Title
Essays in Labor Economics

Author
Lei, Ziteng

Publication Date
2022

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Essays in Labor Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Ziteng Lei

Committee in charge:

Professor Shelly Lundberg, Chair
Professor Peter Kuhn
Professor Clément de Chaisemartin

June 2022

The Dissertation of Ziteng Lei is approved.

Professor Peter Kuhn

Professor Clément de Chaisemartin

Professor Shelly Lundberg, Committee Chair

March 2022

Essays in Labor Economics

by

Ziteng Lei

Acknowledgements

The completion of this dissertation would not have been possible without the help and support from many people. I would like first to thank my advisor Shelly Lundberg, who has always been extremely supportive throughout all stages of my research. I am deeply grateful for her patience, kindness and encouragement for me, and all her detailed and insightful feedback to my research. I would also like to thank Peter Kuhn and Clément de Chaisemartin, who have always been providing constructive and comprehensive comments and suggestions to my research from different perspectives. I am sincerely grateful for having them in my committee and learning from them throughout the past few years.

I would like to thank Kelly Bedard, Youssef Benzarti, Ted Bergstrom, Ignacio Esponda, Erik Eyster, Kyle Meng, Ryan Oprea, Cheng-Zhong Qin, Heather Royer, Dick Startz, Douglas Steigerwald, Alisa Tazhitdinova, Gonzalo Vazquez-Bare, Sevgi Yuksel, along with many other faculty members at UC Santa Barbara. I would like to thank all the staff members in the department, especially the graduate advisor Mark Patterson, who has always been very patient and helpful during the past few years. And I would like to thank all my colleagues and friends in the department, especially Woongchan Jeon, Christopher Malloy, and Sarah Robinson, who have been great office-mates and friends in every aspect. It is a great pleasure for me to have had the opportunity to learn and work in such an excellent place.

I would also like to thank my co-authors, in particular Le Kang, Yang Song and Peng Zhang. It is an honor to have had the opportunity to work with them.

Finally, I would like to thank my family and all my friends for their encouragement and support throughout the journey. I may not be where I am today without you all.

Curriculum Vitæ
Ziteng Lei

Education

2022 (Expected)	Ph.D. in Economics, University of California, Santa Barbara.
2018	M.A. in Economics, University of California, Santa Barbara.
2017	Bachelor of Economics, Peking University.

Research Interests

Labor Economics, Environmental Economics, Applied Econometrics

Publications

"Impact of Air Pollution on Short-term Movements: Evidence from Air Travels in China", with Shuai Chen, Yuyu Chen and Jie-Sheng Tan-Soo, *Journal of Economic Geography 20, no. 4 (2020): 939-968.*

"Vulnerable Boys: Short-term and Long-term Gender Differences in the Impacts of Adolescent Disadvantage", with Shelly Lundberg, *Journal of Economic Behavior & Organization 178 (2020): 424-448.*

"Chasing Clean Air: Pollution-Induced Travels in China", with Shuai Chen, Yuyu Chen and Jie-Sheng Tan-Soo, *Journal of the Association of Environmental and Resource Economists 8, no. 1 (2021): 59-89.*

"Short-run and Long-run Effects of Peers from Disrupted Families", *Forthcoming, Journal of Population Economics*

Working Papers

"Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes", with Le Kang, Yang Song and Peng Zhang

"Are Bartik Regressions Always Robust to Heterogeneous Treatment Effects?", with Clément de Chaisemartin

"Family Companionship and Elderly Suicide: Evidence from the Chinese Lunar New Year", with Hanming Fang, Liguo Lin and Peng Zhang [NBER WP 28566]

Referee

Journal of Economic Behavior & Organization, Journal of Human Resources

Teaching Experience

Teaching Assistant, Department of Economics, University of California, Santa Barbara

ECON 1-Principles of Microeconomics	Fall 2017, Winter 2018
ECON 100B-Intermediate Microeconomic Theory	Winter 2021
ECON 101-Intermediate Macroeconomic Theory	Spring 2018
ECON 210A-Graduate Microeconomics	Fall 2018, Fall 2019, Fall 2020, Fall 2021
ECON 210B-Graduate Microeconomics	Winter 2019
ECON 210C-Graduate Microeconomics	Spring 2020, Spring 2021

Selected Honors and Awards

Louise & Stephen Pahl Graduate Fellowship in Social Sciences	2018-2022
Gretler Fellowship	2019
Research Quarter Fellowship	2019-2020, 2021-2022
Award on "Scholarship, Teaching, Contributions to the Department, or to the Community"	2020-2021
Distinction in Ph.D. Preliminary Examination in Microeconomics	2018
Distinction in Ph.D. Preliminary Examination in Macroeconomics	2018
Distinction in Ph.D. Preliminary Examination in Econometrics	2018

Permissions and Attributions

1. The content of Chapter 1 and Appendix A is the result of a collaboration with Le Kang, Yang Song and Peng Zhang.
2. The content of Chapter 2 and Appendix B is the result of a collaboration with Shelly Lundberg, and has previously appeared in the *Journal of Economic Behavior & Organization*. The journal and Elsevier allow re-use of materials in a student's thesis/dissertation with proper attribution. The published version is available at `https://doi.org/10.1016/j.jebo.2020.07.020`.
   Lei, Ziteng, and Shelly Lundberg. "Vulnerable Boys: Short-term and Long-term Gender Differences in the Impacts of Adolescent Disadvantage." Journal of Economic Behavior & Organization 178 (2020): 424-448.
3. The content of Chapter 3 and Appendix C has previously appeared in the *Journal of Population Economics*. The journal and Springer Nature allow re-use of materials in a student's thesis/dissertation with proper attribution. The published version is available at `https://doi.org/10.1007/s00148-021-00839-0`.
   Lei, Ziteng. "Short-run and long-run effects of peers from disrupted families." Journal of Population Economics (2021).

Abstract

Essays in Labor Economics

by

Ziteng Lei

This dissertation consists of three essays that investigate human capital accumulation, a crucial topic in the field of labor economics, and gender differences in the complex human capital accumulation process in school.

The first chapter, based on joint work with Le Kang, Yang Song and Peng Zhang, documents gender differences in reactions to failure in the National College Entrance Exam, an extremely high-stakes exam that solely determines college admission outcomes for almost all teenagers in China. Using unique administrative data in Ningxia Province and a regression-discontinuity design, we find that students who score just below the tier-2 university cutoff have an eight percentage point higher probability of retaking the exam in the next year, and that retaking improves exam performance substantially. However, the increase in retake probability when confronting the failure of scoring just below the cutoff is more pronounced for men than for women (11 percentage points vs. 5.5 percentage points). The gender disparity in the tendency to retake has important implications for exam performance, college enrollment, and labor market outcomes.

The second chapter, based on joint work with Shelly Lundberg, is motivated by the fact that the growing gender gap in educational attainment between men and women has raised concerns that the skill development of boys may be more sensitive to family disadvantage than that of girls. Using the National Longitudinal Study of Adolescent to Adult Health (Add Health) data we find, as do previous studies, that boys are more likely to experience increased problems in school relative to girls, including suspensions and reduced

educational aspirations, when they are in poor quality schools, less-educated neighborhoods, and father-absent households. Following these cohorts into young adulthood, however, we find no evidence that adolescent disadvantage has stronger negative impacts on long-run economic outcomes such as college graduation, employment, or income for men, relative to women. We do find that father absence is more strongly associated with men's marriage and childbearing and weak support for greater male vulnerability to disadvantage in rates of high school graduation. An investigation of adult outcomes for another recent cohort from the National Longitudinal Survey of Youth, 1997 produces a similar pattern of results. We conclude that focusing on gender differences in behavior in school may not lead to valid inferences about the effects of disadvantage on adult skills.

The third chapter investigates the short-run and long-run effects of exposure to peers from disrupted families in adolescence. Using the National Longitudinal Study of Adolescent to Adult Health (Add Health) data, I find that girls are mostly unaffected by peers from disrupted families, while boys exposed to more peers from disrupted families exhibit more school problems in adolescence and higher arrest probabilities, less stable jobs and higher probabilities of suffering from financial stress as young adults. These results suggest negative effects on non-cognitive skills but no effect on cognitive skills, as measured by academic performance. The dramatic increase in family disruption in the United States should thus receive more attention, as the intergenerational mobility and inequality consequences could be larger than anticipated as a result of classroom spillovers.

Contents

Chapter 1

Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes

1.1 Introduction

Gender disparities in educational outcomes and labor market outcomes have attracted increasing attention. Previous studies have documented that gender differences in non-cognitive traits and attitudes, such as willingness to compete, pressure tolerance, risk aversion, and confidence, may explain an important part of the gender gaps in educational choices and labor market outcomes (see a review article by Delaney & Devereux 2021). However, less is known about the gender difference in reactions to failure and its mechanisms and implications, especially in settings of high-stakes competitions. As people confront competitions throughout their career for college admission, jobs and promotions, failures and setbacks in these competitions are not uncommon for most people. Different responses to failure, such as whether to try again in subsequent competitions or give up, may lead to very different educational achievements and career paths. Therefore, understanding the gender differences in responses to failure is crucial for understanding gender gaps in educational and labor market outcomes.

To the best of our knowledge, this is the first study that documents gender differences in responses to failure in an admission-relevant exam for colleges and for a broad group of individuals. Specifically, we study how men and women respond differently to failures in the National College Entrance Exam (hereinafter referred to as NCEE), an annual exam that solely determines the admission of almost all students into higher education institutions in China. The setting is important for at least two reasons. First, many countries use high-stakes standardized tests to rank students for college admission, and retaking such exams when confronting failures are not uncommon. Studying gender disparities in response to failures in these admission-relevant exams and the related consequences is helpful to understand gender gaps in college enrollment and labor market outcomes. Second, since almost everyone needs to take the NCEE to get into colleges in China, our setting alleviates concerns over sample selection in the sense that individuals who do not like competition may choose not to participate in the competitions in the first place.

Estimating gender differences in responses to failure in the NCEE is challenging, as failures are typically subjective and not randomly assigned. To overcome this challenge, we exploit a unique feature of the NCEE, which is the exogenously determined cutoff for different tiers of universities. The over 2,000 universities in China are classified into four tiers, with NCEE score cutoffs determining the eligibility of application for universities in each tier.[1] We show evidence that these cutoffs are exogenously determined,

[1] The higher education institutions in China are classified into tier-1 key universities, tier-2 regular universities, tier-3 universities, and tertiary technical colleges, by the central government. Only students with the NCEE scores above the tier cutoff can apply for universities in that specific tier. See Section

and students cannot self-select around the cutoff. Around 10 million students take the NCEE to compete for admissions to the highly selective universities each year, with only around 25% of students receiving scores that make them eligible to apply for high-quality universities in the top two tiers.

Our empirical strategy thus is to use the gender differences in the discontinuity in retake probability around the tier-2 cutoff to causally identify gender differences in responses to the arrival of a plausibly exogenous failure.[2] To do so, we obtain a unique dataset that covers the universe of NCEE takers in Ningxia Province during 2002-2010. Before we focus on gender differences, we first show that the tier-2 cutoff indeed generates a large discontinuity in the probability of retaking the NCEE regardless of gender. Specifically, students who score just below the tier-2 cutoff, a signal of entering good universities and educational success, have an eight percentage point higher probability of retaking the NCEE in the next year, almost doubling that of those who score just above the cutoff. In addition, we show that retaking the NCEE generates large returns in terms of exam performance and educational success, since it increases the test scores by 0.47 standard deviations and the relative ranking among competitors by 11 percentage points. These improvements amount to 2.7-5.7% higher wage offer for the first job after college under a simple back-of-the-envelope analysis. These results indicate that the response to failure, specifically whether choosing to retake the exam or not, has crucial consequences for college admission and possibly future labor market prospects.

We then focus on gender differences in reactions to the failure of scoring below the tier-2 cutoff. We find consistent evidence that the cutoff-induced retakes from the regression

1.2 for more discussions.

[2]We focus on the tier-2 cutoff because for students in Ningxia, admission into a tier-2 university is generally regarded as an educational success compared with tier-3 universities or technical colleges. By contrast, falling below the tier-1 cutoff, which indicates that the student is still eligible for admission into tier-2 universities, is much less viewed as a failure in the NCEE. Consequently, the decline in retake probability at the cutoff is dramatic for the tier-2 cutoff, but much less pronounced for the tier-1 cutoff. See Section 1.3 for more discussions.

discontinuity design, which reflect the desire to participate in the competition again inspired by the failure of scoring below the cutoff, are much more pronounced for men than for women. Specifically, the increase in retake probability when falling just below the tier-2 cutoff for males is twice as large as for females (11 vs. 5.5 percentage points, respectively), and the gender differences are statistically significant and robust across various specifications.

We detect several important mechanisms that can help explain why women are less likely to retake the NCEE than men when scoring just below the cutoff. We start by testing whether the returns to retake differ across gender. Surprisingly, we find that the causal returns to retake in terms of exam outcomes for women are similar to or sometimes even higher than those for men. In addition, such gender differences in returns cannot be explained by students rationally self-selecting into retake based on returns. These results suggest that gender differences in returns to retake are unlikely to explain the gender differences in the propensity to retake.

In addition, we find that the gender differences are large and of similar magnitude for individuals from urban and rural households, of different ethnicity, from high-quality and low-quality high schools, from rich and poor counties, and from places with high and low levels of sex ratio. These results show that the gender differences in reactions to failure are not driven by certain groups, but are pronounced for all types of individuals. They also suggest that gender differences in benefits and costs of retake, as well as in social norms and family support, are unlikely to fully explain our results. By contrast, gender differences in non-cognitive traits such as causal attributions, confidence, and risk preferences, are likely to explain our results.

Our paper contributes to three strands of the literature. First, we contribute to the broad literature on gender differences in educational choices and competitions (Niederle & Vesterlund 2007, Buser et al. 2014, Flory et al. 2015, Berlin & Dargnies 2016, Buser et al. 2017, Reuben et al. 2017, Astorne-Figari & Speer 2019, Cai et al. 2019),[3] and specifically on the growing literature that focuses on gender differences in the dynamic evolution of willingness to compete in response to winning and losing (Ellison & Swanson 2018, Buser & Yuan 2019, Landaud & Maurin 2020, Wasserman 2020, Fang et al. 2021). These studies have documented that when confronting failures in competitions, women are less likely to choose competition again than men in lab experiments and in low-stakes high school math competitions in the Netherlands and the U.S. (Ellison & Swanson 2018, Buser & Yuan 2019), in low-stakes Rubik's Cube competitions (Fang et al. 2021), in the entrance exam of highly selected elite science graduate programs in France (Landaud & Maurin 2020), and in local elections in California (Wasserman 2020).

Our paper adds to this strand of literature in three important ways. First, we focus on high-stakes admission-relevant exams, which most countries use to select students for college admission. Thus, our findings can directly speak to gender gaps in college enrollment and possibly future labor market. Second, previous studies focus on a selected group in the sense that individuals who do not like competition may choose not to participate in the competition in the first place. Our setting, however, can greatly alleviate the concern of sample selection because almost everyone needs to take the NCEE to get into colleges in China. Our results can thus enhance the external validity of prior findings substantially. Lastly, our results on the gender differences in returns to retake suggest that differential return is unlikely to be an important driver of the gender differences in the tendency to retake, and our rich tests on various heterogeneous groups further improve our understanding of the potential mechanisms of the gender differences in reactions to failure.

Second, we contribute to a growing body of literature that emphasizes the impor-

[3] Cai et al. (2019) look at how males and females perform differently between a mock exam and the actual NCEE in Anxi County, China. We look at how males and females respond differently when they confront a failure in the NCEE.

tance of non-cognitive skills, such as patience, self-control, and grit on human capital accumulation (Heckman et al. 2006, Borghans et al. 2008, Moffitt et al. 2011, Golsteyn et al. 2014, Sutter et al. 2013, Alan et al. 2019). The NCEE admission cutoff provides a quasi-experimental variation in failure, and thus allows us to study how people react to failure in an extremely high-stakes setting. Our study suggests that grit may play an important role in human capital accumulation, which is consistent with Alan et al. (2019), who demonstrate the importance of grit in a randomized educational intervention program. In addition, our study shows that grit may be different between males and females in a high-stakes environment, which may fundamentally change the educational and career paths for all teenagers in China, and possibly for teenagers in countries that heavily rely on standardized tests for college admission.

Lastly, we contribute to the literature on the causal effects of exam retakes, particularly in the high-stakes settings that are admission-relevant (Krishna et al. 2018, Zhang et al. 2019, Goodman et al. 2020). We find that retaking the NCEE can generate substantial returns despite its high opportunity cost of waiting for another year. More interestingly, although females are less likely to retake the NCEE than males, the returns to retake for females are similar to or sometimes even higher than males.

The remainder of our paper is organized as follows. Section 1.2 describes the institutional background of the NCEE in China and our data. Section 1.3 presents the results on the cutoff-induced discontinuity in retake probability and its causal effects. Section 1.4 presents the results on the gender differences in the NCEE retake behavior. Finally, Section 1.5 concludes.

