern we observe.

Disruptive peers

Another possibility is that an increase in the share of low-income peers also picks up a shift in disruptive behavior. Disruptive behavior causes harm to academic achievement both in the short and the long run (Billings and Hoekstra, 2023; Carrell et al., 2018; Carrell and Hoekstra, 2010; Kristoffersen et al., 2015; Zhao and Zhao, 2021). In this case, we would expect a negative effect of our peer treatment on educational attainment at each point of the income distribution (see evidence in Carrell et al. (2018) and Carrell and Hoekstra (2010)).[3] Yet, in light of our baseline results, predictions based on the effect of an increase in disruptive behavior would only be able to explain our negative estimate on high-income students.

[1] We focus on four items that relate to these interactions from the student self-reported questionnaire at wave I: whether teachers care about students, whether students have trouble getting along with teachers, whether teachers treat students fairly, and a mean scale of the above three items. Higher scores in these outcomes reflect better teacher-student interactions.

[2] Ability tracking is reported by school principles at a school wide level. Ability tracking is not asked for other dimensions.

[3] Carrell and Hoekstra (2010), and Carrell et al. (2018) are the only two studies we are aware of evaluating the effects of disruptive peers on student outcomes across the income distribution. Carrell et al. (2018) is the only study examining long-term student outcomes, such as university attendance or attainment of any degree. Their findings point to disruptive peers bringing about negative effects on both low- and high-income students. Carrell and Hoekstra (2010) confirms similar results on test scores in the short-run, though results are imprecisely estimated for the low-income group.

Table A.15. Disruptive peers: share of low-income peers

	University Graduate		
	(1)	(2)	(3)
Fight in School × Bottom 20	-0.03**	-0.03**	
	(0.02)	(0.02)	
Fight in School × Middle	-0.08***	-0.08***	
	(0.01)	(0.01)	
Fight in School × Top 20	-0.08***	-0.08***	
	(0.02)	(0.02)	
Share of Peers Fighting at School × Bottom 20	-0.01	-0.02	0.16
	(0.14)	(0.14)	(0.17)
Share of Peers Fighting at School × Middle	-0.16	-0.16	-0.03
	(0.12)	(0.12)	(0.15)
Share of Peers Fighting at School × Top 20	-0.73***	-0.55***	-0.38
	(0.19)	(0.20)	(0.24)
SLP_{-ics} × Bottom 20		0.18**	0.20**
		(0.07)	(0.09)
SLP_{-ics} × Middle		0.02	0.01
		(0.06)	(0.07)
SLP_{-ics} × Top 20		-0.22*	-0.15
		(0.11)	(0.12)
Observations	11123	11123	8358
R^2	0.25	0.25	0.25
Only Non-Fighters	No	No	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. Fighting at school is an indicator is equal to one if the last physical fight the student had occurred at school. The share of peers fighting at school is a leave-one-out share calculated at the same school-grade level. We also control for the variance of fighting in school at the school-grade level as we do for income. In column (3), we restrict the sample only to those who report not having fought at school.

To assess this, we repeat our baseline regressions after also controlling for the share of peers who have fought at school disaggregated by a student's own-position in the income distribution. As our sample consists of adolescents, we see fighting at school as a particularly salient in-school disruption. Results are reported in the Table A.15. We estimate regressions first including both those who report having been in a fight and those who have not. We then drop fighters to avoid concerns over individual's choice to fight confounding the effects of peer disruption through spillovers (e.g., see Billings and

Hoekstra, 2023). We find highly consistent estimates for the share of low-income peers across the income distribution in all specifications, suggesting our baseline treatment effects are not driven by changes in disruptive behavior. We reiterate here that our flexible income controls and our disaggregation over income of the peer dispersion (SD) of logged household income may already have picked up a mechanism via peers' disruptive behavior and our results here are consistent with this interpretation.

Parental inputs

Another potential explanation for our results is through parental response to changes in the share of low income peers. Recent evidence in fact points to substitution effects between parental beliefs about school quality and parental time investments (Greaves et al., 2023). If parents can observe their child's peers and infer the distribution of peer quality (through peer income), they may react adjusting their inputs or parenting style.[4] If peer quality is viewed by parents as a signal of school quality, parental response could in part compensate, or even dominate, the negative effect of a decrease in school quality (due to a higher share of low-income peers).[5]

To explore this, we leverage three different measures of parental involvement from our survey, based on whether the child reported to have done any of the following activities with their parents: (a) talking about their school work or grades, (b) working on a project for school, and (c) talking about things they were doing in school. Then, we construct a school-related involvement scale and use it as an outcome. We also build a measure of overall involvement, given by a composite scale of ten items including several activities such as going shopping and playing a sport. Results of this exercise are reported in the Table A.16, where we see no response of parental involvement to variation in the share of low income peers across all different outcomes, suggesting that fluctuations in the share of low-income peers does not trigger any sort of parental response.

[4] Recent literature examines how parental style can directly intervene in children's peer group formation (Agostinelli et al., 2020). However, we abstract from this mechanisms as both our theoretical framework and our identification strategy treat peers as exogenously determined.

[5] Fredriksson et al. (2016) also provide evidence that the response of high-income parents is greater than that of other groups, when there is an increase in class size.

Table A.16. Parental involvement

	School-related Involvement			Overall Involvement
	Mother	Father	Parents	Parents
	(1)	(2)	(3)	(4)
SLP_{-ics} × Bottom 20	-0.06	0.03	-0.10	0.02
	(0.19)	(0.24)	(0.19)	(0.19)
SLP_{-ics} × Middle	-0.03	-0.18	-0.13	0.02
	(0.16)	(0.21)	(0.16)	(0.16)
SLP_{-ics} × Top 20	-0.08	-0.32	-0.18	0.18
	(0.23)	(0.26)	(0.22)	(0.22)
Mean Dep Var	0.04	0.04	0.05	0.05
Observations	10699	8049	11073	11073
R^2	0.052	0.054	0.060	0.103

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. We use three measures: (a) talked about their schoolwork or grades, (b) worked on a project for school, and (c) talked about things they were doing in school to construct the school-related involvement scale for mothers and fathers. Scales for mothers and fathers are averaged to create a parent score. Aggregated involvement in column (4) is a composite scale of ten items including all activities such as going shopping, playing a sport, going to a religious service or church-related event, talking about someone they were dating, going to a movie, talking about a personal problem, and having a serious argument about their behavior. Each scale is standardized.

A.5.2 GPA and Advanced Courses with Maximum Sample

Table A.17. GPA and advanced courses: maximum sample estimates

	GPA		Advanced Courses			
	Self	Transcript	Math	Science	English	More than One
	(1)	(2)	(3)	(4)	(5)	(6)
SLP_{-ics} × Bottom 20	-0.02	0.71***	0.40***	0.30**	0.07	0.54***
	(0.14)	(0.24)	(0.12)	(0.15)	(0.20)	(0.16)
SLP_{-ics} × Middle	-0.11	0.57**	0.15	0.15	0.01	0.26*
	(0.12)	(0.22)	(0.11)	(0.14)	(0.21)	(0.14)
SLP_{-ics} × Top 20	-0.26*	0.01	0.14	-0.16	0.11	0.05
	(0.15)	(0.27)	(0.13)	(0.17)	(0.23)	(0.15)
Edu non-response weights	NA	Yes	Yes	Yes	Yes	Yes
Mean Dep Var	2.77	2.41	0.40	0.45	0.23	0.59
Observations	14185	8326	8343	8304	5937	8353
R^2	0.197	0.282	0.255	0.214	0.255	0.245

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. Column (1) shows the effects of share of low-income peers on self-reported GPA from Wave I In-Home data while column (2) shows the effects on average GPA calculated from the first interviewed year to the end of the high school from Wave III high school transcript data. Columns (3) - (6) show the effects of share of low-income peers on the taking rate of advanced courses of Math, Science, English, and if ever took more than one advanced course. We use specific educational sampling weights constructed to adjust for transcript non-response as well as survey non-response in columns (2) - (6). We use our fully available sample in this table.

A.5.3 Risky Behaviors

Effort in school may also be proxied by risky behaviors. Students who work harder at school may be less likely to engage in such behaviors and vice-versa. There is broad evidence that human capital investment reduces risky behavior (Conti et al., 2010; Cutler and Lleras-Muney, 2010; Kenkel et al., 2006), as well as evidence that the stringency of education dampens risky behavior (Hao and Cowan, 2019). This could be explained by time constraints in case of contemporaneous effects as well as expectation effects, if students anticipate the future cost of engaging in risky behavior in terms of reduced return to human capital.

Add Health provides a range of self-reported risky behaviors that we use from wave

I. We assess our effects of interest on these behaviors in Table A.19. We expect these may measured with a degree of error that could obscure results and caution strong conclusions.

We assess drinking behavior in columns (1) - (3). Frequent drinking is an indicator for an above median report on frequency of drinking in the past year; drinking out is whether one drank without their parents present; and binge drinking is an indicator for having ever binged (5 or more) drinks in a single outing in the past year. Next, in columns (4) - (6), we have the number of days one smoked in the past year (column 4); an indicator for above median marijuana use (column 5); and an indicator for having used hard drugs (column 6). Finally, in column (7), we report a measure for having engaged in unprotected sex.

The results for the share of low-income peers have a generally consistent pattern across outcomes. Qualitatively we see mostly negative point estimates for the bottom 20th group and positive point estimates for the top 20th. Many of these are null effects, though not all, thus we do not want to over-interpret them. Nevertheless, the patterns here are consistent with our results on education and particularly show that even if the high income students did not suffer a significant drop in GPA, they still show behavioral patterns consistent with the result on long-term university graduation.

Table A.18. Risky Behavior Summary Statistics

	Mean	SD	Min	Max
Frequently drinking	0.17	0.38	0	1
Drinking with people other than family	0.41	0.49	0	1
Ever binge drinking	0.29	0.45	0	1
Standardized smoking days during the past month	-0.00	1.00	-0.49	2.51
Frequently using marijuana	0.14	0.34	0	1
Ever using hard drug	0.05	0.22	0	1
Standardized having unprotected sex recently	-0.00	1.00	-0.23	6.41
Observations	11165			

This table presents summary statistics for the risky behaviors in Table A.19 after restricting to our analytic sample. The smoking variable originally ranges from 0 to 30 days, and the unprotected sex variable ranges from 0 to 5 times. Both variables have many zeros (69.9% and 94.4%, respectively) and are highly right-skewed. We standardize them to mean 0 and standard deviation 1.

Table A.19. Risky Behavior Outcomes

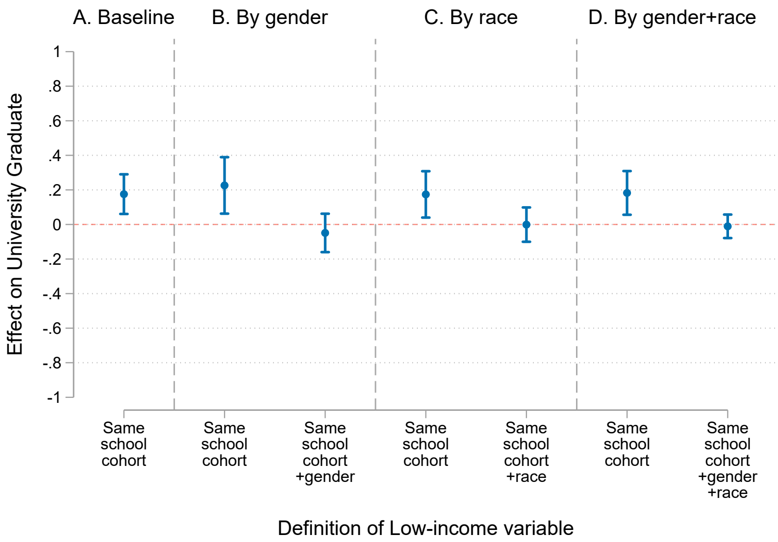
	Frequent Drinking	Drinking Out	Binge Drinking	Smoking	Marijuana	Hard Drug	Unprotected Sex
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SLP_{-ics} × Bottom 20	-0.11	-0.03	-0.09	-0.35*	0.02	-0.00	0.12
	(0.07)	(0.09)	(0.10)	(0.19)	(0.06)	(0.04)	(0.19)
SLP_{-ics} × Middle	-0.06	-0.02	-0.05	-0.05	0.06	0.07**	0.34**
	(0.06)	(0.07)	(0.08)	(0.16)	(0.05)	(0.03)	(0.17)
SLP_{-ics} × Top 20	0.05	0.09	0.10	0.29	0.11*	0.14***	0.46**
	(0.08)	(0.10)	(0.09)	(0.20)	(0.06)	(0.05)	(0.22)
Mean Dep Var	.17	.41	.29	0	.14	.05	0
Observations	11092	11101	10100	9502	11011	11021	11162
R^2	0.083	0.137	0.139	0.134	0.075	0.039	0.038

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. School and cohort fixed effects are included in all specifications. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification in column (2) of Table 2.2. We trim our data to our analytic sample as in Table 2.2 and standardize smoking and unprotected sex outcomes to mean 0 and standard deviation 1.