1.2 Institutional Background and Data

1.2.1 Institutional Background

The NCEE, which is also commonly known as *gaokao*, is an annual examination held on June 7th and 8th that determines the admission of almost all students into higher education institutions in China.[4] The NCEE is highly competitive and often described as the "toughest exam in the world". Around 10 million students compete for the admission slots of the highly selective universities each year.[5] More than 2,000 universities in China are classified into four tiers, with NCEE score cutoffs determining the eligibility of application for each tier. It is estimated that less than 10% of candidates enroll in top-tier universities, and only less than 0.2% of exam takers will be admitted into China's top five universities (Cai et al. 2019). In addition, success in the NCEE has been taught to be the central goal for most students throughout the 12 years of schooling, and has been shown to lead to substantial improvement in labor market outcomes (Jia & Li 2021). Therefore, the NCEE is a high-stakes competition for almost the universe of students in China.

Students choose either the science or art (social science) track after the 10th grade, and they take the NCEE in their corresponding track. The most commonly adopted examination system across the provinces is the 3+X system: "3" represents the three compulsory subjects of Chinese, Mathematics, and English, each accounting for 150 of 750 of the total score. "X" represents the combined science subjects (Physics, Chemistry, and Biology) for science track or combined arts subjects (History, Geography, and Politics) for the art track, accounting for 300 of 750 of the total score. The exams are written and graded at the province level, and the test scores are only comparable within the

[4] Some provinces such as Shandong also have exams on June 9th.

[5] https://www.sohu.com/a/434396300_116509 (in Chinese).

province-year-track. In other words, students only compete with peers within the same province-year-track.

The admission process after the NCEE is hierarchical. The central government designates all higher education institutions into various tiers: tier-1 key universities, tier-2 regular universities, tier-3 universities, and tertiary technical colleges, according to the level of prestige. Tier-1 universities are the most selective universities with the best reputation in China, followed by tier-2 universities, and most tier-1 and tier-2 universities are public universities that are of high quality and charge minimal tuition (Jia & Li 2021). By contrast, tier-3 universities are mostly private universities that are of lower quality and charge high tuition. All tier-1 to tier-3 universities are four-year universities that grant bachelor's degrees, whereas tertiary technical colleges mostly offer programs lasting two to three years. Admission into tier-1 and tier-2 universities is generally considered as an educational success, while admission into tier-3 universities or tertiary technical colleges is often considered as less desirable and a failure in college admission (Zhang et al. 2019).

After the NCEE, provinces announce the track-specific admission cutoff scores for each university tier, based on the score distributions and university quotas assigned by the Ministry of Education. Students then apply to universities by submitting a rank order list.[6] The college assignment is organized sequentially by tier: tier-1 universities first finish their assignment, then tier-2 universities recruit, followed by tier-3 universities and tertiary technical colleges. Students who score above the cutoff score of a given tier are eligible to apply to the universities in that tier, but without a guarantee of being admitted into a school in that tier. The cutoff for tier-1 universities is set as the minimum score for admission into tier-1 universities, which is often lower than the actual admission

[6]Students are aware of the cutoff scores for each tier and their own test scores when they submit their applications in our sample period. See Ha et al. (2020) for more discussions on the timing of the college application submission in China.

cutoff scores for most tier-1 universities. For example, a student scoring just above the tier-1 cutoff who lists only super selective universities may not be admitted into any tier-1 university because her score is lower than the admission cutoffs for the universities on her rank order list. Students scoring below the cutoff score of a given tier will not be eligible to apply for any university in that tier.

If a student is unsatisfied with the exam and admission outcomes, then she can choose not to enroll in the assigned college and retake the NCEE next year, regardless of whether she is currently admitted into a program. As the NCEE is held annually, she must wait a year for the next take. Retakers will be marked so in the administrative records but face no advantages or disadvantages in the competition. There is no official restriction on the number of times to take the NCEE, but taking the NCEE more than two times is rare.

1.2.2 Data

Our administrative data include the test scores and demographic information for the universe of NCEE takers in the Ningxia Province (or Ningxia Hui Autonomous Region) from 2002 to 2010. Our data only have the total test score and do not contain detailed test scores by subject.[7] Ningxia is a small province in China, with around 7 million population. Recently, there are around 60,000 NCEE takers each year in Ningxia, and the number of NCEE takers is comparable to direct-controlled municipalities such as Beijing and Shanghai.[8] We also hand-collect the year-track cutoff points for the tier-1 and tier-2 universities in Ningxia Province from publicly available records.[9]

[7]The test score discussed in this paper is the total score for admission purpose, which is the raw test score plus the "bonus scores" for the students. For example, students of minority ethnicity in Ningxia get "bonus scores" because of their ethnicity. As these "bonus scores" are usually still applicable if they retake the NCEE in the next year, it will not confound the decision to retake.

[8]https://www.163.com/dy/article/FGP06FE50516EN5U.html (in Chinese).

[9]Admission into tier-3 universities is much less competitive that 40% to 50% of students are eligible for a tier-3 or better university (Cai et al. 2019). In addition, we are unable to find complete public

In order to identify whether NCEE takers retake the exam in the following year and their exam performance, we match observations in the two consecutive years based on the name identifier (which uniquely identifies a full name), exact date of birth, gender, ethnicity (Han/Hui/other ethnicity) and exam track (science/art). Individuals who are matched with the observations in the next year are defined to have retaken the NCEE in the next year.[10] Observations with identical information on the variables listed above within each year are dropped from the sample (approximately 0.1% of the sample) as they cannot be uniquely identified. Our final sample consists of 362,592 observations of NCEE takers from 2002 to 2009 and contains information on their exam performance, whether they retake the NCEE in the next year, and if so, their exam performance for the retake exam.[11]

1.3 Cutoff-Induced NCEE Retakes and the Effects on Exam Outcomes

1.3.1 Empirical Strategy

We first investigate the effects of the failure of scoring just below the cutoff and the causal effects of retaking the NCEE on exam outcomes. We focus on gender differences in Section 1.4. To make exam outcomes comparable across different years, we standardize

records of the cutoff points for the tier-3 universities during the sample period. Therefore, we do not focus on the tier-3 cutoffs in this paper.

[10]One may be concerned that our approach does not fully capture the retake behavior of students. For example, if a student chooses to move to another province to retake the NCEE, then she could not be detected in our sample. However, such possibility is unlikely to invalidate our results for two reasons. First, the hukou restrictions for the NCEE takers prevent students from arbitrarily choosing the province to take the NCEE. Second, even if the student can move to another province to take the NCEE, she will likely make such choice before her first take of the NCEE rather than doing so for retakes. As the exam content is often not the same across provinces, moving for retakes is very risky.

[11]Year 2010 is excluded from our analysis because we do not have the data for the next year, and are unable to identify whether the NCEE takers in year 2010 retake the exam in the next year or not.

the test score within each year-track, with a mean of 0 and a standard deviation of 1. We also consider an alternative measure of exam outcomes, the relative ranking of the test score, which measures the proportion of students with lower test scores within the same year-track. This measure is admission-relevant because it is the relative position among all competitors within the same year-track that determines the admission outcomes.

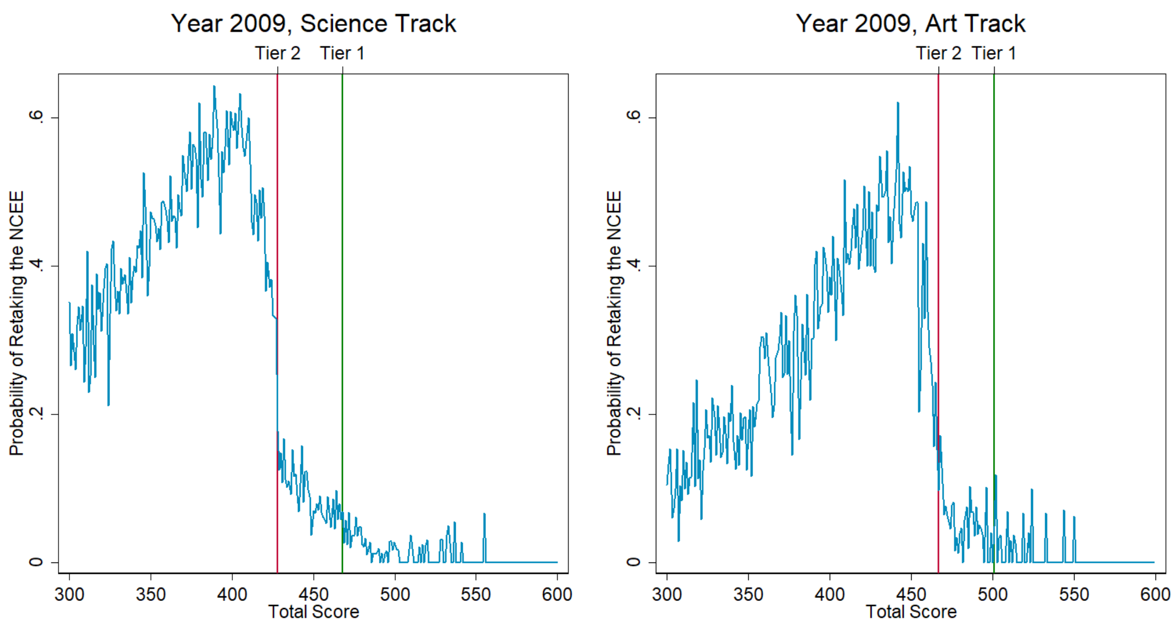
The propensity to retake the NCEE in the next year may be strongly correlated with unobserved student characteristics, such as inherent ability and risk preferences, and these characteristics may also be correlated with exam outcomes. In addition, students who choose to retake the NCEE may be a selective group and very different from the general population. To address endogenous retaking, we exploit the tier cutoffs for university admission and use a regression-discontinuity design to estimate the causal effects of retaking on exam outcomes. An important feature is that the tier cutoffs are exogenously determined by the score distribution and the quota assigned by the Ministry of Education each year. Students are not able to predict the exact cutoff scores, or to manipulate their test scores to be above the cutoffs. We provide evidence in Section 1.3.2.

Figure 1.1 plots the probability of retaking the NCEE in the next year against the test score for NCEE takers of the year 2009, for science track and art track separately. The retake probability measures the proportion of NCEE takers at each score that choose to retake in the next year. The patterns of the results are very similar for other years in our sample period. It is evident that there is a dramatic decline in retake probability at the tier-2 university cutoff, particularly for students in the science track. The retake probability is much lower for students around the tier-1 cutoff, and the decline in retake probability at the tier-1 university cutoff is much less pronounced.[12] This is because

[12]The tier-1 cutoff is generally higher than the tier-2 cutoff by 30-60 points, depending on the year and the exam track, and the cutoff is extremely selective and only 10% of students score above the cutoff.

for students in Ningxia, admission into a tier-2 university is generally regarded as an educational success compared with tier-3 universities or technical colleges (Zhang et al. 2019). By contrast, just falling below the tier-1 cutoff, which indicates that the student is still eligible for admission into tier-2 universities, is much less viewed as a failure in the NCEE. Therefore, we focus on the tier-2 cutoff for the rest of the paper.

Figure 1.1: The Relationship between the Probability of NCEE Retaking and Test Scores



Notes: This figure plots the probability of retaking the NCEE in the next year against the test score for NCEE takers of year 2009, for science track (left) and art track (right) separately. The retake probability measures the proportion of NCEE takers at each score that choose to retake in the next year. The lines in each panel of the figure represent the cutoff scores for tier-2 (left) and tier-1 (right) university admission for each track.

To examine how falling below the tier-2 university cutoff affects the retaking behavior, we estimate the following specification:

$$\begin{aligned} Retake_{i,y,tr} &= \beta\mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) + \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr}) \\ &+ \gamma_2\mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \end{aligned} \tag{1.3.1}$$

where $Retake_{i,y,tr}$ is a binary indicator for whether individual i in year y and track tr (science or art) retakes the NCEE next year. $Score_{i,y,tr}$ is the test score of individual i, and $Cutoff_{y,tr}$ is the cutoff score for tier-2 university admission that varies across year-track. The indicator function $\mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr})$ equals 1 if the test score is below the cutoff. We include a function of the running variable, $Score_{i,y,tr} - Cutoff_{y,tr}$, the distance between the test score and the cutoff, and its interaction with the indicator of below the cutoff. We consider linear and quadratic functions in this parametric specification, as well as the local polynomial non-parametric estimation and inference procedure (Calonico et al. 2014). In the parametric specifications, we control for a set of individual characteristics X_i, including gender, ethnicity, age, household registration (hukou) status, and whether the individual is a first-time taker of the NCEE. Year-by-track fixed effects $\mu_{y,tr}$ are also controlled. For the baseline, we use a 15-point bandwidth and uniform kernel weights. Our results are robust to alternative bandwidths and kernel weights.

The standard errors are two-way clustered at the individual identifier level and the high school-year level. The former accommodates the fact that the same individual may appear multiple times in our estimation sample.[13] For example, if a student retakes once after her first take in the NCEE, and her scores are within the 15-point bandwidth around

[13]The individual identifier is generated based on the name identifier, exact date of birth, gender, ethnicity, and exam track. It uniquely identifies an individual within the sample.

the cutoffs in both years, then she will enter the estimation sample twice.[14] The latter allows arbitrary error correlation between schoolmates in the same school cohort. We follow the recommendation of Kolesár & Rothe (2018) and do not cluster the standard errors by the discrete running variable. However, the results are very similar when the standard errors are clustered at the running variable level (Lee & Card 2008).

1.3.2 Effects of Falling Below the Tier-2 Cutoff on Retake Probability

Table 1.1 shows the summary statistics of the individual characteristics and the indicator of retaking the NCEE in the next year. Column (1) shows the summary statistics for the full sample, and column (2) shows the summary statistics for observations within the 15-point bandwidth, which is our RD estimation sample. Columns (3) and (4) show the summary statistics for the observations below and above the tier-2 cutoff, both still within the 15-point bandwidth. One can find that students below the tier-2 cutoff are more likely to retake the NCEE next year than those above the tier-2 cutoff. Overall, these summary statistics show that retaking the NCEE is not an uncommon choice for students—28% of the NCEE takers (20% for the RD sample) choose to retake next year. The retake probability is also very stable over time in our sample period.

[14]Approximately 92.7% of observations within the 15-point window are individuals that only appear once. Approximately 3.6% of the individuals within the 15-point window appear twice. Less than 0.1% of the individuals appear more than two times in the 15-point window.

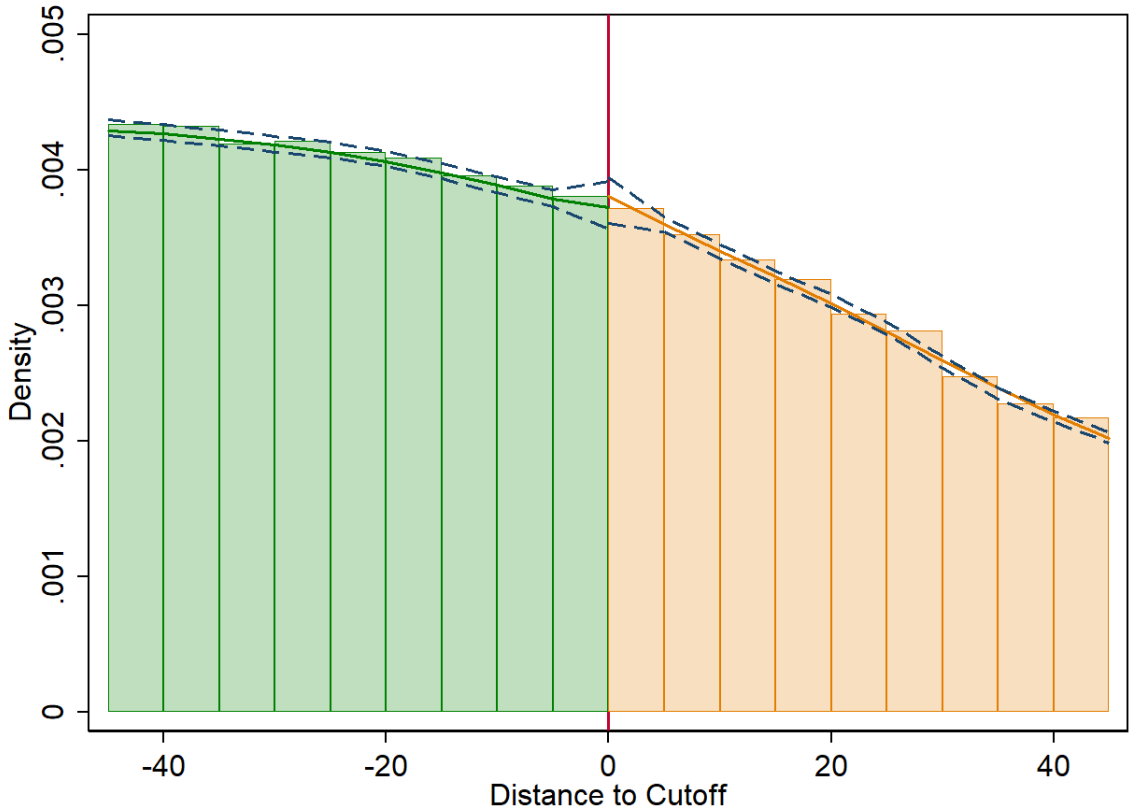
Table 1.1: Summary Statistics

	(1)	(2)	(3)	(4)
Sample	Full	[-15,15]	[-15,0)	[0,15]
Observations	362,592	41,477	21,123	20,354
	Mean (S.D.)			
Male	0.52	0.51	0.51	0.51
	(0.50)	(0.50)	(0.50)	(0.50)
Ethnicity: Han	0.78	0.79	0.79	0.79
	(0.41)	(0.41)	(0.41)	(0.40)
Ethnicity: Hui	0.20	0.19	0.19	0.19
	(0.40)	(0.39)	(0.39)	(0.39)
Urban	0.45	0.46	0.44	0.47
	(0.50)	(0.50)	(0.50)	(0.50)
First-Time Taker	0.73	0.57	0.56	0.58
	(0.44)	(0.50)	(0.50)	(0.49)
Age	19.15	19.16	19.19	19.12
	(1.23)	(1.24)	(1.24)	(1.23)
Retake	0.28	0.20	0.31	0.08
	(0.45)	(0.40)	(0.46)	(0.27)

Notes: This table shows the summary statistics of individual characteristics and the indicator of retaking the NCEE in the next year. Column (1) is for the full sample. Column (2) is for the sample within the 15-point bandwidth around the tier-2 cutoff. Column (3) is for the sample in column (2) that is below the cutoff. Column (4) is for the sample in column (2) that is above or equal to the cutoff.

Before presenting our main results, we present evidence to support the validity of our regression discontinuity design. The density distribution of the running variable around the tier-2 cutoff is shown in Figure 1.2. We apply the manipulation testing procedure proposed by Cattaneo, Jansson & Ma (2018) and obtain a p-value of 0.82, suggesting that there is no evidence of discontinuous density in test scores around the tier-2 cutoff. This confirms our research design because the cutoffs are determined after the NCEE, and students do not have the ability to sort around the cutoffs.

Figure 1.2: Test of Running Variable Density Smoothness around Tier-2 Cutoff



Notes: This figure plots the density of the running variable (the distance to the tier-2 cutoff score) following the manipulation testing procedure in Cattaneo, Jansson & Ma (2018). The bars in this figure represent the density distribution of the running variable over 5-point bins. The straight lines represent the estimated density to the left and to the right of the cutoff using the local polynomial density estimators proposed in Cattaneo et al. (2020). The dashed lines represent the lower and upper bounds of the 95% confidence interval for the estimated density.

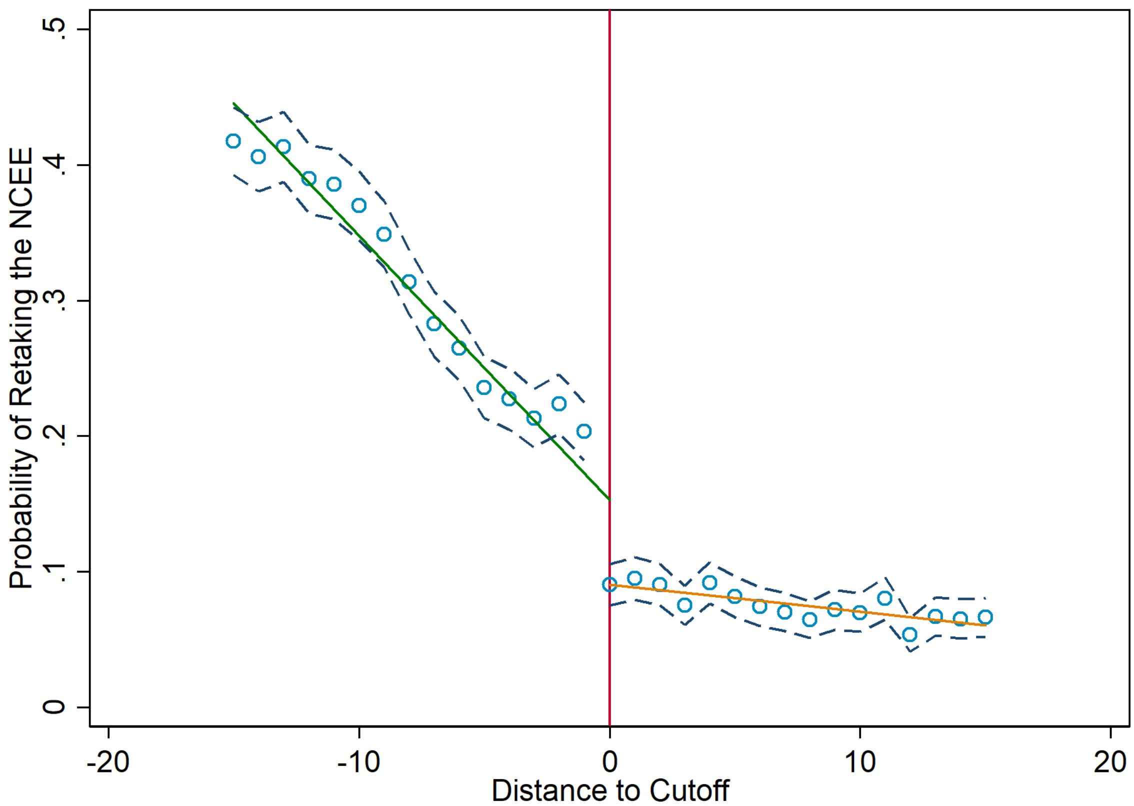
We also plot the individual characteristics of students against the distance to the cutoff in Figure A.1. There is no substantial discontinuous jump for these pre-determined characteristics at the cutoff. The estimation results for the balancing tests are shown in Table A.1. Indeed, there is no consistent evidence showing that a pre-determined characteristic has a substantial discontinuity at the cutoff for both linear and quadratic control specifications.[15] As explained above, students do not have the ability to sort around the cutoffs because of the institutional setting, and there is no reason that students of certain characteristics are more likely to appear on one side of the cutoff. Note that the graders have no information on students and the grading process is highly regulated, and thus discrimination based on individual characteristics is not possible.

Figure 1.3 plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score.[16] There is a notable discontinuity in retake probability around the cutoff point. The retake probability is close to 10% and relatively stable above the cutoff point, but ranges from 20% to 40% below the cutoff point. The estimated discontinuity effect without any covariates is 0.081 when using the local polynomial non-parametric estimation and inference procedure in Calonico et al. (2014), with a robust 95% confidence interval [0.051, 0.095]. Table 1.2 presents the results using the parametric specification (Equation 1.3.1), for both linear and quadratic controls. The results are consistent and show that falling below the tier-2 cutoff increases the probability of retaking the NCEE by eight percentage points, which is almost an 100% increase compared to being above the cutoff. In addition, whether including the individual characteristics in the regression or not barely changes the estimates of our main results, which further suggests that the discontinuity in retake probability at the cutoff is unlikely to be confounded.

[15] There is one coefficient significant at 10% level (first-time taker) when using the linear control specification, and one coefficient significant at 5% level (age) when using the quadratic control specification. However, none of the individual characteristics show significant coefficients under both specifications.

[16] Stata package *rdplot* is used for the regression discontinuity plots. See Calonico et al. (2015) and Calonico et al. (2017) for details.

Figure 1.3: The Probability of NCEE Retaking vs. Distance to Tier-2 University Cutoff Score



Notes: This figure plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Table 1.2: Effects of Below Tier-2 University Cutoff on Retake Probability

	(1)	(2)	(3)	(4)
Variables	Retake			
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0081)	(0.0076)	(0.0122)	(0.0115)
Observations	41,477	41,477	41,477	41,477
R-squared	0.117	0.220	0.117	0.220
Bandwidth	15	15	15	15
Interaction Controls	Linear	Linear	Quadratic	Quadratic
Individual Characteristics	No	Yes	No	Yes
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (2) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (3) and (4) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Columns (1) and (3) do not control for individual characteristics. Columns (2) and (4) control for a set of individual characteristics, including gender, ethnicity, hukou status, whether the individual is a first-time taker, and age. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Our results are robust to alternative specification choices and inference methods. Figure A.2 plots the estimated discontinuity in retake probability at the tier-2 cutoff for alternative bandwidth choices and weighting methods. In addition to the 15-point bandwidth in the baseline, we also consider 10-point, 20-point, and the data-driven optimal bandwidth (Calonico et al. 2014), as well as using the triangular kernel weights instead of the uniform kernel weights in the baseline.[17] Our results remain robust. In addition, Table A.2 shows that our results are not sensitive to using alternative inference methods, including clustering the standard errors by the discrete running variable and allowing error correlation between all NCEE takers in the same high school.

One interesting pattern of Figure 1.3 is that the retake probability declines as the test score approaches the cutoff from below. One may rather expect an increase in retake probability because the regret of missing the cutoff may be larger when getting

[17]The CCT optimal bandwidth (Calonico et al. 2014) is 8.1 points when using the uniform kernel weights.

closer to the cutoff from below. However, higher test score also implies better college admission options in the current year—as the tier-3 or worse colleges may still differ in quality and other characteristics, such as college and major reputation, and location. Therefore, students who have higher test scores would have better outside options other than retaking the exam, and are supposed to have a lower likelihood of retaking the exam, which appears to be the dominant effect empirically. Note that the negative slope pattern is not unique in our study—Landaud & Maurin (2020) also find that the retake probability declines as the ranking increases on both sides of the cutoff in a similar regression discontinuity design investigating the entrance exam of highly selected elite science graduate programs in France. Therefore, the slopes may be negative below the cutoff, especially in settings of educational selection system based on exams where the running variable is positively associated with the outside options. In addition, all these effects only matter for the slopes around the cutoff but not the discontinuity at the cutoff, and should not confound our RD design as we are comparing students just below and just above the cutoff only.

1.3.3 Effects of Falling Below the Tier-2 Cutoff and Retake on Exam Outcomes

To estimate the causal effects of retaking the NCEE on exam outcomes for the cutoff-induced retakers, we first estimate the reduced-form effects of falling below the tier-2 cutoff on exam outcomes:

$$
\begin{aligned}
Y^I_{i,y,tr} &= \beta^I \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) + \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr}) \\
&+ \gamma_2 \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr},
\end{aligned}
\tag{1.3.2}
$$

$$Y^F_{i,y,tr} = \beta^F \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) + \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr})$$
$$+ \gamma_2 \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \tag{1.3.3}$$

where $Y^I_{i,y,tr}$ is the outcome Y in the first year of this two-year period, which is referred to as the "initial outcome". $Y^F_{i,y,tr}$ is the final outcome Y over this two-year period, which is equal to the outcome in the first year for those who do not retake the NCEE in the next year, and is equal to the outcomes in the next year for those who retake the NCEE in the next year. It is the final payoff of the retake decision and is referred to as the "final outcome".[18] The summary statistics of the exam outcomes are shown in Table A.3. The standardized score and ranking are generally higher in the final outcome than in the initial outcome.

We distinguish the initial and final outcomes for ease of interpretation. As the initial outcomes such as test scores are realized before the cutoff is determined, they should not be affected by the cutoff ($\beta^I = 0$). By contrast, β^F identifies the effect of falling below the tier-2 cutoff on the final payoff of the retake decision. Note that we can also use $Y^F_{i,y,tr} - Y^I_{i,y,tr}$ as the dependent variable of the same specification, and the coefficient would be equal to $\beta^F - \beta^I$, which can be interpreted as the reduced-form effects of falling below the tier-2 cutoff on the improvement in exam outcomes through retakes. By definition, $Y^I_{i,y,tr}$ and $Y^F_{i,y,tr}$ only differ for those who choose to retake the NCEE in the next year, and the differences in the effects can only come from retakes. Note that because $\beta^I = 0$,

[18] We restrict the analysis to the retaking decisions and outcomes for next year and do not analyze the decisions to retake for multiple years. Unlike other admission-related exams that can be taken multiple times in a year such as SAT, the NCEE can only be taken once per year, and the decision to retake is better modeled as a sequential decision in each year. In addition, taking the NCEE more than two times is rare—there are only around 4% of the individuals who appear in our sample more than two times.

this coefficient reduces to β^F, and the coefficients when using $Y^F_{i,y,tr} - Y^I_{i,y,tr}$ or $Y^F_{i,y,tr}$ as the dependent variable identify the same parameter of interest, which is confirmed in Table 1.3. We use the specification with $Y^F_{i,y,tr} - Y^I_{i,y,tr}$ as the dependent variable as the baseline specification for measuring the return to retake because it has a clear interpretation as the causal effect on the improvement in exam performance, and can be directly compared with the improvement in exam performance for retakers that are not driven by falling below the cutoff (see Section 1.4 for more discussions).

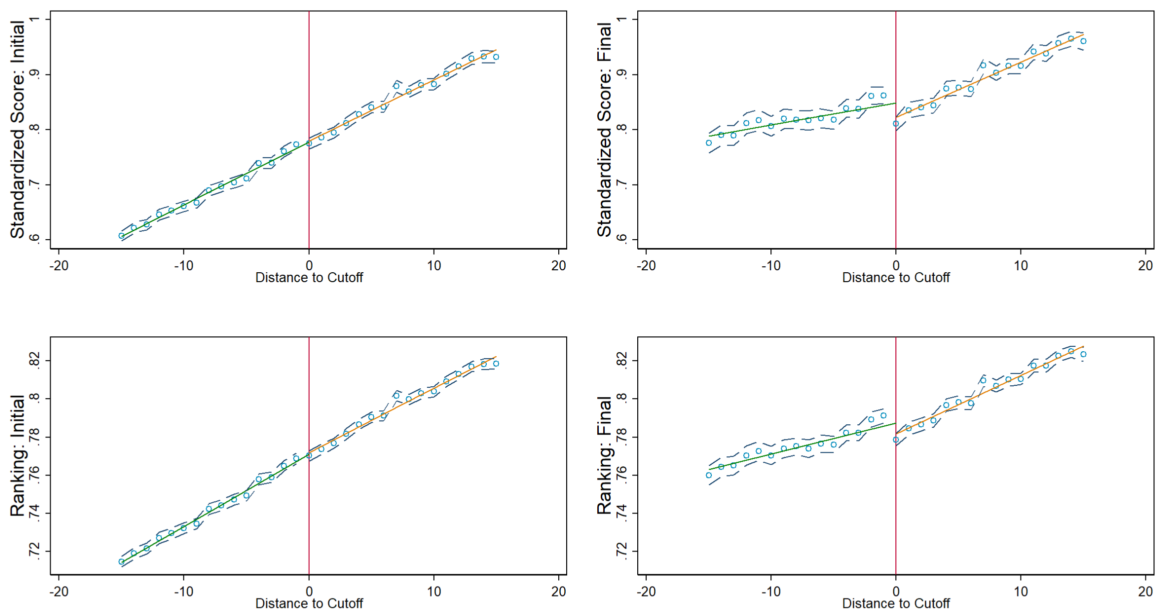
In addition, we can use the discontinuity as an instrument and estimate the following two-stage least square (2SLS) specification:

$$\begin{aligned} Y^F_{i,y,tr} - Y^I_{i,y,tr} &= \beta_{IV} Retake_{i,y,tr} + \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr}) \\ &+ \gamma_2 \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \end{aligned} \tag{1.3.4}$$

where $Retake_{i,y,tr}$ is instrumented by the indicator $\mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr})$ as in Equation 1.3.1. The coefficient β_{IV} estimates the returns to NCEE retake driven by missing the tier-2 university cutoff in terms of exam outcomes.

Figure 1.4 plots the exam outcomes against the distance to the tier-2 cutoff. The left panel of the figure plots the initial outcomes of standardized score and ranking, and the right panel of the figure plots the final outcomes. There is no discernible discontinuity in the initial score and ranking, and the points above and below the cutoff are almost on the same line. This is reassuring because the cutoff is determined after the initial score and ranking outcomes are realized, and should not have any effects on these outcomes. By contrast, there are pronounced discontinuities in the final score and ranking outcomes: students just below the cutoff have higher final payoffs in terms of standardized score and ranking than students just above the cutoff, who have better initial outcomes. The only plausible explanation for these differences is through higher retake probabilities for students scoring just below the cutoff, and retaking improves the exam outcomes substantially.