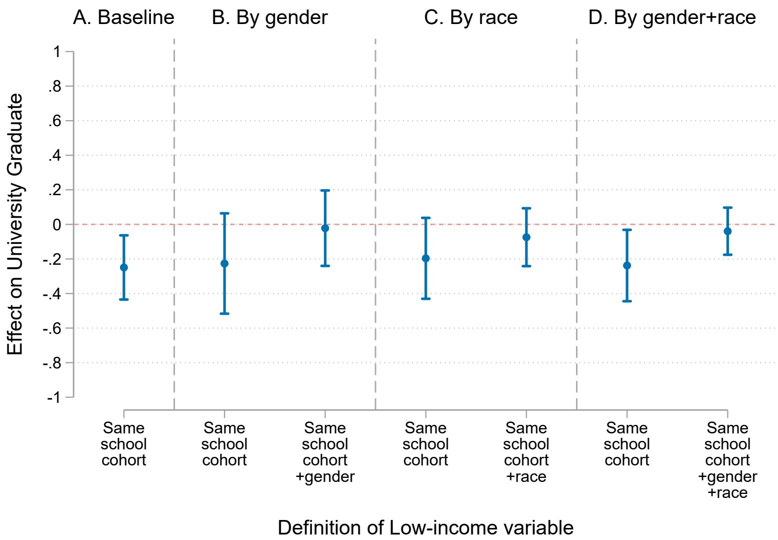
A.5.4 Social Cohesion: Additional Results

Figure A.8. University completion: different definitions of peers groups

(a) Peer effect estimates for bottom 20



(b) Peer effect estimates for top 20



Notes: These figures tests how different definitions of peer groups compare against our baseline effects from the share of low-income peers on university graduation. We always include school and cohort fixed effects as in column (2) of Table 2.2. Panel A presents the estimates for students in the bottom 20th percentile of household income. Panels B presents the estimates for students in the top 20th percentile of household income. In each sub-panel, we include both definitions of the share of low-income peers in the regression. The middle-income students are included in the regression but we omit the estimates here as they are null effects.

A.6 Heterogeneity via a Causal Forest

We want to examine heterogeneity across subgroups in our data that may be relevant for policy, e.g., by gender, single parent homes, and so forth. However, our main results are already heterogeneous by whether a student is from a low, middle, or higher-income family. Thus, further heterogeneity across many dimensions is difficult. While absent a larger sample there is no way to avoid this problem, we can use the recently developed, and data driven, causal forest approach to gain a better idea around how our effects differ across both observable dimensions in our data and the family income groups we have used throughout the paper.

Causal forests change the problem from estimating differences in effects across specific groups to nonparametrically recovering heterogeneous treatment effects across individuals. This approach, pioneered by Athey and Imbens (2016), Athey et al. (2019), and Wager and Athey (2018), adapts regression trees to capture how treatment effects vary across partitions based on feasible combinations of observable control variables. With a binary treatment, this implies estimating differences in potential outcomes at realization of specific values among the observed controls yielding conditional average treatment effects (CATEs). In our case, we recover conditional average partial effects as $E[Cov[Y_i, W_i]|X_i]/Var[W_i|X_i]$ where Y_i is university graduation, W_i is the share of low-income peers, and X_i is our vector of exogenous individual characteristics. We will refer to these as CATEs for simplicity.

Causal forest works by growing trees. Put simply each tree is a partition of leaves whereby each leaf is a subset of observations with particular realizations of characteristics. Leaves are partitioned by maximizing the variance in treatment effects across partitions tuned with cross validation. In the "honest" implementation of Wager and Athey (2018), each tree is grown by randomly splitting the data into training and estimation subsets, using the training data to grow the tree, i.e., find the partitions, and the estimation sample to make the "out of bag" estimation of the treatment effects within partitions. The out of bag estimates are estimated on each leaf and then aggregated across trees. Importantly, Athey et al. (2019) show that treatment effect estimates under unconfoundness and "honesty" are asymptotically normal, allowing the calculation of confidence intervals.[6]

We employ causal forests but with two pre-step modifications. Note that causal forests rely on unconfoundness either via randomization or through conditioning. Thus, step one: we residualize Y, W, and each of our controls removing school and cohort fixed effects and we do this separately with the bottom 20th, middle, high-income groups. Next, we want to investigate heterogeneity within our already defined low, middle, and high-income groups due to our pre-existing focus on these groups motivated from our theory. Thus, step two: we run the causal forest on each of these income groups separately using the residualized variables from step one. Moreover, we employ cluster-robust random forests at the school level as shown in Athey and Wager (2019).[7] Finally, we stack the out of bag CATE estimates across income groups for analysis.

We first demonstrate that the pattern in the CATEs across income groups matches closely to our previous results in panel (a) of Figure A.9. For the bottom 20th income group, the interquartile range falls entirely in the positive domain with a median of 0.234. The middle group falls right around zero. And, finally, the top 20th group has an interquartile range below zero with a median of -0.229.

Next, in panel (b), we check whether our results vary over cognitive ability. We have already discussed the link between income and ability and we have controlled flexibly for ability and school-cohort ability rank. It could be, however, that only a portion of the ability distribution drives our results. For instance, Carlana et al. (2022a) focus on a treatment applied to higher ability disadvantaged students who at pre-treatment tended to hold lower beliefs about their educational possibilities relative to more advantaged students of the similar ability. It is useful for policy then to understand whether an aspiration gap mechanism centers around certain portions of the ability distribution or is relevant across ability types. We, however, expect that this mechanism is relevant across cognitive ability types, per our arguments that capacity is broader than just cognitive ability, meaning students of different ability types are also faced with other

[6]This discussion omits complexities on tuning parameters discussed in Athey and Imbens (2016), Athey et al. (2019), and Wager and Athey (2018).

[7]To implement, we use the *grf* package and *causal_forest* command in *R*.

skills and constraints that our mechanism can operate around.

In panel (b) of Figure A.9, we find rather homogeneous effects across the ability distribution (PVT scores) among the bottom 20th and middle-income groups. For the bottom 20th, effects are always positive and quite similar and for the middle-income group the CATEs are near zero and similar across ability. The top 20th group does show some heterogeneity with effects that are always negative but somewhat mitigated at the top end of the ability distribution. While these students may well have a very high capacity, this pattern is suggestive that very high ability students are likely to complete university for many other reasons or they place less weight on the social environment to determine their reference points. This is proxied by γ in our theory. Students with a high family income but who are not in the top of the ability distribution may still have higher capacity due to better opportunities – or alternatively have high beliefs due their family income such that their beliefs are above their true capacity – and may then be the ones who put more weight on the social environment to determine their reference points.

We then report binscatter plots across income deciles split by gender and by dual vs. single parent homes in panel (c) and (d) of Figure A.9. The effects are generally similar across genders but with females experiencing stronger, more positive, effects in the bottom 20th, and somewhat more negative effects in the top 20th. Students from dual parent homes exhibit a similar pattern, with particularly stronger effects among the top 20th.

Now we turn to evaluate the variation in the CATEs across the set of individual characteristics in Table A.20. Our individual characteristics included in the causal forests correspond to those in the Appendix Table A.1. We split each income group by those with a high or low CATE (above or below the median)[8] and then test mean differences in having a high or low CATE across student characteristics and report a p-value adjusted for multiple hypothesis test bias.

First, the median CATE in each income group matches our expectations and pre-

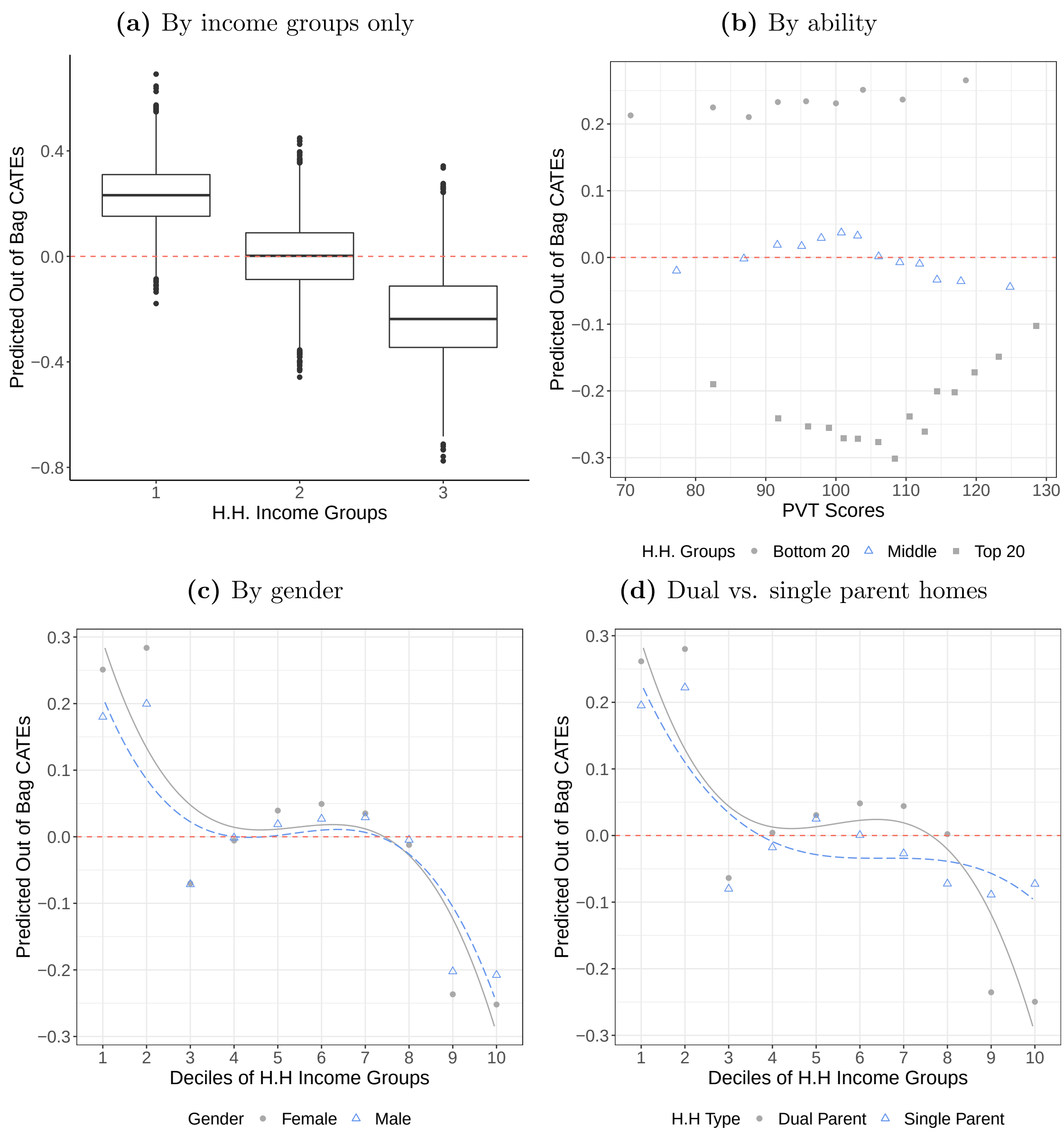
[8]Our approach here is similar to that of Carlana et al. (2022a) except that we split across income groups.

vious results. The median CATE is 0.234 for the bottom 20th, −0.002 for the middle, and −0.229 for the top 20th income group. Second, we see a number of significant differences across high and low median groups in terms of characteristics. Many of these are minimal in magnitude; however, gender and single parent homes stand out.

We find that in the bottom 20th there are significantly more females and more students from dual parent homes with an above median CATE. For the top 20th, we continue to see significant heterogeneity by gender and single parent home status. These differences are significant even after adjusting for multiple hypothesis test bias. Here there is a higher share of females and students from dual parent homes with a below median CATE – as the median here is negative this implies they have a larger magnitude effect in absolute value.

In this case, a reasonable assumption is that adolescents in dual parent homes, and where incomes are high, likely have high capacity through a broader range of opportunities and fewer life stressors. Thus, these students would be farther ahead of their aspiration reference point as the share of low-income peers increases. We cannot, however, make conclusions here and look to these results as suggestive. Possibly a more important takeaway from this exercise is that our results overall are quite consistent across income groups.