Figure 1.4: Exam Outcomes vs. Distance to Tier-2 University Cutoff Score



Notes: This figure plots the exam outcomes against the distance to the tier-2 cutoff score. The upper panel uses the standardized score as the outcome variable. The lower panel uses the ranking within the year-track as the outcome variable. The left panel shows the initial outcomes, i.e., the dependent variables in the current year. The right panel shows the final outcomes, i.e. the final payoffs of the dependent variables, which are equal to the dependent variables in the current year if the individual does not retake the NCEE in the next year, and are equal to the dependent variables in the next year if the individual retakes the NCEE in the next year. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Table 1.3 presents the results using the parametric specifications (Equations 1.3.2-

1.3.4). Panel A shows the effects on initial outcomes, and Panel B shows the effects on final outcomes. There is little evidence on effects on initial exam outcomes.[19] By contrast, falling below the tier-2 cutoff increases the final NCEE score by 0.04 standard deviations and increases the final ranking by 0.9 percentage points. Panel C shows the effects on the differences between the final and initial outcomes, which can be interpreted as the reduced-form estimates, i.e., the effects of falling below the cutoff on the improvement in exam performance, and the estimates are almost identical to Panel B. Panel D shows the 2SLS estimates of the effects of retaking the NCEE on the improvement of exam outcomes, in which we use the indicator of falling below the cutoff as an instrumental variable for retaking. The first-stage KP F-statistics are well above the Stock-Yogo critical value of 16.38 (Kleibergen & Paap 2006), suggesting a strong first stage. The 2SLS results show that retaking the NCEE increases the standardized score by 0.47 standard deviations and increases the ranking by 11 percentage points. Together, they show that retaking the NCEE leads to a substantial improvement in the exam outcomes of students, and the returns to retake are high—students can beat an additional 11% of competitors if they retake the NCEE in the next year. Figure A.3 plots the estimated returns to retake in terms of exam outcomes under different bandwidth and specification choices, and the results are similar.

[19] There is one statistically significant coefficient in column (3) for the initial ranking, when using the linear function specification. This is because the transformation from the raw test score to ranking is not a perfect linear transformation, and the estimated discontinuity happens to be statistically significant at the cutoff. Nevertheless, the point estimate for the discontinuity is small and economically insignificant, and becomes no longer statistically significant when using the quadratic function specification that accounts for the transformation from score to ranking more flexibly.

Table 1.3: Effects of Below Tier-2 University Cutoff and Retake on Exam Outcomes

	(1)	(2)	(3)	(4)
Variables	Standardized Score		Ranking	
Panel A: Initial Outcomes				
Below Cutoff	-0.0001	-0.0001	-0.0004***	-0.0000
	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Panel B: Final Outcomes				
Below Cutoff	0.0392***	0.0374***	0.0090***	0.0097***
	(0.0046)	(0.0073)	(0.0011)	(0.0018)
Panel C: Differences in Outcomes				
Below Cutoff	0.0393***	0.0375***	0.0094***	0.0097***
	(0.0046)	(0.0073)	(0.0011)	(0.0018)
Panel D: Differences in Outcomes, 2SLS				
Retake	0.4730***	0.4978***	0.1127***	0.1293***
	(0.0389)	(0.0679)	(0.0094)	(0.0169)
1st-stage KP F-stat	120.1	42.8	120.1	42.8
Observations	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15
Interaction Controls	Linear	Quadratic	Linear	Quadratic
Individual Characteristics	Yes	Yes	Yes	Yes
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable in columns (1) and (2) is the standardized score. The dependent variable in columns (3) and (4) is the ranking within the year-track. Panel A shows the results using the initial outcomes, i.e. the dependent variables in the current year. Panel B shows the results using the final outcomes, i.e. the final payoffs of the dependent variables, which are equal to the dependent variables in the current year if the individual does not retake the NCEE in the next year, and are equal to the dependent variables in the next year if the individual retakes the NCEE in the next year. Panel C shows the results using the differences between the final outcomes and the initial outcomes as the dependent variables. Panel D uses the same dependent variables as Panel C, but uses a 2SLS specification and uses the indicator of below the cutoff as an instrument for the indicator of retaking the NCEE in the next year. Columns (1) and (3) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (2) and (4) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Gender, ethnicity, hukou status, whether the individual is a first-time taker, age, and year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

One may be concerned that these improvements in exam performance may not translate to meaningful improvements in terms of admission. In order to further illustrate the magnitude of these improvements, we use an indicator of whether the test score is above or equal to the tier-1 cutoff score as the outcome variable.[20] By construction, the

[20]Note that we cannot use an indicator of whether the test score is above or equal to the tier-2 cutoff

initial outcome for this indicator is always equal to 0 within the 15-point bandwidth, as the tier-1 cutoff is generally higher than the tier-2 cutoff by 30-60 points in our sample. However, as shown in Figure A.4, the probability that students are eligible to apply for tier-1 universities is around 5% above the cutoff and around 10% below the cutoff, with a sharp discontinuity at the cutoff when we use final exam scores, i.e., exam scores in the next year for retakers and in the initial year for non-retakers. These results show that despite both ineligible to apply for tier-1 universities in the initial year, students scoring below the tier-2 cutoff are more likely to become eligible to apply for tier-1 universities next year than students scoring above the tier-2 cutoff because of the improvement of exam scores through retake. The 2SLS estimates (Equation 1.3.4) show that retaking the NCEE increases the probability of being eligible to apply for tier-1 universities by 51-62 percentage points for the cutoff-induced retakers, indicating that these improvements in exam performance are consequential for admission—the retakers become eligible to apply for universities of higher quality that they would not be eligible to apply for otherwise.[21] Moreover, these improvements are also consequential for their labor market opportunities—Jia & Li (2021) show that being eligible to apply for tier-1 universities translates to a 5.2-9.2% higher wage offer for the first job after college. Therefore, under a simple back-of-the-envelope analysis, our estimates suggest that retaking the NCEE increases the first-job wage by around 2.7-5.7% for students around the tier-2 cutoff.

To conclude, retaking the NCEE leads to sizable improvements in exam outcomes and a large return in terms of educational success for students. Our estimates (0.47 standard deviations increase) are comparable to and even larger than the estimates of the causal effects of retaking the SAT on the admission-relevant superscore (around

score as the outcome variable, because there is a discontinuity from 0 to 1 in the initial outcome at the cutoff by construction, which violates the continuity assumption required by regression discontinuity design (Cattaneo, Idrobo & Titiunik 2018). Therefore, we use whether the test score is above or equal to the tier-1 cutoff score to evaluate the consequence of the improvement in exam performance.

[21]The coefficients (not reported, available upon request) are all statistically significant at the 1% level.

0.34 standard deviations increase) in Goodman et al. (2020). However, we are unable to estimate the optimal retaking strategy for students because the opportunity cost of retaking the NCEE—postponing the entrance into higher education by (at least) a year—may also be large and heterogeneous for different students. In addition, our estimates are for the local average treatment effects of the students who retake the NCEE because of falling just below the tier-2 cutoff—a group of students performing better than the general population but still have a large room for improvement—and should be carefully interpreted when extrapolating the effects to the general population of all students.

1.4 Gender Differences in the Retaking Decisions

In the previous section, we have documented that students who confront the failure of scoring just below the tier-2 cutoff are more likely to retake the NCEE in the next year, and such cutoff-induced retakes generate large returns in terms of exam performance. In this section, we investigate the gender differences in the retaking decisions when confronting the failure of missing the cutoff, and the mechanisms and explanations for these gender differences.

1.4.1 Empirical Strategy

To investigate the gender differences in the propensity to retake induced by missing the cutoff, as well as the effects of retakes, we first split the sample by gender and estimate the baseline specifications separately. To formally test the statistical significance of the gender differences, we use the full RD sample and estimate the following specification with full gender interactions:

$$Retake_{i,y,tr} = \alpha Male_i + \beta \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) + \delta Male_i \times \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr})$$
$$+ \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr}) + \rho_1 Male_i \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \gamma_2 \mathbf{1}(Score_{i,y,tr} <$$
$$Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \rho_2 Male_i \times \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr}$$
$$- Cutoff_{y,tr}) + \theta_1 X_i + \theta_2 Male_i \times X_i + \mu_{y,tr,male} + \varepsilon_{i,y,tr}, \tag{1.4.1}$$

where $Male_i$ is a binary indicator for being male. With full gender interactions, the slopes are allowed to be different to the left and right of the cutoff, and be different for each gender. The individual characteristics are also interacted with the male indicator to allow for differential effects, and the fixed effects are now at year-by-track-by-gender level. The coefficient δ captures the gender differences in the propensity to retake induced by missing the cutoff, and is equal to the difference in the coefficients for male and female subsamples.

For gender differences in returns to retake, we follow the same strategy as the baseline and estimate the following specification:

$$Y^F_{i,y,tr} - Y^I_{i,y,tr} = \alpha Male_i + \beta_{IV} Retake_{i,y,tr} + \delta_{IV} Male_i \times Retake_{i,y,tr} + \gamma_1 f(Score_{i,y,tr} -$$
$$Cutoff_{y,tr}) + \rho_1 Male_i \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \gamma_2 \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times$$
$$f(Score_{i,y,tr} - Cutoff_{y,tr}) + \rho_2 Male_i \times \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} -$$
$$Cutoff_{y,tr}) + \theta_1 X_i + \theta_2 Male_i \times X_i + \mu_{y,tr,male} + \varepsilon_{i,y,tr}, \tag{1.4.2}$$

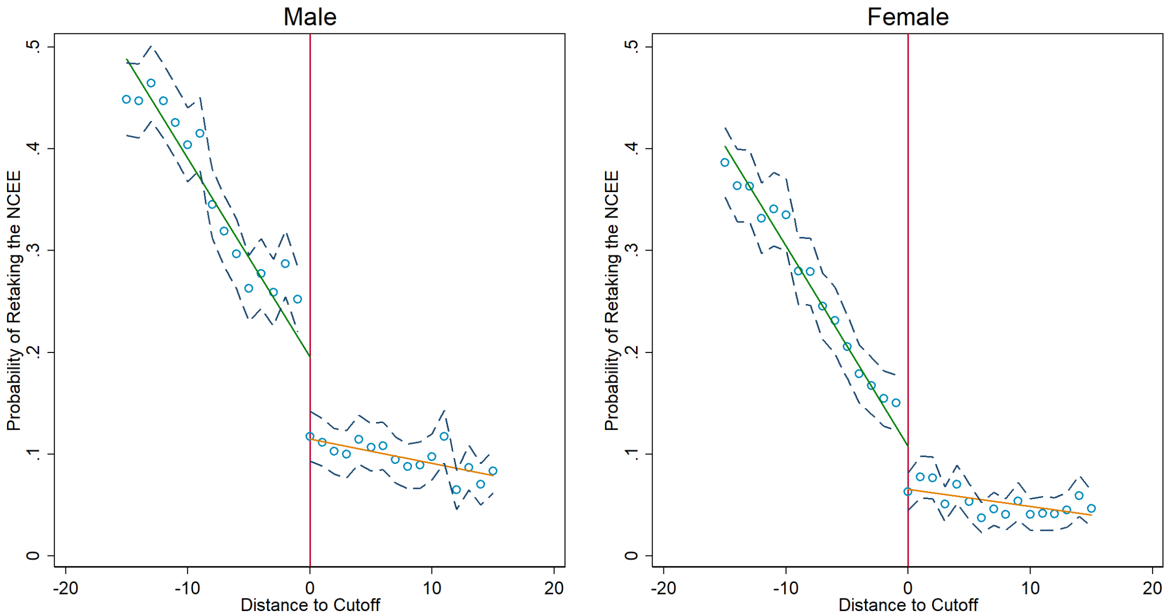
where $Retake_{i,y,tr}$ and $Male_i \times Retake_{i,y,tr}$ are instrumented by $\mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr})$ and $Male_i \times \mathbf{1}(Score_{i,y,tr} < Cutoff_{y,tr})$ as in Equation 1.4.1. The coefficient δ_{IV} estimates the gender differences in the effects of the NCEE retake driven by missing the tier-2 university cutoff on improvements in the exam outcomes, and is equal to the difference in estimated returns for male and female subsamples.

1.4.2 Main Results

We start by showing the predictors of retaking the NCEE in the next year from a linear probability model that regresses the retake indicator on a set of covariates. The results are presented in Table A.4. Columns (1)-(2) show the results for our full sample. The results show that males are two to three percentage points more likely to retake than females, and the gender difference is persistent when more covariates are added. Note that these other covariates are also strong predictors of the retake probability—students of Han ethnicity are much more likely to retake, followed by students of Hui ethnicity, compared with students of other minority ethnicities. In addition, students with urban hukou, repeated takers, older students and students with higher test scores are less likely to retake. Columns (3)-(4) show the results for the sample within the 15-point bandwidth around the tier-2 cutoff. The pattern is similar, and the gender difference is more pronounced—males are six to eight percentage points more likely to retake than females when they score around the tier-2 cutoff.

Given that retaking the NCEE is an endogenous choice that correlates with many unobservable personal traits, the gender differences in retakes could arise from gender differences in many aspects, such as confidence and goal-setting. We focus on the retakes induced by missing the tier-2 cutoff and examine whether males and females differ in the likelihood of retaking when confronting this exogenous failure. Figure 1.5 plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff separately for males and females. It is clear that males have a higher retake probability than females on both sides of the cutoff, and the gender differences are much more pronounced to the left of the cutoff. More importantly, the discontinuity in retake probability at the cutoff is much more pronounced for males than for females.

Figure 1.5: The Probability of NCEE Retaking vs. Distance to Tier-2 University Cutoff Score, by Gender



Notes: This figure plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score separately for males and females. The left panel is for males, and the right panel is for females. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Note that our analysis on gender differences relies on the validity of the regression discontinuity design for each gender. We plot the density distribution of the running variable around the tier-2 cutoff for each gender in Figure A.5, and there is no evidence of discontinuous density in test scores around the tier-2 cutoff for males or females.[22] We

[22]The p-value of the manipulation testing procedure proposed by Cattaneo, Jansson & Ma (2018) is 0.61 for males and 0.84 for females.

also plot the individual characteristics of students against the distance to the cutoff for males and females separately in Figure A.6, and the estimation results for the balancing tests are shown in Table A.5. Again, there is no substantial discontinuous jump for these pre-determined characteristics at the cutoff for both males and females under both linear and quadratic controls, which reassures the validity of our research design.[23]

Table 1.4 presents the results of the parametric specifications in Equation 1.4.1. Columns (1)-(2) present the results for males and females using the linear control separately. Males are 11 percentage points more likely to retake when falling just below the tier-2 cutoff, while females are only 5.5 percentage points more likely to retake when falling just below the cutoff. The gender difference in the retaking probability induced by the cutoff is around 5.6 percentage points, and is statistically significant at 1% level when using the full gender interaction model as in column (3). Columns (4)-(6) similarly show the results using the quadratic control. The results are very similar, and the estimated gender difference is even larger (7.3 percentage points). The gender differences in the effects are quite substantial: the discontinuity effect for males is more than twice of such effect for females.

[23]There is one coefficient significant at 10% level (Han) when using the linear control specification, and one coefficient significant at 5% level (urban) when using the quadratic control specification for males. There is one coefficient significant at 10% level (Han) and one coefficient significant at 5% level (Hui) when using the linear control specification, and no significant coefficient when using the quadratic control specification for females. Again, none of the individual characteristics show significant coefficients under both specifications.

Table 1.4: Gender Differences in the Effects of Below Tier-2 University Cutoff on Retake Probability

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Male	Female	Full	Male	Female	Full
Variables	Retake					
Male*Below Cutoff			0.0558***			0.0732***
			(0.0133)			(0.0203)
Below Cutoff	0.1107***	0.0549***	0.0549***	0.1109***	0.0377***	0.0377***
	(0.0108)	(0.0093)	(0.0093)	(0.0167)	(0.0137)	(0.0137)
Observations	21,162	20,315	41,477	21,162	20,315	41,477
Bandwidth	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track FE	Yes	Yes	By Gender	Yes	Yes	By Gender

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (4) are using only the male sample. Columns (2) and (5) are using only the female sample. Columns (1)-(2) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (4)-(5) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects are controlled in columns (1)-(2) and (4)-(5). Columns (3) and (6) are using the full sample with full gender interactions. Columns (3) controls for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Columns (6) controls for a quadratic function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Columns (3) and (6) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

In Table A.6, we present results using the full RD sample but relaxing the model assumption of full gender interactions to assess the robustness of the results. In column (1), we do not include any individual characteristics and use year-by-track fixed effects, but still allow the slopes to be different to the left and right of the cutoff, and be different for each gender. In column (2), we include individual characteristics, but still do not include gender interactions with these covariates. Then in column (3), the effects of individual characteristics and fixed effects are allowed to vary by gender, and the specification is the same as Equation 1.4.1 and in column (3) of Table 1.4. The advantage of columns (1)-(2) of Table A.6 is that the coefficient of the male dummy will not be absorbed as in the full gender interaction specification, and we can clearly observe the gender difference in retake probability when scoring just above the cutoff. The results in columns (1)-(3) show that, under the linear control specification, the gender difference in the retaking probability induced by the cutoff is robust to the inclusion of covariates and their gender interactions. In addition, columns (1)-(2) show that men are five to six percentage points more likely to retake than women when scoring just above the tier-2 cutoff, and such gender difference becomes around twice larger when scoring just below the cutoff. Columns (4)-(6) show the results using the quadratic control specification, and the results are similar. Figure A.7 plots the estimated coefficients of gender differences (the interaction terms in columns (3) and (6) of Table 1.4) under different bandwidth and specification choices and shows the robustness of the results. Indeed, the estimated gender differences are large and statistically significant across various specifications.

Therefore, we can conclude that males are more likely to be motivated by missing the tier-2 cutoff and retake the NCEE next year than females. Our findings are consistent with previous studies that females are more likely to stop participating in competitions after failures (Buser & Yuan 2019, Wasserman 2020), and our results are for a context with much higher stakes and for a much larger population. By contrast, our results differ from Goodman et al. (2020) who find that females are more likely to retake the SAT than males, although the sense of competition against others is less clear and retake is less costly in SAT than that in the NCEE.