Figure A.9. Causal forest heterogeneity in CATEs by income groups



(a) By income groups only

(b) By ability

(c) By gender

(d) Dual vs. single parent homes

Table A.20. Causal forest: heterogeneity in the CATEs by individual characteristics

	Bottom 20				Middle				Top 20			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	High Predicted CATE	Low Predicted CATE	Diff.	Romano-Wolf p-value	High Predicted CATE	Low Predicted CATE	Diff.	Romano-Wolf p-value	High Predicted CATE	Low Predicted CATE	Diff.	Romano-Wolf p-value
Less than HS Parents	0.302	0.388	-0.086	0.130	0.091	0.132	-0.040	0.031	0.038	0.023	0.014	0.264
HS or GED Parents	0.572	0.507	0.064	0.368	0.546	0.547	-0.001	0.986	0.209	0.342	-0.133	0.003
College Parents	0.127	0.105	0.022	0.768	0.363	0.322	0.041	0.063	0.753	0.635	0.118	0.003
Missing Parents' Education	0.076	0.093	-0.017	0.800	0.035	0.043	-0.008	0.520	0.027	0.006	0.021	0.037
Female	0.700	0.395	0.305	0.003	0.525	0.516	0.008	0.920	0.459	0.564	-0.105	0.017
Age	15.298	15.714	-0.416	0.029	15.560	15.322	0.238	0.016	15.444	15.609	-0.165	0.264
Age Squared	236.792	249.677	-12.885	0.029	244.905	237.707	7.198	0.018	241.020	246.604	-5.583	0.209
Hispanic	0.238	0.229	0.008	0.962	0.120	0.169	-0.049	0.018	0.083	0.053	0.030	0.182
Black	0.326	0.362	-0.037	0.721	0.199	0.156	0.043	0.047	0.092	0.165	-0.073	0.017
Asian	0.028	0.021	0.007	0.876	0.057	0.042	0.014	0.135	0.075	0.052	0.022	0.264
Other Races	0.022	0.021	0.001	0.962	0.018	0.015	0.003	0.920	0.010	0.008	0.002	0.875
Missing Races	0.000	0.001	-0.001	0.883	0.000	0.003	-0.003	0.090	0.005	0.000	0.005	0.264
Child of an Immigrant	0.228	0.212	0.016	0.895	0.153	0.182	-0.030	0.090	0.138	0.115	0.023	0.270
Missing Child of an Immigrant Info	0.002	0.003	-0.001	0.962	0.001	0.001	-0.000	1.000	0.000	0.001	-0.001	0.584
Single Parent Household	0.528	0.707	-0.179	0.003	0.202	0.328	-0.126	0.002	0.187	0.027	0.160	0.003
Family Size	3.928	3.361	0.567	0.003	3.982	3.624	0.358	0.002	3.893	3.907	-0.014	0.875
Ability (AHPVT scores)	96.594	93.130	3.463	0.035	101.719	103.014	-1.295	0.047	108.521	105.925	2.595	0.017
Median CATEs	0.232				0.002				-0.237			
Observations	1090	1090	2180		3460	3460	6920		1033	1032	2065	

Notes: We report summary statistics as the mean for each characteristic split by those above or below the median of CATEs in a specific income group. We also report the difference between the means in columns 3, 7, and 11. Columns 4, 8 , and 12 show the Romano-Wolf p-values adjusted for multiple hypothesis testing. Note that for the top 20 group an above median (high) CATE would imply values closer to zero and a below median (low) CATE implies values that are more negative. See Figure A.9 for reference.

Appendix B

Chapter 3 Appendix

B.1 Beliefs Elicitation Survey: Design and Technical Details

B.1.1 Beliefs Elicitation Survey Design

B.1.2 Beliefs Elicitation Survey Technical Details

Here, we present a few details on the survey which are important for the purpose of our analysis. First, participants cannot proceed the next scenario page without answering all six questions on the first scenario page. Second, when we ask participants what they expect the earnings of the child to be at age 30, we set up a textbox entry, in which we already set a fixed entry as such "**X,000£**". Hence, participants only have to fill in the "X" with their actual expected value, e.g., "32" for "32,000£".

As explained in Subsection 3.3.2, we provide an attention check (cf. Figure B.1) during and a confidence check (cf. Figure B.2) after the beliefs elicitation survey.

Finally, before collecting demographic variables on participants, we provide them with a last attention check, presented in Figure B.3.

Table B.4 below presents the summary statistics for our attention and confidence checks.

Table B.1. Sample Demographics

	Treated	Control	Diff.	Overall
Participant's Characteristics				
Male	0.47	0.49	-0.01	0.48
Female	0.51	0.50	0.00	0.51
Other	0.02	0.01	0.01	0.01
Participant's age	39.16	37.80	1.37	38.43
White	0.87	0.82	0.05	0.84
Other ethnic background	0.13	0.18	-0.05	0.16
Language at home: English	0.99	0.92	0.07*	0.96
Language at home: other	0.01	0.08	-0.07*	0.04
Education: less than a degree	0.47	0.40	0.07	0.43
Education: degree or higher	0.53	0.60	-0.07	0.57
Participant's Employment				
Non-employed	0.17	0.19	-0.02	0.18
Employed	0.83	0.81	0.02	0.82
Job: part-time	0.26	0.19	0.07	0.22
Job: full-time	0.58	0.63	-0.05	0.61
Job: NA	0.16	0.18	-0.02	0.17
Weekly hours worked	30.87	30.13	0.74	30.47
Household Characteristics				
Partner's montly net income	5771.20	6114.89	-343.70	5954.78
Partner's gender: male	0.44	0.47	-0.03	0.45
Partner's gender: female	0.46	0.43	0.03	0.44
Partner's gender: prefer not to say	0.01	0.01	0.00	0.01
No partner	0.09	0.10	-0.00	0.10
Not married	0.43	0.39	0.04	0.41
Married	0.57	0.61	-0.04	0.59
Number of dependent children (aged 0-16) in the household	1.81	1.85	-0.04	1.83
Age of the first (eldest) child	9.79	7.77	2.02**	8.71
Age of the youngest child	6.64	4.60	2.04**	5.55
Nb. of hours spent per day on helping children develop their skills	2.72	2.82	-0.10	2.77
Nb. of hours spent per day doing outdoor activities with children	1.78	1.92	-0.14	1.85

Notes: * $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$. NA stands for "Not Applicable".

Table B.2. Randomization in Set-up

	Mean	SD	Count
Boy (vs. girl)	0.49	0.50	122
Aged 4 (vs. 10)	0.52	0.50	130
10% (vs. 20%) spent on the child's educational activities	0.48	0.50	120
Received the treatment (vs. control)	0.53	0.50	133

Notes: Total $N = 249$. This table presents descriptive statistics for the randomization in set-up. For instance, 49% of our sample (i.e., 122 participants) got displayed, in the hypothetical scenarios, a boy child. 53% received the treatment (i.e., 133 participants out of 249), while the remaining 47% (i.e., 116) participants were part of the control group.

Table B.3. OLS Results – Share Spent on Education (SSE_i) on Expected Outcomes

	(1) $\mathbb{P}$(graduate)	(2) ln(earnings)
$\mathrm{SSE_i}$ = 20% (*ref:10%*)	-0.024	0.003
	(0.022)	(0.030)
$\mathrm{MWL_{j=1}} \times \mathrm{SSE_i}$ = 20% (*ref:10%*)	-0.019***	-0.027***
	(0.006)	(0.010)
Individuals	249	249
Observations	1494	1494
Controls	Yes	Yes

Notes: * p <0.10; ** p <0.05, *** p <0.01. Standard errors in parentheses. Control variables include all hypothetical household income categories, as well as hypothetical child's gender and age.

Figure B.1. Attention Check (1/2)

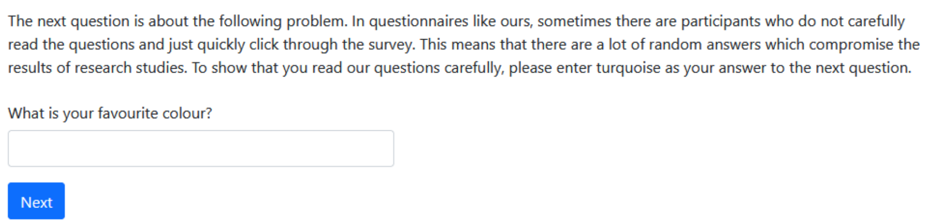


Figure B.2. Confidence Check (1/1)

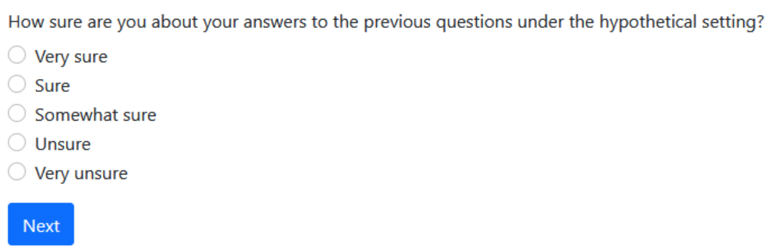
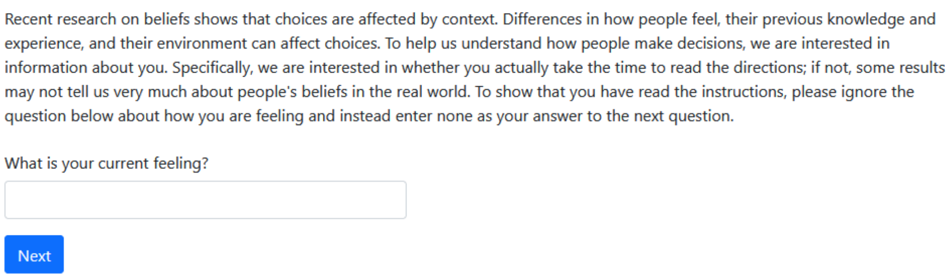


Figure B.3. Attention Check (2/2)



Recent research on beliefs shows that choices are affected by context. Differences in how people feel, their previous knowledge and experience, and their environment can affect choices. To help us understand how people make decisions, we are interested in information about you. Specifically, we are interested in whether you actually take the time to read the directions; if not, some results may not tell us very much about people's beliefs in the real world. To show that you have read the instructions, please ignore the question below about how you are feeling and instead enter none as your answer to the next question.

What is your current feeling?

Next

Table B.4. Attention and Confidence Checks Summary Statistics

	Mean	SD	Min	Max	N
Turquoise check passed	0.99	0.11	0	1	246
Feeling check passed	1.00	0.06	0	1	248
Confidence: very sure	0.04	0.21	0	1	11
Confidence: sure	0.18	0.38	0	1	44
Confidence: somewhat sure	0.41	0.49	0	1	102
Confidence: unsure	0.31	0.46	0	1	76
Confidence: very unsure	0.06	0.25	0	1	16

Notes: Total number of participants: $N = 249$. Interpretation notes: 99% and almost 100% of our participants passed respectively the “turquoise” and “feeling” attention checks. As for the confidence in their answers for the hypotheticals, 63% of our participants report being at least “somewhat sure”.

B.2 Additional Results: Hypothetical Beliefs Elicitation

Table B.5. OLS Results of $\hat{\delta}$ – Gendered Beliefs Stratified by Hypothetical Features

	Child's Gender		Child's Age		Working Hours Profile		Wage Profile		First Shown	
	Boy	Girl	4	10	36 hrs.	42 hrs.	Lower	Higher	Father	Mother
$\mathbb{P}$ (graduate): $\text{MWL}_{j=1}$	-0.025***	-0.004	-0.014**	-0.014**	-0.012*	-0.016**	-0.015**	-0.013**	-0.026***	-0.004
	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)
ln(earnings): $\text{MWL}_{j=1}$	-0.028**	-0.005	-0.017	-0.016	-0.020	-0.013	-0.015	-0.017	-0.026**	-0.007
	(0.013)	(0.012)	(0.011)	(0.014)	(0.014)	(0.011)	(0.012)	(0.012)	(0.013)	(0.012)
Individuals	122	127	130	119	128	121	115	134	118	131
Observations	732	762	780	714	768	726	690	804	708	786
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household Income Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * p <0.10; ** p <0.05, *** p <0.01. Standard errors are clustered at the individual level. We estimate equation 3.3 for each of the feature, i.e., when the hypothetical's child is a girl, a boy, is 4, etc. and report here the $\hat{\delta}$ associated with $MWL_{j=1}$ for each of those regressions. The "first shown" column corresponds to what the participant was shown first $MWL = 1$ (mother works longer hours) or $MWL = 0$ (father works longer hours) in the beliefs elicitation survey.

Heterogeneity by participant's characteristics. We report here additional results for gendered beliefs, by participant's main characteristics, i.e., gender, education, employment status, and voting behavior.

Robustness checks. Here, we ensure the robustness of our key findings by implementing different sample restrictions:

1. We keep participants who report being at least somewhat certain about their answers to the hypothetical beliefs elicitation (column 1);
2. We keep participants who passed at least one out of the two attention checks (column 2);
3. We exclude participants with the 5% lowest and highest response times (column 3).

Table B.6. OLS Results of $\hat{\delta}$ – Beliefs by Participant's Characteristics

	Gender		Degree or More		Employment			Vote UK Elections			
	Male	Female	Yes	No	Part-time	Full-time	None	Consevative	Liberal	Other	None
Panel A: $\mathbb{P}$ (graduate): $\mathrm{MWL_{j=1}}$	-0.020***	-0.008	-0.007	-0.024***	-0.009	-0.020***	-0.002	-0.024	-0.009	-0.008	-0.022**
	(0.008)	(0.007)	(0.007)	(0.009)	(0.012)	(0.007)	(0.010)	(0.015)	(0.007)	(0.017)	(0.011)
Panel B: ln(earnings): $\mathrm{MWL_{j=1}}$	-0.015	-0.015	-0.019*	-0.013	-0.014	-0.016	-0.022	-0.043	-0.000	-0.007	-0.039*
	(0.013)	(0.012)	(0.011)	(0.015)	(0.021)	(0.012)	(0.015)	(0.026)	(0.011)	(0.023)	(0.023)
Individuals	120	126	142	107	55	151	43	47	128	27	47
Observations	720	756	852	642	330	906	258	282	768	162	282
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household Income Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the individual level. We estimate separately each regression for each participant's characteristic, i.e., for when the participant is a man, a woman, has a degree, etc. We display here $\hat{\delta}$ associated with $MWL_{j=1}$ for each of these regressions. Estimates for genders other than male and female are not shown due to sample size issues. For the columns "Vote UK Elections", we asked participants: "Which party did you choose as your primary vote in the last UK General Election?" and provided them with a list of candidate parties. We condensed information as follows – Conservative Party = Conservative, Labour and Green Party = Liberal, Liberal Democrats and any other = Other, None = None.