1.4.3 Understanding Gender Differences in Reactions to Failure

Why are females less inspired to retake after the failure of missing the cutoff? As retaking the NCEE is a risky choice that has high opportunity costs and uncertain returns, such gender differences may be explained by gender differences in several aspects of the decision-making process. First, the returns to retake may be different across gender. For example, if males in general have better performance and higher returns when retaking the NCEE, then it is rational for them to participate in the retakes more frequently. Second, the opportunity costs may be different across gender, as postponing the time of entering higher education and the labor market by a year could have differential impacts on men and women, especially with fertility concerns. Third, the gender differences in retake decisions may also be explained by gender differences in non-cognitive traits and preferences. For example, females may have different causal attribution than males. Men tend to attribute success to internal factors such as talent, and failure to external factors such as luck, whereas women tend to do the opposite (Dweck et al. 1978, Ryckman & Peckham 1987, Beyer 1998). Females who fail the cutoff may be more likely to attribute the failure to own ability and be less confident about the prospect of the retakes, and thus are less motivated to retake than males. Fourth, the gender differences could come from differences in risk preferences (Boring & Brown 2016, Saygin 2016, Reuben et al. 2017), that females may be less motivated to retake than males because of stronger risk aversion. Finally, although the decision to retake is made by students, parents may also have a significant influence on the decision-making process. If the financial or emotional support from parents is weaker for females because of gender differences in social norms, then females may be less likely to retake as well.

Although distinguishing between these competing explanations is hard, we are able to directly test the hypothesis of differential returns by examining whether the returns to retake in terms of exam outcomes are higher for males. Table 1.5 presents the results for exam outcomes using the linear control specification. The results show that the return to retake is on average 0.42 standard deviations in test scores for males, but is 0.58 standard deviations in test scores for females. The difference is statistically significant at the 10% level. When measuring the return in terms of the relative ranking, females also show a larger return than males, although the difference is insignificant. Figure A.8 plots the estimated gender differences in returns to retake in terms of exam outcomes under different bandwidth and specification choices. The results are very robust—the estimates are either negative or statistically insignificant. These results show that females in general have similar or even higher returns to retake than males in terms of exam outcomes. Therefore, higher returns for males in terms of exam outcomes are unlikely to explain the results.

Table 1.5: Gender Differences in the Effects of Retake on Exam Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Male	Female	Full	Male	Female	Full
Variables	Standardized Score			Ranking		
Male*Retake			-0.1566*			-0.0246
			(0.0904)			(0.0223)
Retake	0.4220***	0.5785***	0.5785***	0.1049***	0.1295***	0.1295***
	(0.0460)	(0.0765)	(0.0765)	(0.0119)	(0.0179)	(0.0179)
1st-stage KP F-stat	105.0	35.2	17.6	105.0	35.2	17.6
Observations	21,162	20,315	41,477	21,162	20,315	41,477
Bandwidth	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Linear	Linear	Linear
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track FE	Yes	Yes	By Gender	Yes	Yes	By Gender

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variables are the differences between the final outcomes and the initial outcomes of the exam outcomes (standardized score in columns (1)-(3), the ranking within the year-track in columns (4)-(6)). The indicator of below the cutoff (and its interaction with male dummy) are used as instruments for the indicator of retaking the NCEE in the next year (and its interaction with male dummy). Columns (1) and (4) are using only the male sample. Columns (2) and (5) are using only the female sample. Columns (1)-(2) and (4)-(5) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects are controlled in columns (1)-(2) and (4)-(5). Columns (3) and (6) are using the full sample with full gender interactions. The linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in columns (3) and (6). Columns (3) and (6) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

One may be concerned that the gender differences in returns to retake are in fact driven by gender differences in selection into retake. For example, suppose males have higher returns than females in general, and both males and females select into retake if their returns are sufficiently high, then there would be more males choosing to retake than females, and their average returns become no longer higher than females as they are more selected into retake. Indeed, the implicit assumption of comparing the returns to retake for males and females in our sample is that students do not systematically select into retake based on their expected returns to retake. We believe this assumption is likely to hold because it is hard for students to predict the returns to retake and select into retake, as the performance in NCEE—one of the toughest exams in the world—is hard to predict.[24] In addition, we provide some additional direct evidence against this rational selection hypothesis. Specifically, we compare the returns to retake for the cutoff-induced retakers ("compliers") and for the retakers that will choose to retake regardless of being above or below the cutoff ("always-retakers") at the cutoff.

Let $D_i(1)$ and $D_i(0)$ denote the potential treatment value (whether chooses to retake or not) when the individual i is assigned to be below and above the cutoff, respectively. Under the monotonicity condition (Imbens & Angrist 1994, Cattaneo, Idrobo & Titiunik 2018) $D_i(1) \geq D_i(0)$, the individuals can be classified into three types: "always-retakers" who always choose to retake regardless of being above or below the cutoff ($D_i(1) = D_i(0) = 1$), "compliers" who only choose to retake if being below the cutoff ($D_i(1) = 1$ and $D_i(0) = 0$), and "never-retakers" who never choose to retake ($D_i(1) = D_i(0) = 0$). Then the coefficient β in Equation 1.3.1 estimates the proportion of "compliers", and the coefficient β_{IV} in Equation 1.3.4 estimates the local average treatment effect (LATE) of retake on the improvement of the outcome for the "compliers" at the cutoff under continuity assumptions (Cattaneo, Idrobo & Titiunik 2018).

[24]In addition, the assumption can also be guaranteed to hold when assuming that the expected returns to retake are homogeneous for students scoring just below the cutoff within each gender. Although the expected returns to retake may be heterogeneous in general, it may be plausible to assume that to be homogeneous for students who score just below the cutoff—a group of students showing similar ability and receiving the same feedback signal.

One important feature of our setting is that we can also estimate the LATE of retake on the improvement of the outcome for the "always-retakers" as well. Specifically, we can estimate the mean of the improvement of the outcome for individuals who choose to retake and scoring exactly at the cutoff to recover the LATE for the "always-retakers" at the cutoff.[25] The intuition is that the individuals who choose to retake when scoring above the cutoff must be "always-retakers" by construction.[26] If the rational selection hypothesis is true, then the LATE for the "always-retakers" at the cutoff should be higher than the LATE for the "compliers" at the cutoff, because the former group has strong motivation to retake regardless of the cutoff, and should be the group with the strongest incentive to retake. In Appendix A.1, this intuition is formalized using a simple rational selection model.

We thus estimate and compare the LATE for the "compliers" and the "always-retakers" at the cutoff. The LATE on the improvement of the test score is 0.47 standard deviations for the "compliers" (as shown in Table 1.3, Column (1)), and is 0.40 standard deviations for the "always-retakers" (N=122, s.e.=0.03). Therefore, there is no evidence that the LATE for the "always-retakers" is higher than for the "compliers". When conducting the analysis by gender, men have a LATE of 0.42 standard deviations for the "compliers" (as shown in Table 1.5, Column (1)) and a LATE of 0.40 for the "always-retakers" (N=79, s.e.=0.04), while women have a LATE of 0.58 standard deviations for the "compliers" (as shown in Table 1.5, Column (2)) and a LATE of 0.39 for the "always-retakers" (N=43, s.e.=0.05). Again, the results are inconsistent with the rational selection model as "compliers" show higher LATE than "always-retakers", suggesting that it is unlikely that students can rationally predict their returns and select

[25] See Appendix A.1 for more discussions on these results.

[26] Note that in fact we can similarly use the same method to recover the treatment effect for retakers at any score, but the group of retakers is in general endogenous. Therefore, we focus on the RD estimates of the LATE for the "compliers" at the cutoff when analyzing returns to retake.

into retake. The results are similar when using the relative ranking to measure exam outcomes. Together, these results indicate that our results are not explained by rational selection.

However, the returns we examined here only refer to improvements in exam scores, not other pecuniary or non-pecuniary returns that may be associated with higher scores in the long run. For example, even if the returns to retake are similar for males and females in terms of exam performance, admission into a selective university may translate into higher pecuniary or non-pecuniary returns for males because of labor-market conditions (Cai et al. 2019). Our research design does not allow us to rule out this potential explanation.

1.4.4 Heterogeneous Effects and Mechanisms

In this section, we explore the heterogeneity in gender differences in cutoff effects to further clarify the mechanisms. We focus on the linear control specification throughout the heterogeneous effect analysis, and the results are in general similar when using the quadratic control specification. Table 1.6 presents the estimation results of the gender differences in the cutoff effects on retake probability by individual characteristics. The gender differences are pronounced and similar for Han and minority ethnicity students, and for urban and rural students. These results suggest that the gender differences in the retake decision when confronting the failure may not heavily depend on family background and financial resources, as well as differential cultural and social norms across ethnicities. Besides, these results suggest that the gender differences in labor market returns may only play a limited role in explaining our results, as the return to education for females is substantially higher than the return for males in rural areas but is similar to or slightly lower than the return for males in urban areas (Wang & Wu 2018), but the gender differences in reactions to failure are very similar among urban and rural students. In addition, the gender differences are similar for students in the science track, where females are less represented, and students in the art track.[27] Finally, the gender differences are presented for both first-time takers and repeated takers. These results show that the gender differences in reactions to failure are not driven by certain groups of individuals, but are pronounced for all types of individuals.

We also present the estimated gender differences in the cutoff effects on retake probability by age in Figure 1.6.[28] The results show that the gender differences are pronounced for younger cohorts, but much smaller for older cohorts, especially for those above 21 years old. This implies that the gender differences in opportunity cost may not play an important role in explaining our results. As women in older cohorts may face larger marital market social norm pressure, we may expect that the gender difference in the opportunity cost of spending another year is likely larger for older cohorts, which would predict more pronounced gender differences in retake probability for older cohorts. However, our results show a decline rather than an increase in the gender differences as the cohort becomes older. Note that the age of an individual is highly correlated with the probability that the individual is a repeated taker, and we are unable to distinguish the age differences from the repeated taker differences.

[27] In our sample, the proportion of males is around 60% in the science track, and is around 35% in the art track.

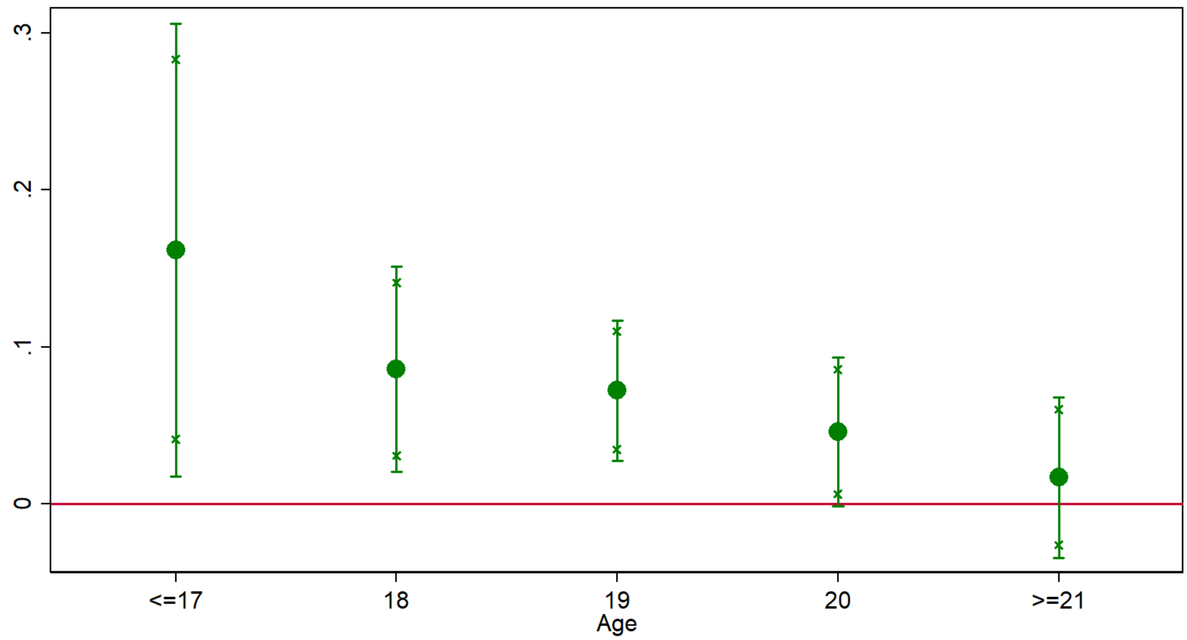
[28] A typical student enters primary school at age 6-7, and thus attends their first NCEE at age 18-19. Most of the observations in the regression discontinuity sample are of age 18 or 19. Observations of age 17 or below are likely to be individuals who enter primary school early or skip grades. Observations of age 20 or above are likely to be individuals who enter primary school late, repeat grades, or are retaking the NCEE.

Table 1.6: Gender Differences by Individual Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Han	Minority	Urban	Rural	Art	Science	First-Time Taker	Repeated Taker
Variables					Retake			
Male*Below Cutoff	0.0506***	0.0739***	0.0504**	0.0627***	0.0498*	0.0587***	0.0822***	0.0276**
	(0.0150)	(0.0247)	(0.0205)	(0.0188)	(0.0260)	(0.0158)	(0.0216)	(0.0136)
Below Cutoff	0.0683***	0.0017	0.0646***	0.0464***	0.0567***	0.0531***	0.0902***	0.0077
	(0.0109)	(0.0151)	(0.0121)	(0.0135)	(0.0140)	(0.0117)	(0.0137)	(0.0093)
Observations	32,896	8,581	18,885	22,592	12,556	28,921	23,643	17,834
Bandwidth	15	15	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track-Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Column (1) uses the sample of observations with Han ethnicity. Column (2) uses the sample of observations with minority ethnicity (Hui and other minority ethnicities). Column (3) uses the sample of observations with urban hukou status. Column (4) uses the sample of observations with rural hukou status. Column (5) uses the sample of observations in art track. Column (6) uses the sample of observations in science track. Column (7) uses the sample of observations that are first-time takers. Column (8) uses the sample of observations that are repeated takers. All columns control for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Figure 1.6: Gender Differences in the Effects of Below Cutoff on Retake Probability, By Age



Notes: This figure plots the estimated gender differences in the effects of below the tier-2 cutoff on retake probability for different age groups. The sample consists of observations within the 15-point bandwidth around the cutoff. The linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in all regressions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects, are also controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Table 1.7 presents the estimation results of the gender differences in the cutoff effects on retake probability by high school and county characteristics. In columns (1)-(2), we divide the sample based on the quality of the high school.[29] High-quality high schools

[29]The quality of the high school is measured by the median of the standardized NCEE score of the students in the high school, separately measured for each year-track. A student in our RD sample is defined to be in high-quality school if the quality of her high school is above or equal to the median of the quality of high school in the RD sample in the given year-track. There are a small proportion of NCEE takers (less than 1%) that do not have valid information on high school, and they are excluded

have better educational resources and peer groups, and students in high-quality high schools may have higher returns and lower non-pecuniary costs if they choose to retake. However, the results show that the gender differences are large and of similar magnitude in high-quality and low-quality high schools. In columns (3)-(4) and (5)-(6), we divide the sample based on the sex ratio of the high school cohort (columns (3)-(4)) and the sex ratio of the county (columns (5)-(6)).[30] Students in places with different levels of sex imbalance may face different gender-related social norms and family support. For example, in places with high boys-to-girls ratios, girls may receive little family support and thus are less likely to retake. However, we find large and similar gender differences when the students are exposed to places with different levels of sex ratio, suggesting that social norms and family support are unlikely to fully explain our results. Finally, in columns (7)-(8), we divide the sample based on the GDP per capita of the counties.[31] Students in poor counties may have lower economic returns and higher opportunity costs of retake, and such differences in benefits and costs of retake between rich and poor counties may differ across gender. However, the results show that the gender differences are again large and of similar magnitude for students in rich and poor counties, indicating that benefits and costs may play a limited role in explaining the gender differences.

All these results on heterogeneous effects show that the gender differences in reactions to failure are not driven by certain groups of individuals, but are pronounced for all types

from this analysis.

[30]The sex ratio of the high school is measured by the proportion of male students in the high school, separately measured for each year. A student in our RD sample is defined to be in school with high sex ratio if the sex ratio of her high school is above or equal to the median of the sex ratio of high school in the RD sample in the given year. The NCEE takers without valid information on high school are again excluded from this analysis. The sex ratio of the county is measured by the proportion of males in the total population. A student in our RD sample is defined to be in county with high sex ratio if the sex ratio of her county is above or equal to the median of the sex ratio of county in the RD sample in the given year.

[31]A student in our RD sample is defined to be in county with high GDP if the GDP per capita of her county is above or equal to the median of the GDP per capita of county in the RD sample in the given year.

Table 1.7: Gender Differences by High School and County Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	High Quality High School	Low Quality High School	High Sex Ratio High School	Low Sex Ratio High School	High Sex Ratio County	Low Sex Ratio County	High GDP County	Low GDP County
Variables				Retake				
Male*Below Cutoff	0.0642***	0.0461***	0.0553***	0.0556***	0.0466**	0.0638***	0.0478***	0.0604***
	(0.0205)	(0.0177)	(0.0195)	(0.0179)	(0.0181)	(0.0195)	(0.0172)	(0.0210)
Below Cutoff	0.0527***	0.0598***	0.0579***	0.0528***	0.0657***	0.0427***	0.0586***	0.0550***
	(0.0128)	(0.0134)	(0.0144)	(0.0116)	(0.0129)	(0.0131)	(0.0112)	(0.0155)
Observations	21,098	20,059	21,035	20,122	22,286	19,191	21,785	19,692
Bandwidth	15	15	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track-Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Column (1) uses the sample of observations in high-quality high schools. Column (2) uses the sample of observations in low-quality high schools. Column (3) uses the sample of observations in high schools with high male-female sex ratio. Column (4) uses the sample of observations in high schools with low male-female sex ratio. Column (5) uses the sample of observations in counties with high male-female sex ratio. Column (6) uses the sample of observations in counties with low male-female sex ratio. Column (7) uses the sample of observations in counties with high GDP per capita. Column (8) uses the sample of observations in counties with low GDP per capita. All columns control for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

of individuals. These results provide substantial support for the external validity of our findings. In addition, these results suggest that gender differences in returns, opportunity costs and family support are less likely to explain our results, while gender differences in causal attributions, confidence, and risk preferences may well explain our results, as these gender differences in non-cognitive traits may exist for all types of students. Nevertheless, we are unable to fully distinguish between these potential explanations.

1.4.5 Implications

We conclude this section by doing a simple back-of-the-envelope calculation for the economic meaning of the gender gap in retake tendency. One may be concerned that the returns to retake at the cutoff may be different from the general population. Therefore, we conduct the calculation for the sample within 15-point bandwidth around the tier-2 cutoff only, as extrapolating the estimated returns may be more plausible within this sample. Females have a 0.06-unit higher standardized score in terms of the final outcome over the two-year period than males (0.89 vs. 0.83) in this sample. If the gender gap in retake probabilities vanishes (7.82 percentage points, as shown in Table A.4, Column (4)), and assuming the returns to retake can be extrapolated to this sample (0.4730, as shown in Table 1.3, Column (1)), then females would have an additional 0.037-unit advantage in terms of the final standardized score in this counterfactual case. This case would expand the current gender gap in exam performance in this sample by 60%. Despite having better exam performance on average, females have a 3.3-percentage point lower probability of finally being eligible to apply for higher-quality tier-1 universities over the two-year period than males (6.3% vs. 9.6%) in this sample because of a lower tendency to retake. If the gender gap in retake probabilities vanishes, then females would be substantially more represented in high-quality universities. These effects may have important implications for the gender disparities in the labor market.

1.5 Conclusion

We document the gender differences in reactions to failure in high-stakes competition in an important field setting—the NCEE in China. Using unique administrative data on the universe of NCEE takers in Ningxia and exploiting a regression-discontinuity design, we show that students who score just below the tier-2 cutoff have an eight percentage point higher probability (an almost 100% increase compared to being above the cutoff) of retaking the NCEE in the next year. We then exploit the discontinuity in the probability of retaking the NCEE around the cutoff to address endogenous retaking and estimate the causal returns to retaking the NCEE. The results show that retaking the NCEE increases the test scores for admission by 0.47 standard deviation, and increases the relative ranking among competitors by 11 percentage points. Our results show that retaking the NCEE generates large returns in terms of exam performance and educational success.

We then document large gender differences in the propensity to retake in the next year. We find consistent evidence that women are less likely to retake the NCEE than men with similar exam performance. The cutoff-induced retakes from the regression discontinuity design, which reflect the desire to participate in the competition again inspired by the exogenous failure of scoring below the cutoff, are also much more pronounced for men than for women. Our results suggest that these gender differences are not explained by gender differences in returns to retake in terms of exam outcomes, and are unlikely to be fully explained by gender differences in benefits, opportunity costs and family support, but may be explained by gender differences in non-cognitive traits, such as causal attribution, confidence, and risk preferences. Our estimates suggest that if females are equally likely to retake as males, females would have better final exam performance and be substantially more represented in the high-quality universities, which may in turn, have important implications for the gender equality in the labor market.

Unfortunately, due to data limitations, we cannot examine the effects of retaking NCEE on long-term outcomes, such as labor market and marital outcomes. In addition, we are unable to fully disentangle the potential explanations for the gender differences in reactions to failure in the NCEE. Further research is needed to test these hypotheses, which could be important for policy designs to address the gender gap.

Chapter 2

Vulnerable Boys: Short-term and Long-term Gender Differences in the Impacts of Adolescent Disadvantage

2.1 Introduction

In 1990, the proportion of young women (aged 25 to 29) who had completed four-year college degrees reached near equality with young men in the United States after steadily increasing for several decades. By 2014, the long-standing gender gap in educational attainment had not just disappeared but reversed—favoring women by a substantial margin. More than 37 percent of young women now have at least a four-year college degree, compared to less than 31 percent of young men (U.S. Census Bureau 2016). Similar gender gaps in education are opening up around the world, with young women completing tertiary degrees at higher rates than men in almost all OECD countries (OECD 2015).

Rising female educational attainment has been a consequence of the removal of barriers to women's schooling and market work that had discouraged investments in women's human capital. However, the emergence of a female advantage in higher education rather than parity, even though women continue to have lower employment rates and shorter work hours than men, has been unexpected. Some studies find a gender gap in benefits to education, such as a higher college wage premium for women than for men (Dougherty 2005), but a consensus seems to be emerging that the principal source of the college gap lies in gender differences in the non-pecuniary costs of educational persistence. These cost differences are reflected in a persistent male disadvantage in school performance at all levels and are due, some argue, to lower levels of non-cognitive skills among boys and the resulting "behavioral advantage" of girls (Fahle & Reardon 2018).

An extensive literature in education and the social sciences has documented gender differences in the academic and behavioral outcomes of boys and girls in elementary and secondary school (Buchmann et al. 2008, DiPrete & Buchmann 2013, Salisbury et al. 1999). These gender gaps are not new phenomena: girls have consistently outperformed boys in grades and have been less likely to get in trouble at school (Duckworth & Seligman 2006). Recent studies interpret the observed gender differences in academic performance, grade repetition, special education placement, homework hours and school reports of disruptive behavior as indicative of gaps between the non-cognitive skills of boys and girls (Becker et al. 2010, Goldin et al. 2006, Jacob 2002). Gender gaps in social and behavioral skills appear to develop early—girls begin school with more advanced learning and social skills than boys, and this advantage grows over time. These early skill gaps, in turn, explain much of the gender differential in later academic achievement and educational attainment (DiPrete & Jennings 2012, Owens 2016).

Economists have focused on possible causes of this gender gap in behavior, including the possibility that the development of capabilities that enhance academic achievement, such as self-control, is more sensitive to family disadvantage among boys than is the skill development of girls. Autor & Wasserman (2013) suggest that the increased prevalence of single-parent families and decreased contact with a stable male parent may have a particularly negative impact on boys and contribute to the growing gender gap in education and to male labor market difficulties, either because boys are more vulnerable to the loss of parental time and financial resources, or due to role model effects of the same-sex parent.[1] Two recent studies report empirical evidence consistent with this hypothesis. Bertrand & Pan (2013) find that the gender gap in early behavior problems and school suspensions is much larger for the sons and daughters of single mothers than for children in two-parent households. They interpret this as evidence that the non-cognitive skills of boys are adversely impacted by non-traditional family arrangements, and suggest that boys' greater tendency to act out and develop conduct problems might be particularly relevant to their relative absence in college. Autor et al. (2019) examine the effects of several dimensions of family disadvantage, including mother's education and marital status, an SES index, neighborhood income and school quality on school performance, behavior, and on-time high school graduation for a large sample of children in Florida. They find that indicators of family disadvantage tend to have significantly greater effects on school outcomes for boys, compared to girls.

In this study, we move beyond K-12 achievement and behavior to assess the role of excess male vulnerability to adverse childhood environments in explaining gender gaps in college graduation and other adult outcomes. This requires data that permits us to link family structure and characteristics of schools and neighborhoods in childhood with

[1] A few studies have found that boys do worse, emotionally and academically, following a divorce (Hetherington & Kelly 2002), but meta-analysis of (correlational) studies of father absence and child wellbeing by Amato & Gilbreth (1999) finds no support for the hypothesis that boys benefit more than girls from paternal involvement. Prevoo & Ter Weel (2015) examine the impact of family disruption on children's personality development and find that behavior problems and self-esteem among teenage boys are more responsive to parental death, but not family disruption in general, compared to girls. The vulnerable male story is also difficult to square with the findings of Bailey & Dynarski (2011) that the growing gender gap in college attendance rates is driven primarily by increases in the education of daughters of high-income parents.

longer-term outcomes, including final educational attainment, and we use rich longitudinal data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). In particular, we examine the association between family disadvantage in adolescence and outcomes that include behavior in school, mental health and educational aspirations in adolescence, and educational attainment, employment, income, marriage and fertility in adulthood, for a recent cohort of young adults.

We find, as do previous studies, that boys are more sensitive than girls to father absence in terms of problems with schoolwork and interactions in school. Girls, however, are more likely than boys to respond to father absence with increased levels of depression, and are particularly negatively affected by residence with a stepfather. When we turn to educational attainment and other adult economic outcomes in later waves of Add Health, we find that family structure in adolescence does not have differential effects on the college graduation rates, income, and job stability of men and women. Family structure in adolescence does differentially affect men and women in terms of family decisions in young adulthood, including marriage and fertility. Results using data of another recent cohort of young adults from the National Longitudinal Survey of Youth, 1997 produce a similar pattern of results. In addition we find, as do Autor et al. (2019), that the behavior of adolescent boys is more responsive to other indicators of disadvantage, such as poor quality schools and less-educated neighborhood, than is the behavior of girls. These differential environmental effects, however, also fail to persist into adulthood.

A few other studies have focused on gender differences in the adult impacts of early life disadvantage. Closest to our design is Brenøe & Lundberg (2018), who are able to assess some long-term effects of family disadvantage with Danish administrative data. Linking entire population cohorts from birth into adulthood, they find that family disadvantage, particularly low maternal education, has more negative effects on school-age outcomes of boys relative to girls, as expected. Administrative data provides few adolescent measures;

the key outcome is a marker for completing primary school on time. Long-term effects are quite different: early disadvantage, particularly low parental education, tends to have stronger impacts on the educational attainment, employment, and earnings of adult women, compared to adult men.[2]

Slade et al. (2017), who also use the Add Health data, find a stronger association between nontraditional family structure in childhood and health outcomes, including depression, self-reported health and smoking, for girls compared to boys. They also find that many of the effects of father absence on health and mental health outcomes in adolescence, including depression, are no longer significant in young adulthood. Autor et al. (2019) find that mother's marital status and education have larger effects on son's high school graduation than daughter's, but are not able to follow subjects further into college and employment. Fan et al. (2015) use Norwegian administrative data to show that mother's employment early in a child's life is more negatively associated with the educational attainment of sons than daughters, suggesting that they are more adversely affected by a reduction in maternal time. Finally, Gould et al. (2011) present contrary results in a different environment–Yemenite child refugees in Israel who were placed in more modern environments achieved higher education and employment rates, but these benefits accrued largely to women.

Our results, which show that short-run differential impacts of disadvantage on boys and girls are not reflected in long-term outcomes, cast doubts on interpreting the gender differences in the impacts of disadvantage in school as evidence on excess male vulnerability in terms of non-cognitive skills. If adolescent disadvantage has differential impacts on non-cognitive skills for boys and girls, we would expect the gender differences to persist into adulthood, affecting long-run outcomes such as college graduation and labor mar-

[2]Family structure and parents' marital status at birth tend to have weak and inconsistent effects on later outcomes, but there is less variation in these indicators than in U.S. data and the comprehensive Danish social welfare system may mitigate the impacts of family disruption on children.

ket performance that are highly correlated with non-cognitive skills. Instead, school-age boys and girls appear to respond to adolescent environments and resources with distinct, gender-typical behaviors that haven't been previously noted in this context, rather than developing a skill gap with implications for adult economic outcomes, such as the gender gap in college graduation. Though non-traditional family structures and adverse environments in adolescence are associated with lower educational attainment and poor labor market outcomes for both men and women, we fail to reject the hypothesis that adolescent disadvantage does not have differential effects across gender on college graduation rates and labor market outcomes. Disadvantage in adolescence does have some distinct effects on marriage and fertility for men and women, but we find little evidence supporting a general pattern of excess male vulnerability.

2.2 Data

2.2.1 Add Health Sample

The National Longitudinal Study of Adolescent to Adult Health (Add Health) has collected a rich array of longitudinal data on the social, economic, psychological and physical well-being of young men and women in the U.S. from adolescence through young adulthood.[3] The Add Health study began in 1994-95 with a nationally-representative school-based survey of more than 90,000 students in grades 7 through 12. The students were born between 1976 and 1984 and attended one of 132 schools in the sampling frame.

[3]This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

About 20,000 respondents were followed in subsequent surveys; the last completed survey (Wave IV) was conducted in 2007-08 when the respondents were between 24 and 32 years of age.

Most of our analysis is based on a subsample of white, non-Hispanic men and women. Father-absent households are much more prevalent in the Black and Hispanic Add Health samples, and school and neighborhood characteristics are also very different on average. Because our focus is on gender differences in the impact of adolescent environments, and these may differ between, for example, households with foreign-born vs. native-born parents or schools with disciplinary regimes of varying harshness, we have chosen to focus on a more homogenous core sample. Key results for the Black subsample are presented in section 2.4.5; the Hispanic subsample is too small to support a separate analysis.

Table B.1 illustrates the selection of our Add Health analysis sample: the columns present mean values for the full Add Health Wave I sample, the full sample with complete data on key variables, the white, non-Hispanic sample, this sample with only those living with biological mother at Wave I and with non-missing maternal characteristics, and finally the sample remaining in the survey at Wave IV. This, the final analysis sample, contains 3,868 non-Hispanic white women and 3,459 non-Hispanic white men. With the exception of the race/ethnic subsampling, these sample restrictions have very little impact on the demographics of the sample.[4] The descriptive statistics by gender for this sample are summarized in Table B.2.

[4]The discrepancy in the male-female sample sizes is the result of consistently lower rates of both contact and response for male Add Health sample members. Attrition from the panel is higher for men than for women: in section 2.3 we show that this differential attrition is independent of family structure and examine other gender selection issues.

Adolescent Outcomes

The Add Health Wave I survey collected an array of student-reported variables, including experiences in school, health, personality, and relationships with parents, siblings, friends and others. Most respondents were between the ages of 12 and 17. We use self-reports of Math and English grades and of school problems, including school suspensions, to generate academic and behavioral outcomes that are similar to those in previous studies such as Autor et al. (2019) and augment this with a standard depression scale and reports of educational aspirations. The principal advantage offered by the Add Health data over the Danish data used by Brenøe & Lundberg (2018), who also compared in the impact of family disadvantage in adolescence and adulthood is this broad array of school-age outcomes, compared to a restricted set of measures available in administrative data.

- *School Problems*: Students were asked about problems they experience in school, including trouble getting along with teachers and other students, trouble getting homework done and trouble paying attention in class (coded 0-4 from "never" to "every day"), how many times they have been absent without an excuse, and whether they have ever received an out-of-school suspension. Factor analysis was used to aggregate these measures into a standardized school problems index.[5] Misbehavior in school, including absenteeism, inattention, and suspensions, are strongly predictive of adult labor market outcomes and criminal behavior (Segal 2013, Lundberg 2017*b*) and school suspensions are a key outcome in both Bertrand & Pan (2013) and Autor et al. (2019). In our sample, the proportion of boys who report ever having been suspended is more than twice that of girls, and their level of school problems is 0.3 standard deviations higher (Table B.2).

[5]The data supports a single-factor model of school problems, and a Cronbach's Alpha of 0.6694 indicates an adequate degree of internal consistency for this index.

- *Depression*: Wave I respondents were asked how often during the past week they felt sad, lonely, depressed, blue, happy, or hopeful. These six items (plus 13 more) are the components of CES-D, a standard depression scale (Radloff 1977). Factor analysis indicated that a single factor is appropriate for these 19 items and was used to form a standardized depression index. Table B.2 shows that reported depression among adolescent girls is much higher than among boys.

- *Grades and Aspirations*: As a measure of academic achievement, we use student reports of their math and English grades in the most recent grading period. Girls' mean English grade is 0.375 grade points higher than boys', but their math grades are only slightly higher on average (Table B.2). Educational aspirations in Wave I are based on student responses (on a 5-point scale) as to how much they want to attend college, and how likely they think it is that they will attend college. Add Health respondents are very optimistic about their college prospects on average, but girls are substantially more likely to report that they both want and expect to attend. Fortin et al. (2015) find that much of the gender gap in high school achievement can be attributed to the gender difference in educational expectations, particularly those linked to career plans that include a graduate degree. Add Health students were also asked about how likely they are to be married at age 25, on a 5-point scale.