Table B.7. OLS Results of $\hat{\delta}$ – Beliefs Robustness

	(1) At Least Somewhat Certain		(2) Passed At Least (1/2) Check		(3) Response Time	
	$\mathbb{P}$ (graduate)	ln(earnings)	$\mathbb{P}$ (graduate)	ln(earnings)	$\mathbb{P}$ (graduate)	ln(earnings)
$MWL_{j=1}$	-0.016***	-0.029***	-0.012***	-0.016**	-0.015***	-0.022***
	(0.006)	(0.011)	(0.005)	(0.008)	(0.005)	(0.008)
Individuals	152	152	236	236	216	216
Observations	912	912	1416	1416	1296	1296

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the individual level. Column 1 "At least somewhat certain" corresponds to the first restriction, keeping only individuals who reported being somewhat certain about their answers to the beliefs elicitation. Column 2 "passed at least (1/2) check" corresponds to the second restriction, keeping only participants who passed at least one out of the two attention checks. Finally, column 3 "response time" corresponds to the third sample restriction, excluding participants with the 5% lowest and highest response times.

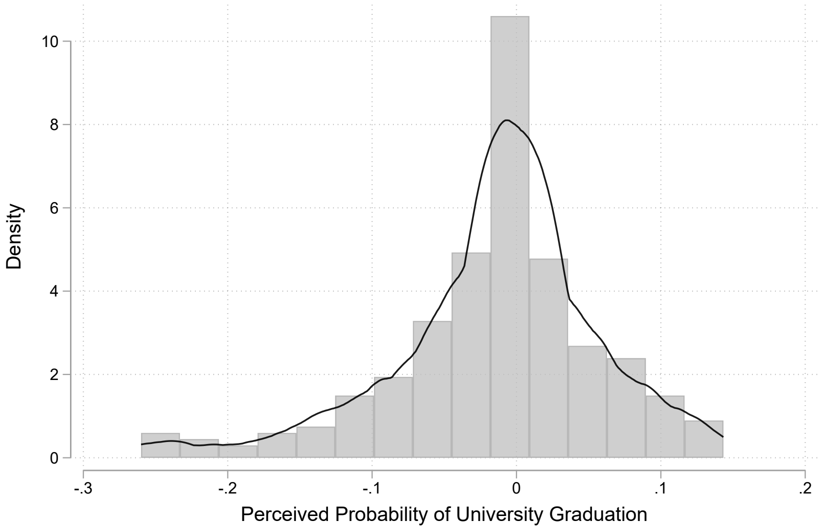
B.3 Perceived Returns Estimation Strategy and Additional Results

Distribution of perceived returns. Figure B.4 below presents the distribution of our perceived returns on both dimensions: 1) the probability of the child graduating from University, and 2) the log of the child's expected earnings at age 30.

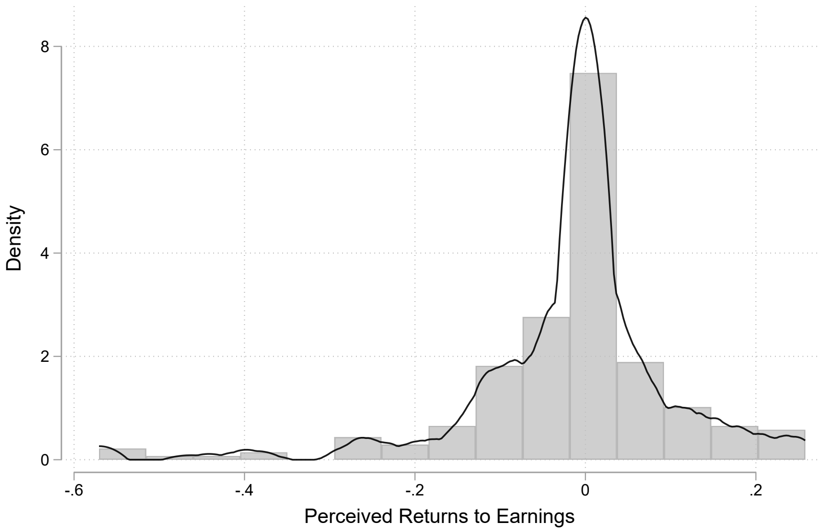
Relationships of perceived returns. We present below a scatter plot between our two thetas (for the probability of graduating from University and the log of expected earnings) with a line of best fit.

Figure B.4. Distribution of Perceived Returns

(a) Panel A: $\mathbb{P}$(graduate)



(b) Panel B: ln(earnings)



Notes: $N = 249$. Histogram of $\theta_{graduate,i}$ (Panel A) and $\theta_{earnings,i}$ (Panel B), both winsorized at the 1% and 99% levels to account for outliers, with kernel density plot.

Figure B.5. Relationships of Perceived Returns

Notes: Relationship of perceived returns for the log of the child's expected earnings, $\theta_{earnings,i}$, at age 30 (x-axis) and the probability of the child graduating from University, $\theta_{graduate,i}$ (y-axis). Both perceived returns are winsorized at the 1% and 99% levels to account for outliers.

Table B.8. Determinants of Individual-level Implicit Costs

	(1) θ_{graduate}	(2) θ_{earnings}
Female	0.005	-0.005
	(0.011)	(0.017)
Has at least a degree	0.025**	0.001
	(0.010)	(0.017)
Age $\geq$ median	-0.006	-0.020
	(0.010)	(0.015)
Other ethnic background	-0.024*	-0.023
	(0.013)	(0.021)
Employment (ref: part-time)		
Full-time	-0.006	-0.001
	(0.014)	(0.024)
Not Applicable	0.010	-0.006
	(0.015)	(0.025)
Nb. of Dependent Children in the HH (ref: 1 or less)		
2 or more	-0.002	-0.010
	(0.010)	(0.017)
Vote at the last UK General Elections (ref: conservative)		
Liberal	0.016	0.042*
	(0.014)	(0.024)
Other	0.015	0.034
	(0.020)	(0.030)
None	0.004	0.005
	(0.016)	(0.031)
Average Return	-.0141	-.0162
Individuals	249	249
Observations	1494	1494

Notes: * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors. This table presents regressions of parental beliefs on the perceived returns of mothers working longer hours than fathers for (1) the expected probability to graduate, and (2) the log of the child's expected earnings at age. Estimates for genders other than male and female are not shown due to sample size issues. Both perceived returns are winsorized at the 1% and 99% levels to account for outliers. For the columns "Vote UK Elections", we asked participants: "Which party did you choose as your primary vote in the last UK General Election?" and provided them with a list of candidate parties. We condensed information as follows – Conservative Party = Conservative, Labour and Green Party = Liberal, Liberal Democrats and any other = Other, None = None.

Table B.9. Relationship between Perceived Returns and Participant's Behavior

	Skill Time			Out Time			ln(Hours Worked)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: All participants ($N = 249$)									
$\theta_{\text{graduate,i}}$	-2.838		-2.903	-0.139		0.618	-0.065		-0.052
	(2.559)		(3.128)	(1.259)		(1.689)	(0.415)		(0.510)
$\theta_{\text{earnings,i}}$		-0.799	0.075		-0.691	-0.877		-0.030	-0.013
		(0.985)	(1.282)		(0.680)	(0.933)		(0.193)	(0.243)
Mean Dep. Var		2.7711			1.8514			3.4847	
Panel B: Control group ($N = 116$)									
$\theta_{\text{graduate,i}}$	-6.728		-6.801	1.797		2.631	-0.699		-1.261
	(4.507)		(6.087)	(1.655)		(2.346)	(0.559)		(0.895)
$\theta_{\text{earnings,i}}$		-2.296	0.085		-0.050	-0.971		0.146	0.603
		(1.685)	(2.718)		(1.123)	(1.513)		(0.291)	(0.511)
Mean Dep. Var		2.7155			1.7759			3.4847	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We run OLS regressions on three main outcomes, reflecting participant's behavior with their child(ren), and in the labour market. Skill Time corresponds to the number of hours they spend, per day, helping their child(ren) develop their skills, while Out Time is the number of hours they spend, per day, doing outdoor activities with their child(ren). ln(Hours Worked) corresponds to the log of the number of hours worked, per week. These regressions look at the associations between these outcomes and their perceived returns on the probability of the child graduating from University (columns 1), the expected earnings at age 30 (columns 2), and both dimensions (columns 3). Control variables include participant's gender, age, a dummy variable for ethnicity (white vs. non-white), and a dummy variable for having a degree or less.

Table B.10. Relationship between Perceived Returns and Participant's Behavior, by Gender

	Skill Time			Out Time			ln(Hours Worked)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: Men ($N = 120$)									
$\theta_{\text{graduate,i}}$	-1.344		-1.938	0.063		0.273	0.470		0.069
	(3.286)		(4.097)	(2.111)		(2.933)	(0.554)		(0.769)
$\theta_{\text{earnings,i}}$		-0.015	0.615		-0.128	-0.217		0.412	0.388
		(1.133)	(1.559)		(0.785)	(1.259)		(0.278)	(0.388)
Mean Dep. Var		2.7711			1.8514			3.4847	
Panel B: Women ($N = 126$)									
$\theta_{\text{graduate,i}}$	-3.762		-3.491	0.027		0.941	-0.449		-0.252
	(3.828)		(4.440)	(1.608)		(2.172)	(0.559)		(0.630)
$\theta_{\text{earnings,i}}$		-1.306	-0.369		-0.991	-1.243		-0.316	-0.246
		(1.676)	(1.978)		(1.170)	(1.516)		(0.269)	(0.308)
Mean Dep. Var		2.7155			1.7759			3.4847	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We run OLS regressions on three main outcomes, reflecting participant's behavior with their child(ren), and on the labour market for men and women respectively. Skill Time corresponds to the number of hours they spend, per day, helping their child(ren) develop their skills, while Out Time is the number of hours they spend, per day, doing outdoor activities with their child(ren). ln(Hours Worked) corresponds to the log of the number of hours worked, per week. These regressions look at the associations between these outcomes and their perceived returns on the probability of the child graduating from University (columns 1), the expected earnings at age 30 (columns 2), and both dimensions (columns 3). Control variables include participant's gender, age, a dummy variable for ethnicity (white vs. non-white), and a dummy variable for having a degree or less.

B.4 Information Treatment: Design and Technical Details

B.4.1 Information Treatment Construction

For both the information treatment construction and the incentivized beliefs about behavioral problems, we make use of the same dataset – the Millennium Cohort Study (MCS). In this section, we briefly present the dataset as well as the data manipulations to get our final datasets , corresponding to different time points.

The Millennium Cohort Study. The Millennium Cohort Study (MCS) follows the lives of around 19,000 young people ($N = 18,818$) born across England, Scotland, Wales and Northern Ireland in 2000-02. The MCS provides multiple measures of the cohort members' physical, socio-emotional, cognitive and behavioral development over time, as well as detailed information on their daily life, behavior and experiences. Alongside this, rich information on economic circumstances, parenting, relationships and family life is available from both resident parents.[1] For the purpose of our analysis, we use parents' reported information on various variables, such as their education, their employment status and number of hours worked per week, as well as the joint net household income, with which we merge their child's information we are interested in (i.e., non-cognitive outcomes and GCSEs pass rates later on).

Data management. We construct two different datasets, respectively, for the incentivized beliefs and for the information treatment. On the one hand, the incentivized beliefs correspond to the number of primary school age children out of 100 having more behavioral problems than the median student of their own gender. Therefore, we make use of sweeps 1 (9 months of the child), 3 (age 5), and 4 (age 7). Parental education is reported at sweep 1, and we consider sweeps 3 and 4, which correspond to primary school ages. We don't go further than sweep 4 because behavioral problems (see the next paragraph for more details) are measured at sweep 4. In turn, we merge information reported from sweeps 1 to 4, excluding sweep 2 (age 3, not corresponding

[1] See the MCS website for a more detailed description of the survey.

to primary school age). We keep only England as well as dual parents families. This leaves us with a sample of 6,787 children.

On the other hand, the information treatment (see the "outcomes" paragraph for further details) corresponds to the share of 5 or more GCSEs passed, by mother's working hours during primary school (i.e., ages 5 and 7). To obtain this information, we make use of the same sweeps as above but also include sweeps 7 (age 17), when the GCSEs outcomes are measured. In turn, we merge parents' reported information from sweeps 1, 3, 4 and 7 (excluding sweep 2 again for the same reasons as above). Finally, we keep only England as well as dual parents families, leaving us with a sample of 5,457 children.

Outcomes. The different outcomes have been constructed at the individual (child) level, at different sweeps.

First, to obtain the *number of primary school children out of 100 that have more behavioural problems than the median child of their own gender*, we used the age 7 sweep (i.e., sweep 4), in which parents respond to the Strengths and Difficulties Questionnaire (SDQ). More precisely, we focus on the externalising score (which we refer to as the 'behavioural problems'), which ranges from 0 to 20 and is the sum of the conduct and hyperactivity scales[2]. Once we have calculated this score for every child aged 7, we create an individual dummy variable taking the value 1 if the child is above the median of this score – meaning that s/he has more behavioral problems than the median – and taking the value 0 otherwise.

Second, to obtain the *share of 5 or more GCSEs passed*, we used the age 17 sweep (i.e., sweep 7) in which young pupils have been asked about their educational attainment. At the age of 17, we expect students to have passed their GCSEs. Since the exam conditions and requirements vary depending on which country you live in, we restrict our analysis to England only. In England, particularly, students are expected to take 9 subjects in GCSEs, among which 3 of them are compulsory – Maths, English and Science. Maths earns you 1 GCSE, English 2 GCSEs (English Language and Lit-

[2]See the Early Intervention Foundation website.

erature) and Combined Science, which is worth 2 GCSEs[3]. In turn, we calculate the within-person number of GCSEs passed and create a dummy variable taking the value 1 if they have achieved 5 or more A*- C (4) grades at GCSE; otherwise this dummy variable takes the value 0.