Adult Outcomes

Our principal goal is to examine whether deficits in early skills and aspirations due to family disadvantage have long-term implications for gender gaps in economic and social outcomes in adulthood. The Wave IV survey collected an array of adult outcomes, including educational attainment, employment and income, and family histories.

- *Educational Attainment*: Highest educational attainment is collected when most respondents are between 25 and 31 years of age. Most, though not all, will have completed their final level of formal schooling at this point. We focus on the attainment of a 4-year college degree, since the rising returns to education in recent decades have largely been restricted to college graduates. Though there is a gender gap in high school graduation and college attendance as well, the college graduation gap has received the most attention given its substantial implications for lifetime income. However, we also examine high school graduation, imputed years of schooling, and a categorical education variable that ranges from 0=less than high school to 5=post-graduate degree. As expected, we see a moderate gender gap in high school graduation rates and a substantial one in college graduation in our analysis sample (Table B.2).

- *Employment and Income*: Deficits in non-cognitive skills and limited schooling are likely to lead to adverse labor market outcomes. We examine self-reported before-tax personal earnings, and define respondents as currently employed if they report working for pay at least 10 hours a week. We also consider two other aspects of employment histories: number of times fired[6] and satisfaction with current or last job[7], as these may reflect non-cognitive skills. A dummy variable for financial stress is based on respondent reports that they have faced difficulties in paying bills in the past year.[8] This measure is likely to be driven both by economic resources

[6]The survey question asks, "Thinking back over the period from 2001 to the previous year, how many times have you been fired, let go or laid off from a job?"

[7]The survey question asks "How satisfied (are/were) you with this job, as a whole?" on a 5 point scale.

[8]The financial stress dummy is set equal to one if there is a positive response to at least one of: "without phone service because you didn't have enough money", "didn't pay the full amount of the rent or mortgage because you didn't have enough money", "were evicted from your house or apartment for not paying the rent or mortgage", "didn't pay the full amount of a gas, electricity, or oil bill because you didn't have enough money", "had the service turned off by the gas or electric company, or the oil company wouldn't deliver, because payments were not made", and "worried whether food would run

and by skills associated with managing those resources. Male respondents are, as expected, more likely to be employed than women, and also report higher earnings and lower levels of financial stress. Men are more likely to report being fired than women and have roughly equivalent levels of job satisfaction.

- *Marriage and Children Ever Born*: Marriage and fertility histories are collected in the Wave IV survey. Half of the Add Health respondents have never been married at the time of Wave IV survey; we focus on a dummy variable for ever married as our key outcome. The respondents were also asked about the number of times they have been pregnant/have made a partner pregnant, and the live births resulting from these pregnancies. Add Health men are less likely to have been married and report fewer children than women, reflecting the expected gender differences in the timing of family formation.

- *Depression*: The Wave IV survey includes a shorter (11 item) version of the depression instrument in Wave I, and we include this in the analysis to see how persistent family and environmental influences on this mental health indicator are.[9]

Indicators of Disadvantage in Adolescence

- *Father Absence in Wave I:* A large literature has documented the empirical relationship between single parenthood, family instability, and a child's prospects for success in adulthood (McLanahan & Sandefur 1994, Lopoo & DeLeire 2014, Woessmann et al. 2015), though causal inference has been difficult given the confounding effects of unobserved parental, child, and environmental characteristics. The restricted economic and parental resources in single-parent families are very

out before you would get money to buy more".

[9]Factor analysis indicated that a single factor is appropriate for these 11 items and was used to form a standardized depression index.

likely, however, to limit investments in children and adolescents. Though nearly 90 percent of the Add Health respondents were living with their biological mother in Wave I, almost 10 percent were also living with a step-father or other father figure rather than their biological or adoptive father, and nearly 22 percent were living with no father figure at all. Girls were 15 percent more likely to be living with no father in the household than were Add Health boys.

- *School Quality:* Autor et al. (2019) find that poor quality schools have a particularly disadvantageous effect on school outcomes for boys and that this environmental influence is distinct from the effects of family disadvantage. The Add Health Study includes a school administrator questionnaire that can be used to construct a standardized index of school quality for the schools attended in Wave I.[10] The components of the index are average daily attendance, class size, percentages of new and of experienced teachers, the share of teachers with a Masters' degree, grade 12 dropout rates, percentages of students with standardized achievement tests at, below, or above grade level, and the share of 12th graders who enrolled in a 2-year or 4-year college the next year.

- *Neighborhood:* A growing literature finds that neighborhood characteristics appear to have long-term causal impacts on economic outcomes for children, and that boys may be differentially sensitive to these forces (Chetty et al. 2016, Chetty & Hendren 2018, Autor et al. 2019). The Add Health Contextual Database provides an array of community characteristics that enable researchers to investigate contextual influences for a wide range of adolescent behaviors.[11] We use an indicator for "educated neighborhood" defined as the proportion of individuals aged 25+ with

[10]Factor analysis indicates a single-factor model is appropriate, with a Cronbach's Alpha of 0.6229.

[11]For most respondents participating in the Add Health in-home survey, Wave I home locations were identified. When possible, these residence locations have been geocoded in order to link them to contextual data that is available from many other sources.

a college degree or more at the census tract level.

Maternal Characteristics

Maternal characteristics are included as control variables in most regressions. The educational attainment of the respondent's biological mother is divided into 4 categories, "less than high school", "high school degree", "some college", and "college degree or more". We also included indicators for whether the biological mother is foreign-born and young (under age 22) at the birth.

2.2.2 NLSY97 Sample

We examine the robustness of our analysis of Add Health with the National Longitudinal Survey of Youth 1997 (NLSY97), which provides comparable measures of adult (but not comparable adolescent) outcomes for a set of birth cohorts similar to Add Health. NLSY97 is a representative longitudinal study with surveys from 1997 (round 1) to 2015-2016 (round 17). The cohort was born between 1980 and 1984, with respondents aged between 12 and 18 at the time of the first interview and between 30 and 36 at round 17. As in our Add Health models, we analyze a subsample of non-Hispanic white women and men. Details regarding the sample, variable construction, and a comparison with the Add Health sample are in the Data Appendix.

2.3 Empirical Strategy

Obtaining a causal estimate of the difference in the impacts of father absence and other indicators of family disadvantage on outcomes for boys and girls requires that the distribution of male and female children across households with and without fathers be identical in terms of their potential outcomes with a father present. For any outcome Y

for boys (b) and girls (g), we can define possible outcomes in alternative family structures as:

$$Y_b = (1 - D)Y_b(0) + DY_b(1)$$

$$Y_g = (1 - D)Y_g(0) + DY_g(1)$$

where $Y_i(0)$ is the potential outcome of child i if his or her father is present in the household ($D = 0$), and $Y_i(1)$ is the potential outcome if his or her father is absent ($D = 1$). Y_i is the observed outcome.

In general, the causal impact of father absence cannot be determined by comparing outcomes for children in different types of household due to the confounding effects of unobserved parental, child, and environmental influences. The average difference in outcomes between boys in father-absent and father-present households is:

$$E(Y_b|D = 1) - E(Y_b|D = 0)$$

$$= E(Y_b(1)|D = 1) - E(Y_b(0)|D = 1) + E(Y_b(0)|D = 1) - E(Y_b(0)|D = 0)$$

The first term is the average causal impact of father absence for boys raised in father-absent households; the second term is selection bias–the difference between potential outcomes in the father-present state between boys who were raised in that state and boys who were not. This will generally be non-zero, and any estimate of the effect of father absence will be biased if there are unobserved differences in child capabilities and mother characteristics in father-present and father-absent households. This is true for girls as well.

However, if the selection terms are identical for boys and girls, an estimate of the gender difference in the effects of father absence will be unbiased. If we have:

Assumption 2.3.1 (Non-differential Selection)

$$E(Y_b(0)|D=1) - E(Y_b(0)|D=0) = E(Y_g(0)|D=1) - E(Y_g(0)|D=0)$$

then the gender difference in the causal effects of father absence is identified (Equation 2.3.1). Alternatively, we can define under Assumption 2.3.1 an unbiased estimate of the causal effect of father absence on gender gaps (Equation 2.3.2).

$$[E(Y_b|D=1) - E(Y_b|D=0)] - [E(Y_g|D=1) - E(Y_g|D=0)]$$

$$= [E(Y_b(1)|D=1) - E(Y_b(0)|D=1) + E(Y_b(0)|D=1) - E(Y_b(0)|D=0)]$$

$$-[E(Y_g(1)|D=1) - E(Y_g(0)|D=1) + E(Y_g(0)|D=1) - E(Y_g(0)|D=0)]$$

$$= E(Y_b(1) - Y_b(0)|D=1) - E(Y_g(1) - Y_g(0)|D=1) \tag{2.3.1}$$

$$= E(Y_b(1) - Y_g(1)|D=1) - E(Y_b(0) - Y_g(0)|D=1) \tag{2.3.2}$$

The main econometric specification in this paper is:

$$Y_i = \alpha_0 + \alpha_1 Male_i + \alpha_2 NF_i + \alpha_3 OF_i + \beta_1 Male_i \times NF_i + \beta_2 Male_i \times OF_i + \gamma X_i + \varepsilon_i \tag{2.3.3}$$

where NF_i is a dummy variable equal to one if child i lived in a household with no father figure in the baseline survey and OF_i is equal to one if a non-biological, non-adoptive father, such as a step-father, lived in the household. X_i includes maternal characteristics and the child's birth cohort. Standard errors are clustered by the school attended in Wave I. The coefficients of interest–β_1 and β_2–identify the gender difference in the causal effects of father absence and other father, under Assumption 2.3.1.

Therefore, the interpretation of the results as causal relies on the key identification assumption that selection into father absence and other father households is identical for boys and girls. Estimates may not be unbiased if father absence and child gender are not independent, and the fact that girls are more likely to live in father-absent households than boys raises the possibility of selection on child or maternal characteristics. One mechanism driving this gap may be parental decisions about marriage and custody that are child gender-specific: fathers are more likely to co-reside with, seek custody of, and marry the mothers of their sons rather than their daughters (Lundberg & Rose 2003, Dahl & Moretti 2008, Lundberg 2005). Another may be through the effects of parental circumstances on the gender mix of offspring: evidence is mounting that prenatal stress (which may be related to partnership status) has differential impacts on the mortality of male and female fetuses, though the effects are small (Almond & Edlund 2007, Hamoudi & Nobles 2014, Norberg 2004).

Table B.4 presents a test for the identification assumption. If selection into father absent and other father households is identical for boys and girls, then the gaps in pre-determined characteristics between father-absent/other-father household and father-present household should be the same for boys and girls. We do find that mother's characteristics and family income is often significantly associated with family structure, but there is no evidence that the selection is different across gender: none of the interaction terms are significantly different from 0.[12] Another possible concern is selective attrition in this longitudinal study. We define attrition as the sample size reduction when conditioning on appearance in Wave IV (i.e. the change from column (4), Table B.1 to

[12]Autor et al. (2019) and Brenøe & Lundberg (2018) show that there are no gender differences in the effects of family disadvantage on outcomes at birth, which suggests an absence of selection on child capability, but Add Health does not have similar early measures except for one measure (self-reported birth weight, column (7)).

column (5), Table B.1). As shown in Column (8), Table B.4, boys and adolescents in father-absent households are more likely to attrit before Wave IV, but attrition is not different across gender/family structure categories. These results indicate that there is no evidence that our identification assumption is violated in this sample.[13,14]

2.4 Results

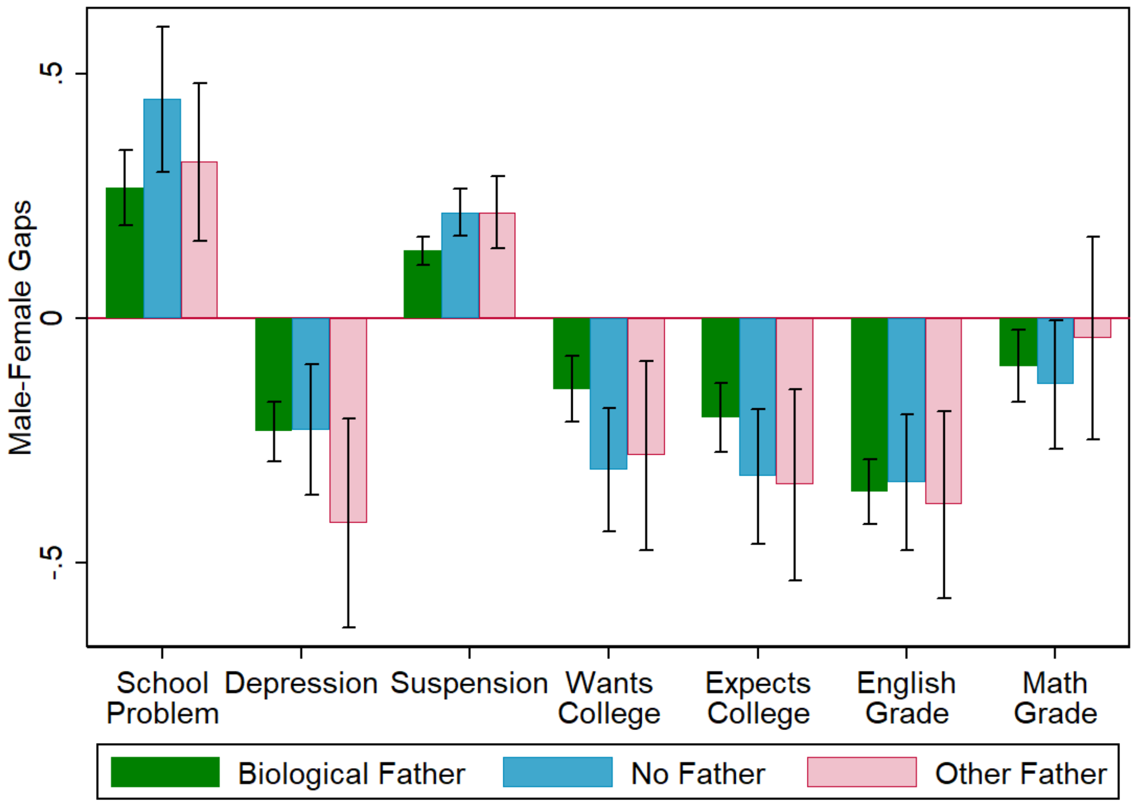
2.4.1 Adolescent Outcomes

Gender gaps in the behavior and achievements of school children in the Add Health sample are shown in Figure 2.1 for the three family structure groups: biological father present, no father, and other father. The gender gaps in the school problems index, depression index, school suspensions, and educational aspirations are relatively larger for respondents in non-traditional families. Therefore, Figure 2.1 provides some raw evidence that the effects of family structure on adolescent outcomes may differ by gender.

[13]We also did similar identification tests for NLSY97 sample (not reported here). None of the interaction terms is significantly different from 0, and so there is also no evidence of violation of the identification assumption in this sample.

[14]An alternative approach to identifying gender differences in the effects of disadvantage is to use family fixed-effects, as do Lundberg (2017*a*) and Brenøe & Lundberg (2018). The subsample of gender-discordant siblings in these data, however, consists of only 351 observations from 168 families and models for adolescent outcomes have large standard errors so we have not included these results.

Figure 2.1: Gender Gaps in Adolescent Outcomes by Family Structures, Add Health Non-Hispanic White Sample



Notes: This figure displays the coefficient of “Male” dummy, when regressing the outcomes on “Male” dummy and a constant, with its 95% confidence interval. “No Father” and “Other Father” refer to living arrangements at Wave I. “School problems” is a standardized index based on factor analysis of the variables in Table B.5. “Depression” is a standardized index based on factor analysis of the variables in Table B.6. Grades are student-reported and range from 1=D or lower to 4=A. College desires/expectations are standardized measures based on a 0-4 scale. All models are weighted by Wave I weights.

Estimates of Equation 2.3.3 showing the effects of father absence in Wave I on adolescent outcomes are reported in Table 2.1. Columns (1) and (2) show that living in a father absent household or step-father household is positively associated with school problems and school suspensions, and that this association (particularly for no-father households) is significantly greater for boys than for girls.[15] The results for school suspensions in particular are strongly consistent with the findings of both Bertrand & Pan (2013) and Autor et al. (2019).[16] In these results we see some evidence of differential male susceptibility to non-traditional family structures.

A different picture emerges when we look at another set of Wave I self-reports: depression in adolescence. Column (3) shows the effects of family structure on the depression index. Boys are significantly less likely than girls to report experiencing negative emotions, and youth in no-father and step-father families are more likely to make such reports. We find that depression is more strongly associated with living with a step-father or other father figure for girls than for boys, and the interaction term is also significant for several depression index components (Table B.6). Depression is one example of an "internalizing" response to stress that is more common for girls, as opposed to the "externalizing" or disruptive behavior more typical of boys (Leadbeater et al. 1999).

Family structure does not appear to have any differential effect on self-reported grades in English and Math, though we find the usual pattern that boys' grades are lower than girls, particularly in English (Column (4)-(5), Table 2.1). When asked in Wave I about their college plans, Add Health boys are less likely than girls to report either that they want to attend college or that they expect to attend college (Column (6)-(7), Table 2.1). In this case, living in a household with no father appears to have a more severe effect on the college intentions of boys–they are substantially less likely to report a strong desire to attend college than girls in similar families.