Estimation and predictors. To derive both final figures, we first regress our outcomes on educational and income fixed effects. We then predict the residuals, to which we add back the mean of our outcome variables[4]. For the educational fixed effects, we use parents' reported information at sweep 1 on their highest education achievement. The final education variable we construct is a 3-category variable (high, medium and low), as presented in Table B.11. We include both mother's and father's educational fixed effects in the regressions, for both outcomes.

Table B.11. Education Coding Scheme

3-category coding	9-category coding	Questionnaire items included
High education	1. Higher degree	Higher degree (A)
	2. Bachelor's degree	First degree (A) Professional qualifications at degree level (V)
	3. HE below degree	Diplomas in higher education (A) Nursing or other medical qualifications (V)
Medium education	4. A-level	A/AS/S level (A) NVQ/SVQ/GSVQ Level 3 (V)
	5. Trade apprenticeship	Trade apprenticeship (V)
	6. GCSE A-C	O-level/GCSE grades A-C (A) NVQ/SVQ/GSVQ Level 2 (V)
Low education	7. GCSE D-G	GCSE grades D-G (A) NVQ/SVQ/GSVQ Level 1 (V)
	9. None	None of these (A & V)

Notes: We exclude category 8, which corresponds to "other qualification including oversea", for consistency.

For the income fixed effects, we proceed in two different ways, depending on the information we want to calculate. For the share of children that have more behavioral problems than the median child of their own gender, we use income reported by the

[3]See this website for a full description.

[4]For the share of children having more behavioral problems than the average, the outcome is the above-defined dummy variable taking the value 1 if the child is above the median of this score; 0 otherwise. For the share of 5+ GCSEs passed, the outcome is the dummy variable taking the value 1 if the child has achieved 5 or more A*- C (4) grades at GCSE; 0 otherwise.

parents at sweep 4 (since our outcome is also measured at sweep 4, i.e., age 7 of the child). The variable provided by MCS is a banded variable, including 19 categories, ranging from less than £1,600 a year, to £100,000 or more. There are too many categories to include this variable as fixed effects. Therefore, we divided individuals in quintiles, roughly with respect to the original variable.

For the share of 5+ GCSEs passed, we use the income reported by the parents at sweeps 3 and 4 (corresponding to primary school ages, i.e., ages 5 and 7). Like previously, the income variable is categorical and contains too many categories to be included as fixed effects. Thus, for each sweep and for each individual, we create a new variable who takes the reported income mode and average this new variable across sweeps to create a final version presented in quintiles.

Finally, we create a dummy variable to determine whether the mother works full-time (i.e., 35 hours a week and more) or part-time (less than 35 hours a week). For the share of primary school age children that have more behavioral problems than the median child of their own gender, we use the reported mother's working hours at sweep 4 (time of the outcome measurement). For the share of 5+ GCSEs passed, we take the average of reported mother's working hours at sweeps 3, 4 (corresponding to ages 5 and 7, i.e., primary school age) and end up with a dummy variable equal to 1 if, on average, the mother was working 35 hours or more across those 2 sweeps, 0 if the was working, on average, less than 35 hours.

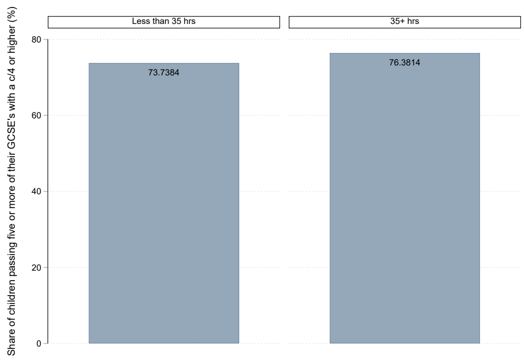
Thus, our final calculated figures, weighted and unweighted, are presented in the figure B.6 below.

B.4.2 Incentivized Beliefs: Associations with Participant Characteristics

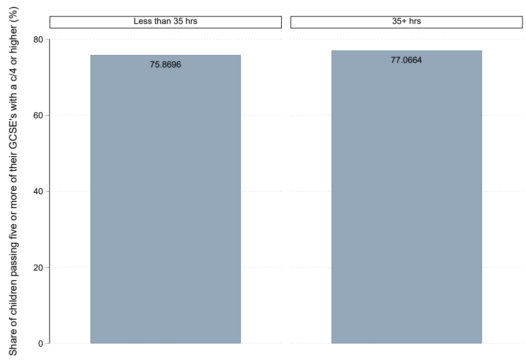
We present in Figure B.7 the correlates of participants' beliefs with individual characteristics. Panel A presents the correlates of the GCSE pass rate belief (i.e., the passing rate of five or more GCSEs with at least C/4, when the mother worked 35 hours or more per week) with participant's main characteristics. We calculate these correlations for the full sample. Panel B presents the correlates of the incentivized behavioral belief

Figure B.6. GCSEs and Behavioral Problems Statistics

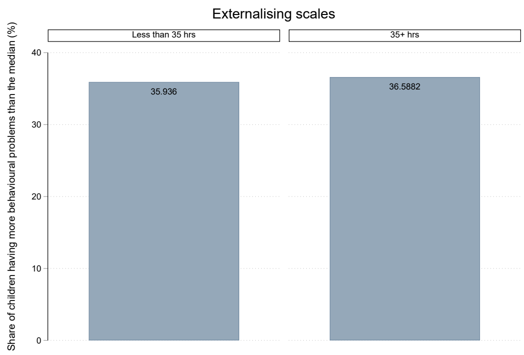
(a) GCSE Unweighted



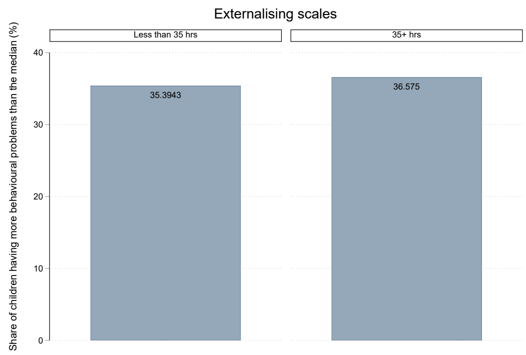
(b) GCSE Weighted



(c) Behavioral Problems Unweighted



(d) Behavioral Problems Weighted



Notes: We present unweighted versions in our survey. The weighted versions are almost identical.

(i.e., about the share of children having more behavioral problems than the median when the mother worked 35 hours or more per week) for the control group only with participant's main characteristics.

Our sample size does not allow us to display significant differences by individual characteristics, but a few points are interesting to note here concerning the direction of some estimates. First, being a woman – as opposed to being a man – is negatively associated with the prior beliefs on GCSEs, but positively associated with the incentivized beliefs. This goes in line with the result we highlight earlier, stating that men tend to hold gendered beliefs, especially concerning the expected probability of graduation. Second, being non-white correlates negatively with the prior beliefs, and positively with the incentivized beliefs. Finally, relative to conservative, voting liberal is negatively associated with both beliefs.

B.4.3 Predictors of Beliefs and Norms

See Table B.12 on the association between beliefs and self-reported norms.

B.4.4 Additional Beliefs, Gender Norms and Policy Views

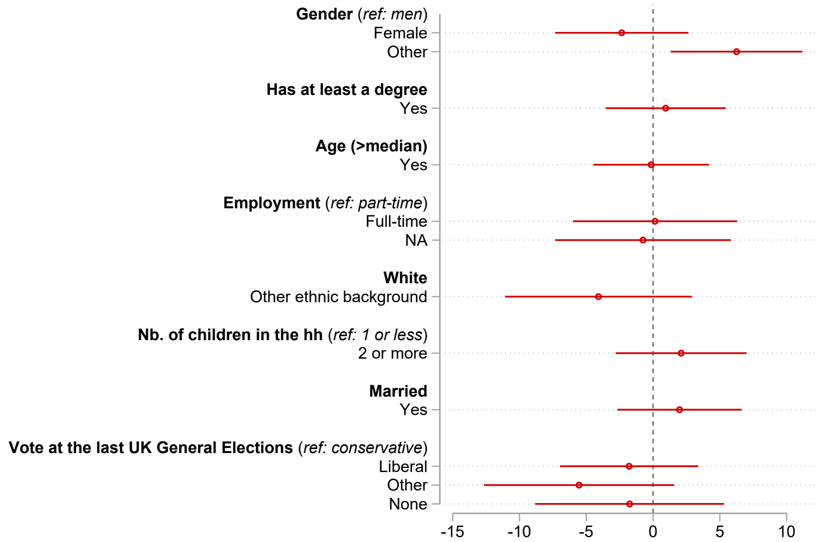
Table B.13 presents the distribution of the beliefs variables about mothers versus fathers decisions and preferences in the investment to children's skills as are defined in the main text Table 3.4.

As stated in the main text, for each question, we define a variable to be equal to 1 if the participant replied “mother” as being the more efficient on the dimension considered (which corresponds to a gendered belief), 0 if they either replied “father” or “both equally”.

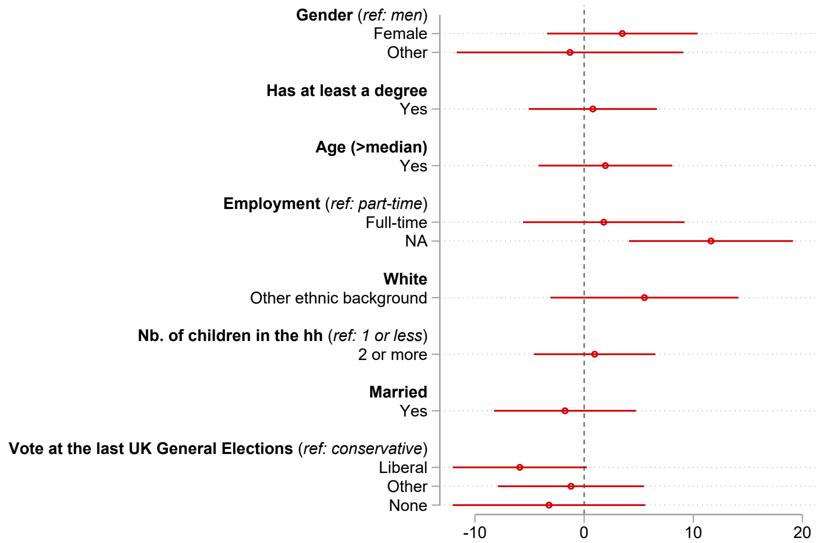
Policy views. We are interested in collecting – without participants knowing that this is linked to our main survey – a set of self-reported policy views on the government spending, taxes, childcare policies, and paternity leave policies. We invited participants back, one week later, by sending them a generic invitation from Prolific to undertake a 5-minute survey that did not reveal the link to the main survey. Table B.14 presents the questions and answer modalities.

Figure B.7. Correlates of Beliefs and Individual Characteristics

(a) Panel A: Correlates with Prior Beliefs on GCSE pass rates



(b) Panel B: Correlates with Incentivized Beliefs on Behavioral Problems



Notes: $N = 249$. The dots indicate the mean values of the estimated multiple regression coefficients. The dependent variable in Panel A is participant's prior belief on the passing rate of five or more GCSEs with at least C/4, when the mother worked 35 hours or more per week. The dependent variable in Panel B is the share of children having more behavioral problems than the median when the mother worked 35 hours or more per week. For gender, the "other" category groups participants who do not identify either as a cisgender man nor as cisgender woman. Lines indicate 95% confidence intervals.

Table B.12. The Association between Beliefs and Self-reported Norms

	Gender Norms Score			$Q1_{bis}$ – Productivity			$Q2_{bis}$ – Resource Allocation			$Q3_{bis}$ – Resource Allocation			$Q4_{bis}$ – Preferences			$Q5_{bis}$ – Preferences		
	Continuous			=1 if mother			=1 if mother			=1 if father			Father's # Hours			Mother's # Hours		
$\theta_{graduate}$	-0.020			-0.013			1.103			0.364			-4.134			-2.309		
	(0.279)			(0.671)			(0.683)			(0.309)			(4.373)			(3.879)		
$\theta_{earnings}$		0.188			-0.548			0.403			0.390			0.370			-0.400	
		(0.169)			(0.380)			(0.462)			(0.272)			(0.926)			(1.524)	
Prior Beliefs: GCSEs			0.002*			-0.001			0.006**			-0.001			-0.022			-0.031*
			(0.001)			(0.003)			(0.003)			(0.002)			(0.018)			(0.017)
Observations	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

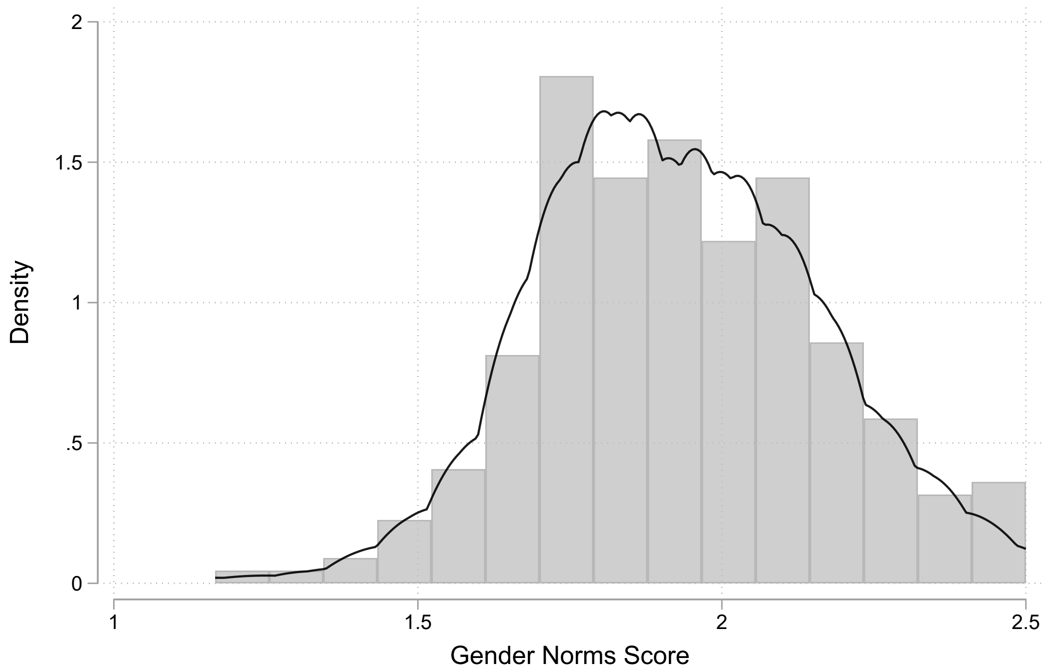
Notes: Control group only. * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors in parentheses. Results of OLS regressions of beliefs, i.e., $\theta_{earnings}$, $\theta_{graduate}$, and the prior beliefs on GCSEs when the mother works longer hours, on gender norms score (derived from Table 3.5), self-reported norms on productivity, resource allocation, preferences. Control variables include participant's gender, age, a dummy for ethnicity, and a dummy for having a degree or less. Note that we use the recoded versions, as described in Section 3.4.1, of the first 3 questions of Table 3.4. $Q1_{bis} = 1$ if the participant thinks the mother would be the most effective if only parent can be involved in time spent on helping the child with educational activities. $Q2_{bis} = 1$ if the participant thinks the mother would allocate more money to the child, if only one parent was allowed to make the resource allocation decisions. $Q3_{bis} = 1$ if the participant thinks the father would be more likely to make the overall resource allocation decisions. $Q4$ and $Q5$ are continuous values corresponding, respectively, to the time the participant thinks the father ($Q4$) and the mother ($Q5$) separately spend helping their children develop educational and social skills.

Table B.13. Distribution of Channels Related to Gender Norms

	Mean	SD	Min	Max	N
Q1 – Productivity (=1 if mother)					
Mother	0.22	0.42	0	1	56
Father	0.12	0.33	0	1	31
Both equally	0.65	0.48	0	1	162
Q2 – Resource Allocation (=1 if mother)					
Mother	0.63	0.48	0	1	158
Father	0.12	0.32	0	1	29
Both equally	0.25	0.43	0	1	62
Q3 – Resource Allocation (=1 if mother)					
Mother	0.57	0.50	0	1	143
Father	0.07	0.25	0	1	17
Both equally	0.36	0.48	0	1	89
Preferences (in hours/day)					
Q4 – Out time	1.85	1.91	0	15	249
Q5 – Skill time	2.77	2.87	0	20	249

Notes: This table presents descriptive statistics for the gender norms questions, presented in Table 3.4. Questions are as follows: **Q1** – In a family where the mother and father have the same education level, in time spent on helping a child with educational activities, if only one parent can be involved, which parent do you believe would be the most effective? **Q2** – In a family where the mother and father have the same education level, in money spent on helping a child with educational activities, if only one parent is allowed to make the resource (money) allocation decisions, which parent do you believe would allocate more money to the child? **Q3** – In a family where the mother and father have the same income level, who do you believe would be more likely to make the resource (money) allocation decisions? **Q4** – In a typical family where the father works full time, how many hours per day on average do you think the father spends helping their children develop educational and social skills? **Q5** – In a typical family where the mother works full time, how many hours per day on average do you think the mother spends helping their children develop educational and social skills?

Figure B.8. Gender Norms Score Distribution



Notes: Histogram of the gender norms score variable with kernel density plot. The minimum (maximum) value can be 1 (5), corresponding to traditional (liberal) attitudes towards gender norms.

Table B.14. Policy Views Variables – Obfuscated Follow-up

	Answer modalities	Final indicator
Q1 – **Public Spending**: Do you think the overall amount of government spending should be increased, decreased, or remain the same?	1. Strongly increased 2. Somewhat increased 3. Kept at its present level 4. Somewhat decreased 5. Strongly decreased	**Q1**$_{bis}$ = 1 if answered 1 or 2; 0 otherwise.
Q2 – **Taxes**: Do you think the overall amount of taxes raised by the government should be increased, decreased, or remain the same?		**Q2**$_{bis}$ = 1 if answered 1 or 2; 0 otherwise.
Q3 – **Wealth**: People feel differently about how far a government should go. Here is a phrase which some people believe in and some don't: Do you think the government should or should not redistribute wealth by heavy taxes on the rich?	1. Yes, redistribute by heavy taxes on the rich 2. No, should not redistribute wealth 3. No opinion	**Q3**$_{bis}$ = 1 if answered 1; 0 otherwise.
Q4 – **Childcare**: Research shows that after having children, women experience a drop in labour earnings (Kleven et al., 2019). This is often explained by the fact that, due to childcare responsibilities, they sort into jobs that offer lower wages but are more flexible and do not require long hours. Do you think that the government should offer policies such as subsidized childcare and universal childcare to help women with children work longer hours?	1. Strongly agree 2. Agree 3. Neither agree nor disagree 4. Disagree 5. Strongly disagree	**Q4**$_{bis}$ = 1 if answered 1 or 2; 0 otherwise.
Q5 – **Paternity Leave**: Research shows that after having children, women experience a drop in labour earnings (Kleven et al., 2019). This is often explained by the fact that, due to childcare responsibilities, they sort into jobs that offer lower wages but are more flexible and do not require long hours. Do you think that the government should offer policies such as paternity leave to help women with children return sooner to work after the childbirth and work longer hours?		**Q5**$_{bis}$ = 1 if answered 1 or 2; 0 otherwise.

B.5 Additional Results: Information Treatment Effects

Heterogeneity by participant's characteristics. We report in Table B.15 below additional results for the information treatment effects, by participant's main characteristics, i.e., gender, education, employment status, and voting behavior.

Beliefs on gender norms. We further look at whether beliefs about gender norms respond to the information treatment. In Table B.16, we report results across the self-reported gender norm questions. We lack enough variation in these measures to say anything efficiently. Yet, we point out some patterns. The pattern in the point estimates suggests that there may be a fair degree of heterogeneity across initial beliefs.

Policy views – obfuscated follow-up. As mentioned earlier in the paper, we are interested in collecting – without participants knowing that this is linked to our main survey – a set of self-reported policy views on government spending, taxes, childcare policies, and paternity leave policies. We end up with 227 participants (out of 249, circa 91%) with valid information. The questions asked to participants are reported in Table B.14 in the Appendix. As we try to encompass participants' policy views into more or less liberal views, but mainly related to women's behavior in the labor market, we focus on Questions 4 and 5 related, respectively, to childcare and paternity leave policies. We recode these two questions so that they equal 1 if the participant replied 1 ("strongly agree") or 2 ("agree"); otherwise 0. We further focus, in column 3, on an indicator reflecting support (or not) for both policies, equal to 1 if the respondent answered supportively of both policies, 0 otherwise.

Table B.15. OLS Results of $\hat{\gamma}$ – Information Treatment Effects by Participant's Characteristics

	Gender		Degree or More		Employment			Vote UK Elections			
	Male	Female	Yes	No	Part-time	Full-time	None	Consevative	Liberal	Other	None
ATE: γ	-0.032	-0.057**	-0.047**	-0.040	-0.019	-0.035*	-0.114***	-0.073***	-0.012	-0.084***	-0.074*
	(0.020)	(0.023)	(0.019)	(0.026)	(0.029)	(0.021)	(0.027)	(0.026)	(0.024)	(0.029)	(0.041)
Individuals	120	126	142	107	55	151	43	47	128	27	47

Notes: * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors in parentheses. We estimate separately equation 3.4 for each participant's characteristic, i.e., for when the participant is a man, a woman, has a degree, etc. We display here $\hat{\gamma}$ associated with D_i for each of these regressions. Estimates for genders other than male and female are not shown due to sample size issues. For the columns "Vote UK Elections", we asked participants: "Which party did you choose as your primary vote in the last UK General Election?" and provided them with a list of candidate parties. We condensed information as follows – Conservative Party = Conservative, Labour and Green Party = Liberal, Liberal Democrats and any other = Other, None = None.

Table B.16. Gender Norms Updating and Information Treatment Effects

	Gender Norms Score	$Q1_{bis}$ – Productivity	$Q2_{bis}$ – Resource Allocation	$Q3_{bis}$ – Resource Allocation	$Q4_{bis}$ – Preferences	$Q5_{bis}$ – Preferences
	Continuous	=1 if mother	=1 if mother	=1 if father	Father's # Hours	Mother's # Hours
All participants: ATE: γ	-0.016	-0.039	0.016	-0.002	-0.001	0.028
	(0.029)	(0.064)	(0.061)	(0.041)	(0.237)	(0.308)
Panel A: by GCSEs Beliefs						
GCSEs under-estimators: γ_1	-0.023	-0.051	0.075	-0.038	0.051	-0.075
	(0.036)	(0.080)	(0.076)	(0.050)	(0.376)	(0.442)
GCSEs over-estimators: γ_2	-0.011	-0.008	-0.090	0.055	-0.048	0.264
	(0.048)	(0.108)	(0.102)	(0.072)	(0.218)	(0.373)
Panel B: by $\theta_{\text{graduate,i}}$						
$(\theta_{\text{graduate,i}} \geq 0)\times$ Treat: γ_1	0.004	-0.069	-0.071	-0.019	-0.171	-0.006
	(0.040)	(0.089)	(0.085)	(0.055)	(0.236)	(0.392)
$(\theta_{\text{graduate,i}} < 0)\times$ Treat: γ_2	-0.039	-0.004	0.107	0.014	0.135	0.016
	(0.043)	(0.094)	(0.089)	(0.065)	(0.442)	(0.474)
Panel C: by $\theta_{\text{earnings,i}}$						
$(\theta_{\text{earnings,i}} \geq 0)\times$ Treat: γ_1	-0.024	-0.047	0.013	0.020	-0.075	0.034
	(0.037)	(0.083)	(0.078)	(0.054)	(0.340)	(0.463)
$(\theta_{\text{earnings,i}} < 0)\times$ Treat: γ_2	0.003	-0.025	0.039	-0.030	0.146	0.086
	(0.048)	(0.104)	(0.099)	(0.064)	(0.296)	(0.338)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individuals	249	249	249	249	249	249

Notes: * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors in parentheses. This table presents regressions of equation 3.4 for all participants. It also provides regressions of equation 3.5, by prior beliefs (Panel A), by average perceived returns for the probability of the child to graduate from University (Panel B), and by average perceived returns for the log of expected earnings (Panel C). The outcomes considered are the following: the gender norms score (column 1), productivity (column 2), resource allocation (column 3), and preferences (column 4 & 5). $Q1_{bis} = 1$ if the participant thinks the mother would be the most effective if only parent can be involved in time spent on helping the child with educational activities. $Q2_{bis} = 1$ if the participant thinks the mother would allocate more money to the child, if only one parent was allowed to make the resource allocation decisions. $Q3_{bis} = 1$ if the participant thinks the father would be more likely to make the overall resource allocation decisions. $Q4$ and $Q5$ are continuous values corresponding, respectively, to the time the participant thinks the father ($Q4$) and the mother ($Q5$) separately spend helping their children develop educational and social skills. Control variables include participant's gender, age, a dummy for ethnicity, and a dummy for having a degree or less.

Table B.17. Information Treatment Effects and Support for Policies

	(1)	(2)	(3)
	Q4_{bis} – Childcare	Q5_{bis} – Paternity Leave	Supportive of Both
All participants: ATE: γ	-0.008	-0.064	-0.091
	(0.048)	(0.065)	(0.066)
Panel A: by GCSEs Beliefs			
GCSEs under-estimators: γ_1	-0.063	-0.126	-0.183**
	(0.065)	(0.086)	(0.087)
GCSEs over-estimators: γ_2	0.072	0.024	0.040
	(0.071)	(0.097)	(0.101)
Panel B: by $\theta_{\text{graduate,i}}$			
$(\theta_{\text{graduate,i}} \geq 0)\times$ Treat: γ_1	-0.021	-0.082	-0.117
	(0.073)	(0.092)	(0.093)
$(\theta_{\text{graduate,i}} < 0)\times$ Treat: γ_2	-0.001	-0.059	-0.080
	(0.065)	(0.093)	(0.095)
Panel C: by $\theta_{\text{earnings,i}}$			
$(\theta_{\text{earnings,i}} \geq 0)\times$ Treat: γ_1	0.069	-0.052	-0.117
	(0.065)	(0.084)	(0.087)
$(\theta_{\text{earnings,i}} < 0)\times$ Treat: γ_2	-0.128*	-0.063	-0.045
	(0.067)	(0.106)	(0.107)
Controls	Yes	Yes	Yes
Individuals	227	227	227

Notes: * p <0.10; ** p <0.05, *** p <0.01. Robust standard errors in parentheses. This Table presents regressions of equation 3.4 for all participants. It also provides regressions of equation 3.5, by prior beliefs (Panel A), by average perceived returns for the probability of the child to graduate from University (Panel B), and by average perceived returns for the log of expected earnings (Panel C). The outcomes considered are final indicators $Q4_{bis}$ and $Q5_{bis}$, presented in the last column of Table B.14, and they are coded such as 1 reflect more liberal attitudes towards women working. In column 3, we create an indicator reflecting support (or not) for both policies, equal to 1 if the respondent answered supportively of both policies, 0 otherwise. Control variables include participant's gender, age, a dummy for ethnicity, and a dummy for having a degree or less.

Appendix C

Chapter 4 Appendix

C.1 Description of Psychological Scales

CFPS uses different scales of mental health distress in different survey waves. One such indicator is the Kessler Psychological Distress Scale (K6), developed by Kessler et al. (2002), which was asked in the 2010 and 2014 surveys. Respondents reported their experiences in the past month on items in Table C.1. We reverse code each item to score as 0 (never), 1 (once a month), 2 (2-3 times a month), 3 (2-3 times a week) and 4 (Almost every day) and aggregate them to a final score ranging from 0 to 24, with higher scores indicting greater depressive symptoms. While a score of 13 usually defines serious mental illness (Kessler et al., 2003), we use a lower threshold of K6 ≥ 5 to indicate moderate mental distress (Prochaska et al., 2012).

Another mental health indicator used in CFPS is the Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977b). The full 20-item CES-D was included in the 2012 and 2016 surveys, while an 8-item version was asked in the 2018 and 2020 surveys. Respondents rated their past-week status on items in Table C.1. Each item was scored as 1 (never(less than one day)), 2 (sometimes (1-2 days)), 3 (often(3-4 days)), and 4 (most of the time (5-7 days)). We reverse code items 4, 8, 12, 16 and aggregate those items, with the 20-item version ranging from 0 to 60 and the 8-item version from 0 to 24. Higher scores indicate more sever depression. The CES-D20 categorizes scores as follows: ≤ 16 indicates no to mild depression, 17-23 indicates moderate depression, and ≥ 24 indicates severe depression (Bi et al., 2023). In this paper, we use a used CES-D20 cut-off of 16, corresponding to an CES-D8 cut-off score of 7, as these scores effectively identify individuals at risk of clinical depression in the Chinese context (Bi et al., 2023).

Both K6 and CES-D are frequently used to evaluate psychological distress and serious mental illness (Kessler et al., 2003; Kim et al., 2016; Weissman et al., 1977). Because these scales do not appear in all survey waves, we construct a consistent variable called "distress". A child is coded as 1 (distressed) if their K6 score is ≥ 5 (Prochaska et al., 2012), CES-D8 score is ≥ 7, or CES-D20 score is ≥ 16 (Bi et al., 2023).

Table C.1. Items of psychological scales

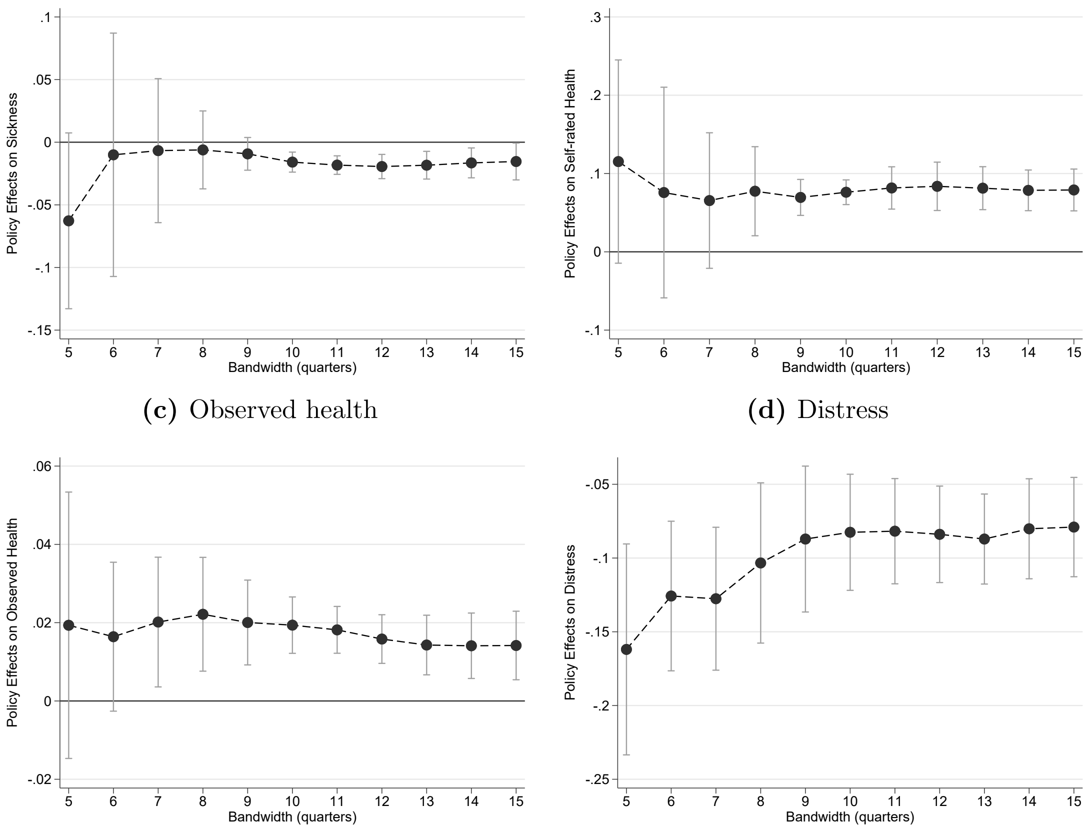
	2010	2012	2014	2016	2018	2020
K6: Please select according to your statuses in the past month.						
(1) Feel depressed and cannot cheer up.	X		X			
(2) Feel nervous.	X		X			
(3) Feel agitated or upset and cannot remain calm.	X		X			
(4) Feel hopeless about the future.	X		X			
(5) Feel that everything is difficult.	X		X			
(6) Think life is meaningless.	X		X			
CES-D: Please select according to your statuses in the past week.						
(1) I am worried about some trivial things.		X		X		
(2) I have a poor appetite and do not want to eat.		X		X		
(3) I feel depressed despite the help from relatives and friends.		X		X		
(4) I find myself not worse than others.		X		X		
(5) I cannot concentrate on things.		X		X		
(6) I am in a low spirit.		X		X	X	X
(7) I find it difficult to do anything.		X		X	X	X
(8) I find the future promising.		X		X		
(9) I feel that I have been a loser for a long time.		X		X		
(10) I feel scared.		X		X		
(11) I cannot sleep well.		X		X	X	X
(12) I feel happy.		X		X	X	X
(13) I talk less than usual.		X		X		
(14) I feel lonely.		X		X	X	X
(15) I find that people are not friendly to me.		X		X		
(16) I have a happy life.		X		X	X	X
(17) I cried or I want to cry.		X		X		
(18) I feel sad.		X		X	X	X
(19) I find that others do not like me.		X		X		
(20) I feel that I cannot continue with my life.		X		X	X	X
Number of items	6	20	6	20	8	8

Notes: This table presents detailed items of the K6 scale and the CES-D scale and in which wave they were elicited.

C.2 Robustness Checks

Figure C.1. Sensitivity of results to bandwidth choices

(a) Sick **(b)** Self-rated health

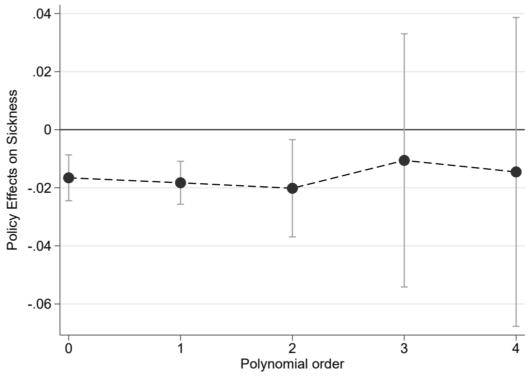


(c) Observed health **(d)** Distress

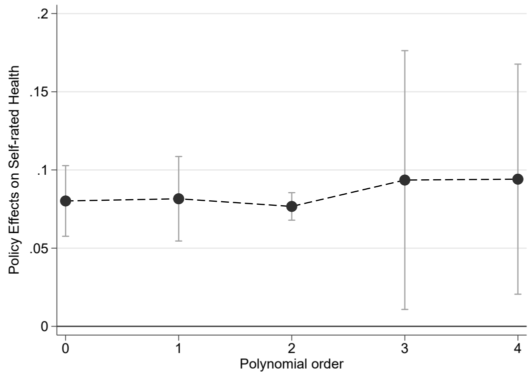
Notes: Each sub-graph reports coefficient estimates and confidence intervals for different bandwidths from 5 to 15 quarters. Each dot indicates the RD estimate using the specified bandwidth. Capped spikes represent 90% confidence intervals of the estimates.

Figure C.2. Sensitivity of results to different orders of polynomial

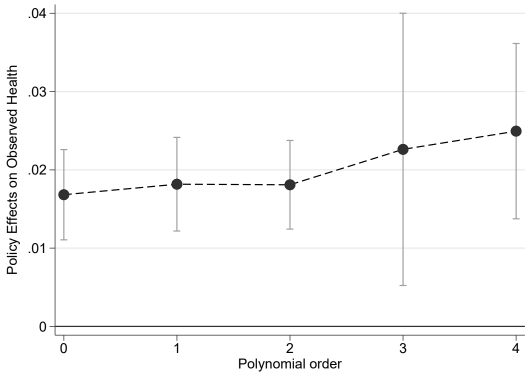
(a) Sick



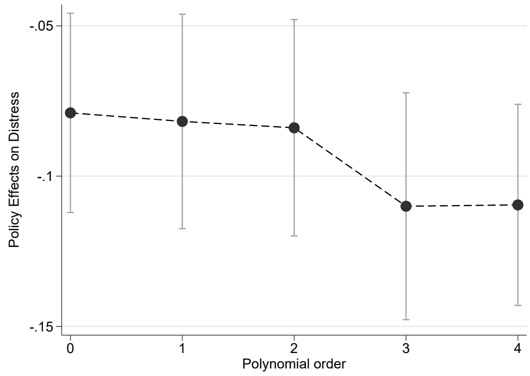
(b) Self-rated health



(c) Observed health



(d) Distress



Notes: Each dot represents the RD estimate using the specified order of RD polynomial. Capped spikes represent 90% confidence intervals of the estimates.

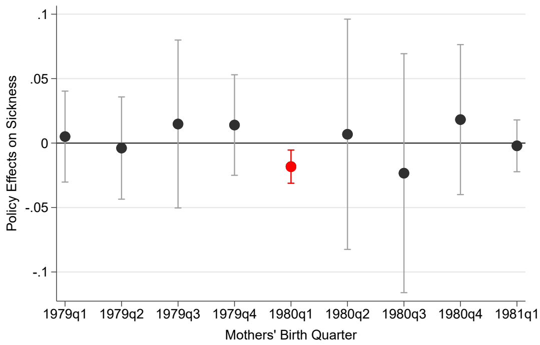
Table C.2. Robustness to different specifications

	Linear Interaction (1)	Quadratic Interaction (2)	No weights (3)	Panel weights (4)	Donut 1Q (5)
Panel A. Dependent variable is: Sick					
Policy	-0.022	-0.018	-0.023	0.014	-0.006
	(0.012)	(0.048)	(0.021)	(0.012)	(0.021)
Policy × Running quarters	-0.006**	-0.005			
	(0.002)	(0.016)			
Policy × Running quarters^2		0.000			
		(0.003)			
Mean	0.263	0.263	0.263	0.253	0.263
Observations	3,056	3,056	3,274	2,907	2,938
R^2	0.104	0.104	0.089	0.106	0.101
Panel B. Dependent variable is: Self-rated health					
Policy	0.079***	0.143*	0.091**	0.193**	0.057
	(0.013)	(0.057)	(0.032)	(0.073)	(0.046)
Policy × Running quarters	0.013	0.059			
	(0.008)	(0.033)			
Policy × Running quarters^2		0.003			
		(0.005)			
Mean	0.379	0.379	0.379	0.378	0.378
Observations	1,176	1,176	1,266	1,251	1,130
R^2	0.123	0.124	0.114	0.211	0.145
Panel B. Dependent variable is: Observed health					
Policy	0.019**	0.043***	0.006	0.009	0.041***
	(0.006)	(0.010)	(0.011)	(0.009)	(0.010)
Policy × Running quarters	0.003	0.031**			
	(0.002)	(0.009)			
Policy × Running quarters^2		0.001			
		(0.001)			
Mean	0.980	0.980	0.980	0.983	0.980
Observations	1,564	1,564	1,666	1,529	1,500
R^2	0.070	0.075	0.068	0.167	0.081
Panel B. Dependent variable is: Distress					
Policy	-0.083**	-0.103***	-0.089**	-0.217**	-0.078*
	(0.023)	(0.020)	(0.023)	(0.067)	(0.033)
Policy × Running quarters	0.005	0.021*			
	(0.005)	(0.010)			
Policy × Running quarters^2		-0.002			
		(0.001)			
Mean	0.123	0.123	0.123	0.125	0.125
Observations	1,175	1,175	1,265	1,250	1,129
R^2	0.254	0.255	0.237	0.251	0.263

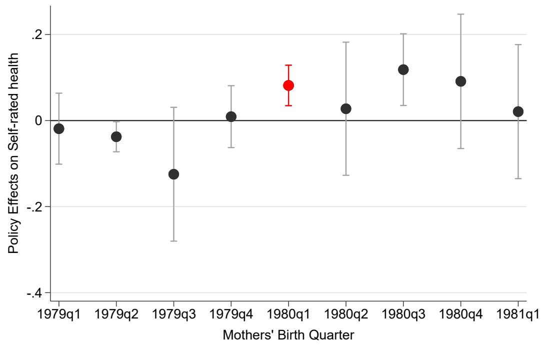
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The policy cut-off is 1980Q1. Regressions include children of mothers born within 11 quarters around the policy cut-off. CFPS panel weights are used in column (4).

Figure C.3. Placebo 1980 Q1 cut-offs

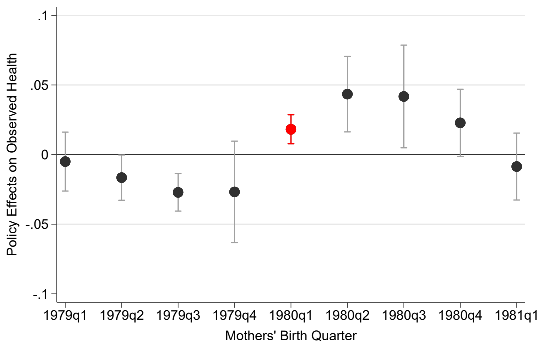
(a) Sick



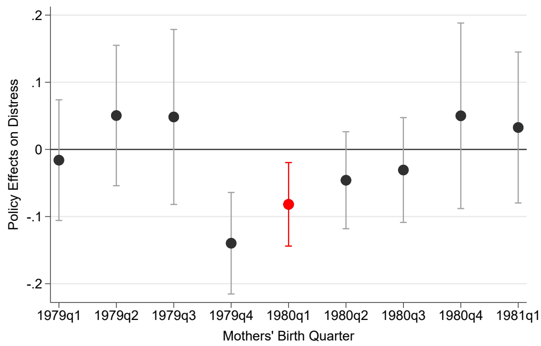
(b) Self-rated health



(c) Observed health



(d) Distress



Notes: This figure tests different policy cut-offs up to 4 quarters prior and post the policy cut-off employed in this paper, at a 1-quarter frequency. The policy cut-off we choose for this paper is 1980Q1, marked in red.

C.3 Results on child growth

Growth indicators. In addition to the main results, we also look at children's growth indicators, which are often used for younger kids to indicate their general health. We use the same specification as our main analysis but exclude controls for the children's gender and age, as the Z-scores already take these factors into account.

Table C.3 shows the effects of the policy on various indicators of child growth. The results generally show null effects on child growth indicators, except for a slightly significant increase in the probability of being overweight, as derived from the BMI Z-score. These additional results suggest that children born to mothers who were born right before or after the policy cut-off have similar growth patterns, except for a marginally higher probability of being overweight. This is generally consistent with some papers that also find null effects of sibling size on children's height and BMI (Zhong, 2014).

Table C.3. Growth indicators for children

	(1) Height-for-age Z-score	(2) Weight-for-age Z-score	(3) Body Mass Index Z-score	(4) Overweight	(5) Obese
Policy	-0.062	0.204	0.175	0.070*	0.004
	(0.039)	(0.111)	(0.157)	(0.034)	(0.015)
Mean	0.138	0.241	0.242	0.094	0.062
Observations	2,801	2,047	2,832	2,832	2,832
R^2	0.149	0.140	0.099	0.051	0.095

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. Z-scores are generated using the WHO Child Growth Standards, accounting for age and gender. Overweight is defined as a BMI-for-age Z-score above 2, while obese is defined as a BMI Z-score above 3.

Weight categories for mothers. Table C.4 shows the effects of the policy on the body mass index (BMI) and different weight categories for mothers. The policy has no significant effect on the BMI of mothers in general. However, we see a small but significant increase in the probability of being in the overweight category for mothers born after the policy cut-off date. This is consistent with the literature looking at the effects of the policy on health in middle age (Islam and Smyth, 2015; Wu and Li, 2012). Meanwhile, we see a reduction in the probability of being obese for these mothers, and null effects on the probability of being in the healthy weight or underweight categories.

These results suggest that the policy had a noticeable effect on the weight distribution of mothers, specifically by increasing the likelihood of being overweight. A possible explanation for the increase in the overweight category but not in obesity could be that the policy led to improved economic conditions and access to food and nutrition, which caused mothers to gain weight and move from a healthy weight to overweight. However, the same improvements in economic conditions and access to health care may have prevented the extreme weight gain that leads to obesity, explaining the decline in obesity rates. This shift in weight categories reflects nuanced changes in maternal health outcomes influenced by policy, highlighting the complex interplay between fertility policy and health behaviors. This effect on mothers' weight outcomes could also be transmitted to their children, as shown in the above Appendix Table C.3.

Table C.4. Weight categories for mothers

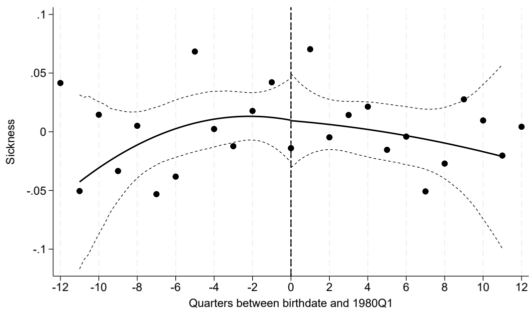
	(1) Body Mass Index	(2) Healthy weight	(3) Overweight	(4) Obese	(5) Underweight
Policy	0.005	-0.056	0.073***	-0.032*	0.015
	(0.119)	(0.030)	(0.016)	(0.014)	(0.023)
Mean	22.529	0.631	0.226	0.065	0.078
Observations	2,612	2,612	2,612	2,612	2,612
R^2	0.197	0.078	0.095	0.088	0.075

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. Healthy weight is defined as a BMI larger than or equal to 18.5 but smaller than 24. Overweight is defined as a BMI above 24, while obese is defined as a BMI above 28.

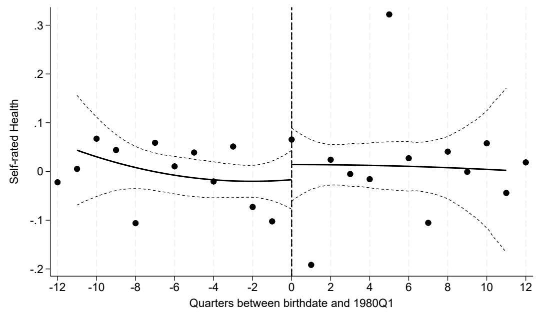
C.4 Additional tables and figures

Figure C.4. Quadratic polynomial: RD plots for all outcomes

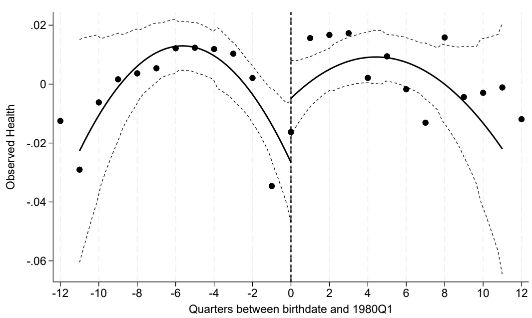
(a) Sick



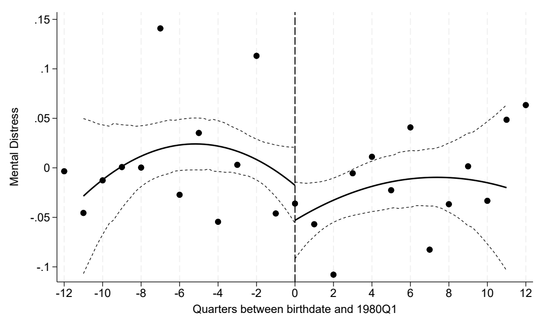
(b) Self-rated health



(c) Observed health



(d) Distress



Notes: The points depict binned residuals from a main regression of the outcome variable on a quadratic polynomial in birth quarter, along with other control variables. Solid lines display quadratic polynomial regression fit, separately estimated on each side of the cut-off, with dashed lines indicating 90% confidence intervals.

Table C.5. Effects from fathers' side: Family income and expenditure

	(1) Total Income	(2) Total Exp.	(3) Med. Exp.	(4) Public Ins.	(5) Commercial Ins.	(6) Commercial Ins. Spending
Policy	-0.211***	0.017	-0.245	-0.032*	-0.086**	-0.556*
	(0.036)	(0.076)	(0.135)	(0.013)	(0.027)	(0.229)
Mean	10.726	10.943	5.542	0.721	0.206	1.315
Observations	2,542	2,491	1,149	2,533	2,530	2,528
R^2	0.333	0.329	0.222	0.166	0.098	0.106

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the fathers' birth years. The first two are taken from family level expenditure and the rest are directly on children. Total family income comprises five components: wage income, total/net business income, property income, transfer income, and other income. We take natural logs of total income and expenditure (columns 1 and 2), medical expenditure (column 5), and commercial insurance spending (column 6). Public and commercial insurance in columns (3) and (4) are binary variables.

Table C.6. Policy effects on fathers' demographic characteristics

	(1) No siblings	(2) Number of children	(3) College+
Policy	-0.048	0.107**	0.037
	(0.032)	(0.033)	(0.089)
Mean	0.215	1.751	0.220
Observations	2,147	2,158	2,158
R^2	0.331	0.327	0.280

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the fathers' birth years. No siblings is a dummy variable and College+ is column (3) takes 1 if a father has a college degree or higher.

Table C.7. Policy effects on fathers' health status

	(1) Discomfort	(2) Chronic Disease	(3) Self-rated health	(4) Unhealthy	(5) Observed health	(6) Distress
Policy	-0.102***	-0.069	0.124***	0.012	-0.000	-0.021
	(0.017)	(0.047)	(0.022)	(0.025)	(0.008)	(0.036)
Mean	0.188	0.095	0.242	0.163	0.983	0.161
Observations	1,723	1,720	2,153	2,153	1,752	1,723
R^2	0.092	0.077	0.224	0.116	0.056	0.150

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the fathers' birth years. Discomfort takes 1 if a father reported physical discomfort in the last two weeks. Chronic disease is a dummy variable indicating whether a father was diagnosed with a chronic disease in the past six months. Self-rated health is a binary variable, with 1 indicating good health, while Unhealthy takes 1 if they rated themselves as very unhealthy. Interviewer-observed health is a binary variable, with 1 indicating good health. Distress is a binary variable where 1 indicates psychological distress.

Table C.8. Effects from fathers' side: Interaction between parents and children

	Interviewers' observation		Children's Response		Parents' Response			
	(1) Active Communication	(2) Care about Education	(3) Quarrel	(4) Heart-to-heart Talk	(5) Give up watching TV	(6) Discuss	(7) Homework Check	(8) TV Restriction
Policy	-0.007	0.015	-0.287	-0.432	0.010	-0.033	0.020	-0.013
	(0.018)	(0.017)	(0.165)	(0.520)	(0.023)	(0.027)	(0.022)	(0.020)
Mean	0.843	0.843	1.140	2.506	0.533	0.426	0.655	0.582
Observations	1,729	1,774	626	609	1,294	1,296	1,271	1,296
R^2	0.550	0.400	0.168	0.129	0.290	0.260	0.469	0.191

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the fathers' birth years. The first two variables are dummy variables, showing interviewers' observations on whether parents communicate with their child actively and on whether home environment indicates parents care about their child's education. The next two variables are only reported by those aged 9-15. Quarrel refers to the number of times children quarrelled with their parents last month (column 3). Heart-to-heart talk refers to the number of times children had a heart-to-heart talk with parents last month (column 4). The last four variables are all dummy variables constructed based on parents' responses: whether parents often give up watching TV to avoid disturbing their child (column 5), whether parents often discuss happenings at school with their child this semester (column 6), whether parents often ask their child to finish homework or check their child's homework (column 7) and whether parents restrict their child from watching TV or restrict the type of TV programs their child could watch (column 8).

Table C.9. Effects from mothers' side: Interaction between parents and children, restricted to children above 9 years old

	Interviewers' observation		Children's Response		Parents' Response			
	(1) Active Communication	(2) Care about Education	(3) Quarrel	(4) Heart-to-heart Talk	(5) Give up watching TV	(6) Discuss	(7) Homework Check	(8) TV Restriction
Policy	0.053***	0.023	0.788***	0.422	0.062	-0.113	-0.034	-0.062**
	(0.010)	(0.024)	(0.187)	(0.450)	(0.034)	(0.064)	(0.033)	(0.016)
Mean	0.803	0.801	1.335	2.527	0.545	0.430	0.670	0.577
Observations	1,088	1,114	1,090	1,022	1,423	1,426	1,425	1,426
R^2	0.598	0.425	0.127	0.127	0.300	0.280	0.499	0.176

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and clustered at the mothers' birth years. The first two variables are dummy variables, showing interviewers' observations on whether parents communicate with their child actively and on whether home environment indicates parents care about their child's education. The next two variables are only reported by those aged 9-15. Quarrel refers to the number of times children quarrelled with their parents last month (column 3). Heart-to-heart talk refers to the number of times children had a heart-to-heart talk with parents last month (column 4). The last four variables are all dummy variables constructed based on parents' responses: whether parents often give up watching TV to avoid disturbing their child (column 5), whether parents often discuss happenings at school with their child this semester (column 6), whether parents often ask their child to finish homework or check their child's homework (column 7) and whether parents restrict their child from watching TV or restrict the type of TV programs their child could watch (column 8).