The results in this section both reinforce and expand upon the findings of previous

[15] Table B.5 shows the determinants of the school problems index and its components. Male students report higher incidence of all individual school difficulties except absences. The male/no-father interaction is significantly predictive of school suspensions, reported problems paying attention in school, and the overall school problems index.

[16] Bedard & Witman (2020) find that the gender gap in diagnosis and treatment of ADHD is much larger in non-traditional families, another result that suggests parents in traditional families may find it easier to cope with male behavioral difficulties in early life.

Table 2.1: Adolescent Outcomes, Add Health Non-Hispanic White Sample

VARIABLES	(1) School Problem Index	(2) Ever Suspended from School	(3) Depression Index	(4) English Grade	(5) Math Grade	(6) Wants to Attend College	(7) Expects to Attend College	(8) Chance of Being Married by Age 25
Male	0.263***	0.136***	-0.240***	-0.361***	-0.0925**	-0.143***	-0.216***	-0.119***
	(0.0384)	(0.0140)	(0.0295)	(0.0320)	(0.0390)	(0.0319)	(0.0342)	(0.0297)
Male*No Father	0.179**	0.0741***	-0.000464	0.0365	-0.0261	-0.147**	-0.0943	0.0232
	(0.0787)	(0.0263)	(0.0663)	(0.0771)	(0.0720)	(0.0652)	(0.0763)	(0.0771)
Male*Other Father	0.0466	0.0682*	-0.205*	0.00603	0.0738	-0.102	-0.0930	-0.0232
	(0.0884)	(0.0368)	(0.115)	(0.0969)	(0.120)	(0.104)	(0.105)	(0.0895)
No Father	0.175***	0.102***	0.184***	-0.248***	-0.198***	-0.00407	-0.167***	-0.243***
	(0.0569)	(0.0191)	(0.0489)	(0.0529)	(0.0518)	(0.0476)	(0.0528)	(0.0525)
Other Father	0.146**	0.0366*	0.272***	-0.0827	-0.144	0.0215	-0.0502	-0.174***
	(0.0702)	(0.0191)	(0.0969)	(0.0641)	(0.0886)	(0.0619)	(0.0769)	(0.0627)
Constant	-0.999***	-0.505***	-1.651***	3.478***	3.974***	1.478***	0.169	1.370***
	(0.353)	(0.143)	(0.260)	(0.333)	(0.376)	(0.264)	(0.277)	(0.285)
Observations	7,203	7,325	7,311	7,068	6,755	7,309	7,308	7,291
R-squared	0.050	0.120	0.063	0.091	0.038	0.090	0.136	0.020
Mean of dependent variable	0.000	0.207	0.000	2.918	2.789	0.000	0.000	0.000
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "No Father" and "Other Father" refer to living arrangements at Wave I. "School problems" is a standardized index based on factor analysis of the variables in Table B.5. "Depression" is a standardized index based on factor analysis of the variables in Table B.6. Grades are student-reported and range from 1=D or lower to 4=A. College desires/expectations and expectation of chance of being married by age 25 are standardized measures based on a 0-4 scale. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave I weights.

studies that show excess vulnerability of school-aged boys in the face of family disadvantage and father absence. The gender gap in school problems is much greater for adolescents who are not living with both biological parents, and this pattern is consistent with earlier studies that find increasing gender gaps in schools suspensions and externalizing behavior. Examining an aspect of problematic internalizing behavior, depression, indicates that girls may have distinctive responses to family disadvantage not reflected in standard measures of school achievement and disciplinary outcomes. These contrasting results show that our conclusions about which gender is more sensitive to father absence may depend on which school outcomes we are measuring.

What are the mechanisms underlying these results? One possible source of these differences may be sex differences in early developmental trajectories that have implications for skills in adolescence. Beginning in preschool, girls are more mature than boys in language skills and emotional regulation, and this may increase their resilience in some adverse circumstances. The absence of a stable, same-sex parent may also have distinct behavioral effects on boys, as the presence of a step-father appears to have for girls.[17] A key implication here is that disadvantaged boys may be left with a deficit of skills, particularly non-cognitive skills, that can further disadvantage them in adulthood. In contrast, a cultural explanation is provided by DiPrete & Buchmann (2013), who argue that developing a masculine self-image may involve a rejection of school values, and that this "oppositional culture" may be particularly relevant for boys with absent or low-education fathers. The effects of father absence on boys' desire to attend college may provide some support for this mechanism.

Finally, parental investments in low-resource environments may differ by child gender. Though a large literature shows that, on average, fathers spend more time with sons

[17] Others have argued that school environments, with predominantly female teachers, fail to adapt to the learning needs of boys (Dee 2007).

than with daughters, and that this gap grows with age (Lundberg 2005), Bertrand & Pan (2013) find that single mothers spend less time with sons than daughters and report less emotional closeness with sons in early school years. Such a result suggests a parental investment variant of the Trivers-Willard hypothesis from evolutionary biology: parents who are maximizing reproductive success invest more in male offspring in good conditions but more in females in poor conditions (Trivers & Willard 1973). Explicit attempts to test for evidence of Trivers-Willard patterns in modern families, however, have not found it to be well-supported (Keller et al. 2001).

The Add Health survey has limited direct measures of parental inputs, but does include multiple indicators of the quality of the parent-child relationship, which may be related to parental investments. Adolescent reports in Wave I about their relationships with parents do not show any evidence of such distinctive boy-girl responses to father absence (Table B.7). Children in father-absent families are less likely to report that their parents care about them, that their family generally has fun, that their mothers are warm and loving towards them, and that they are satisfied with their relationship with their mother, but we find no significant gender differences in these family structure effects.

2.4.2 Educational Attainment

Do the gender-differential effects of father absence in adolescence persist into adulthood? The effects of father absence in adolescence on several measures of educational attainment from Wave IV of the Add Health Study, when the respondents are in their late twenties and early thirties, are reported in Tables 2.2-2.3. Table 2.2 shows that being male has a large negative effect on the probability that an Add Health respondent receives a 4-year college degree. In the initial model with no other covariates, the college gender gap is 7 percentage points and controlling for mother's characteristics (Columns

(2)-(6)) has little effect on this gap. The coefficients on dummy variables for living in a family with no father figure or with a non-biological (step) father figure in Wave I are also large and negative.

Table 2.2: College Graduation, Add Health Non-Hispanic White Sample

VARIABLES	(1) Living with Bio-mon in Wave I	(2)	(3)	(4) Mother High School	(5) Mother Some College	(6) Mother College Grad
Male	-0.0702***	-0.0893***	-0.0891***	-0.0600***	-0.0916*	-0.136***
	(0.0169)	(0.0178)	(0.0179)	(0.0215)	(0.0489)	(0.0338)
Male*No Father		0.0397		0.0352	-0.0212	0.0691
		(0.0296)		(0.0358)	(0.0826)	(0.0750)
Male*Other Father		0.0366	0.0365	0.0231	0.00942	0.0802
		(0.0423)	(0.0423)	(0.0460)	(0.0962)	(0.101)
No Father		-0.149***		-0.130***	-0.133***	-0.235***
		(0.0209)		(0.0241)	(0.0467)	(0.0556)
Other Father		-0.126***	-0.127***	-0.107***	-0.0958	-0.207***
		(0.0264)	(0.0264)	(0.0333)	(0.0639)	(0.0670)
Male*No Father Recently			0.0326			
			(0.0421)			
Male*No Father Always			0.0368			
			(0.0329)			
No Father Recently			-0.106***			
			(0.0283)			
No Father Always			-0.192***			
			(0.0235)			
Constant	0.155	-0.123	-0.105	-0.00228	-0.0192	0.590**
	(0.217)	(0.168)	(0.167)	(0.204)	(0.299)	(0.261)
Observations	7,327	7,327	7,327	3,932	1,468	1,922
R-squared	0.006	0.171	0.172	0.037	0.038	0.063
Mean of dependent variable		0.368		0.235	0.362	0.643
Mother's characteristics	NO	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at Wave I. "No Father Always" means the adolescent has not lived with his/her biological father since the age of 5. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave IV weights.

Non-traditional family structures do not, however, have differentially negative impacts on the college graduation rates of young men (Column (2)). The interaction effects, expected to be negative if boys are more vulnerable to father absence, are instead positive and insignificant. Column (3) decomposes the "no father" group into young adults who, though they did not live with their biological father at Wave I, did do so after the age of 5 (No Father Recently) and those who never lived with their father after age 5 (No

Father Always). The latter status, as expected, has a larger negative association with college graduation but the gender interaction effects are once again positive, small and insignificant. Columns (4)-(6) report results from the core model for subsamples based on mother's education level, and the pattern is similar–negative effects of non-traditional family structures, but no evidence that the college graduation rates of men are more strongly affected by father absence than is college graduation by women.

The results in Table 2.3 show similar patterns in the determinants of high school graduation. Men are less likely to graduate from high school than women, living with no father or a step-father in Wave I has a strong negative association with graduation. There is weak support for greater male vulnerability to disadvantage: the interaction term of Male and No Father is marginally significant at the 10% level. The impact of father absence before age 5 (No Father Always) on men's high school graduation, compared to women's, is also marginally significant at a 10% level. Columns (4)-(6) split the sample by mother's education and show that the effects of family structure on men's high school graduation appear to be concentrated in families in which the mother had some college education, though none of the interaction terms are individually significantly in these subsamples.

Table B.8 reports the results of key models for two alternative measures: years of completed education and a categorical measure of educational attainment that ranges from 0=less than high school to 5=post-graduate degree. The pattern of coefficients is very similar to that for college graduation: substantial negative effects of being male and living without a father in adolescence, but no differential impacts of family structure by gender. The interaction effects are small relative to the male effects and are not significant at conventional levels.

In general, the evidence from the Add Health cohorts of young adults strongly suggests that, though being male and living in a household without a biological father in adoles-

Table 2.3: High School Graduation, Add Health Non-Hispanic White Sample

	(1)	(2)	(3)	(4)	(5)	(6)
		Living		Mother	Mother	Mother
VARIABLES		with Bio-mom in Wave I		High School	Some College	College Grad
Male	-0.0251***	-0.0235***	-0.0235***	-0.0258*	-0.0110	-0.0140**
	(0.00821)	(0.00837)	(0.00837)	(0.0151)	(0.0130)	(0.00576)
Male*No Father		-0.0327*		-0.0298	-0.0720	-0.00242
		(0.0196)		(0.0291)	(0.0452)	(0.0266)
Male*Other Father		0.0171	0.0170	0.0433	-0.0171	-0.0203
		(0.0272)	(0.0273)	(0.0484)	(0.0360)	(0.0310)
No Father		-0.0336**		-0.0592***	-0.0257	-0.0124
		(0.0129)		(0.0209)	(0.0181)	(0.00969)
Other Father		-0.0379**	-0.0380**	-0.0663**	0.00108	0.000956
		(0.0176)	(0.0176)	(0.0324)	(0.0178)	(0.00198)
Male*No Father Recently			-0.0116			
			(0.0240)			
Male*No Father Always			-0.0597*			
			(0.0332)			
No Father Recently			-0.0293*			
			(0.0155)			
No Father Always			-0.0380*			
			(0.0204)			
Constant	0.796***	0.607***	0.612***	0.761***	0.809***	0.906***
	(0.0859)	(0.0774)	(0.0788)	(0.117)	(0.126)	(0.0497)
Observations	7,327	7,327	7,327	3,932	1,468	1,922
R-squared	0.003	0.080	0.081	0.023	0.032	0.012
Mean of dependent variable		0.935		0.900	0.958	0.988
Mother's characteristics	NO	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at Wave I. "No Father Always" means the adolescent has not lived with his/her biological father since the age of 5. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave IV weights.

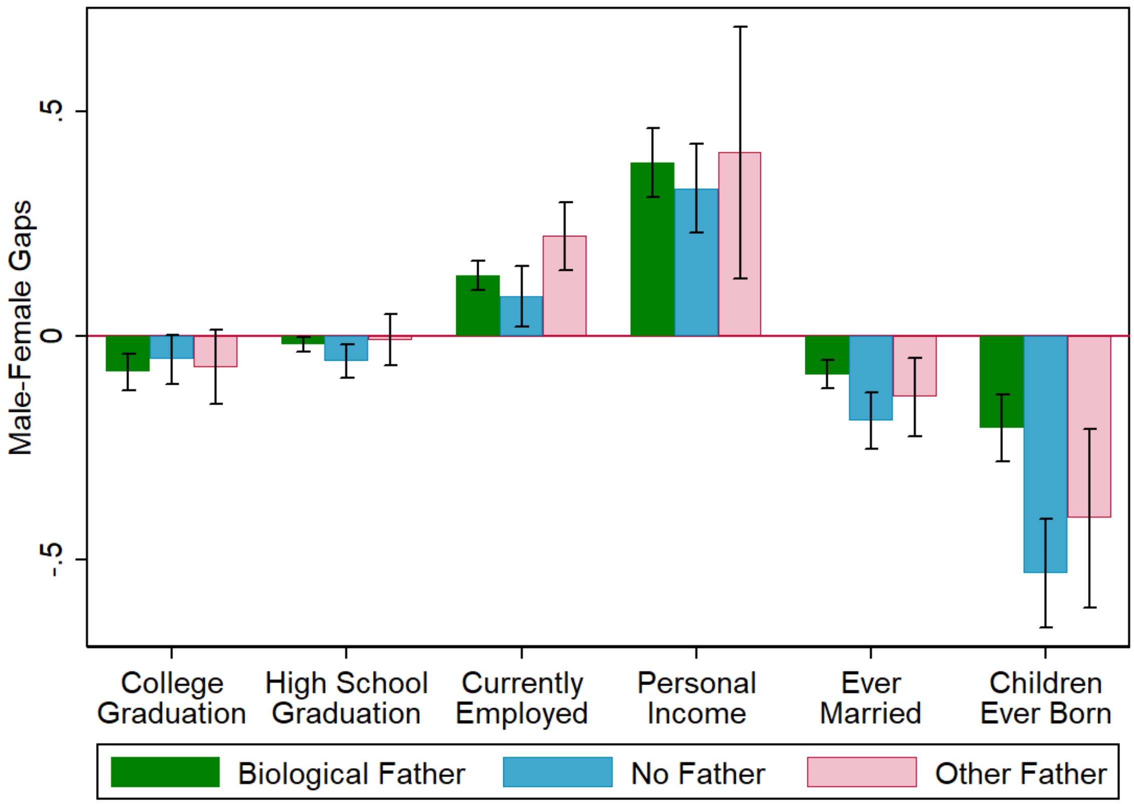
cence are negatively associated with educational attainment, young men do not appear to be differentially affected by father absence when we focus on long-term outcomes such as college graduation. There is some limited evidence that high school graduation, which is likely to be more directly affected than college graduation by grade school achievement and misbehavior, may be a hurdle for which father presence is more important for boys.

2.4.3 Other Adult Outcomes: Labor Market, Family, and Mental Health

Gender gaps in key adult outcomes for different family structures are shown in Figure 2.2. Most gender gaps are prevalent across all types of family, though gaps for family outcomes appear to be larger for respondents who grew up in non-traditional families. Table 2.4 reports the estimated effects of father absence in adolescence on adult outcomes.

In general, growing up in a father-absent or a step-father household is associated with negative labor market and other economic outcomes for both men and women, including employment, income, job satisfaction, job terminations, and a measure of financial stress (Columns (1)-(5)). For most of these outcomes, the effects do not differ significantly by gender. With the exception of number of times fired, the gender/father presence interaction terms are small and all are imprecisely estimated. One exception is that, though other father figures have a weakly negative association with the likelihood of being currently employed for women, this is not true for men: the stepfather-male interaction term is positive and significant at the 5% level. Also, the negative effect of a no-father household on financial stress is significantly smaller for men than for women. This evidence shows that the effects of non-traditional family structure on most labor market outcomes, as with educational attainment, do not differ across genders. When they do differ significantly, it is to the benefit of men rather than women.

Figure 2.2: Gender Gaps in Adult Outcomes by Family Structures, Add Health Non-Hispanic White Sample



Notes: This figure displays the coefficient of "Male" dummy, when regressing the outcomes on "Male" dummy and a constant, with its 95% confidence interval. "No Father" and "Other Father" refer to living arrangements at Wave I. "Currently Employed" is a dummy for whether is currently working at Wave IV. "Personal Income" is the total personal earnings before tax in 2006/2007/2008. "Ever Married" is a dummy for whether has been married up to Wave IV. "Children Ever Born" is the self-reported total number of child births up to Wave IV. All models are weighted by Wave IV weights.

Men typically marry at later ages than women, and we find that male Add Health respondents are less likely to have ever been married than female respondents at the same age (Column 6) and also report fewer children (Column 7). Marriage probabilities for women who grew up in father-absent and step-father households are not significantly different from those in bio-father households, but father absence is associated with a

Table 2.4: Other Adult Outcomes, Add Health Non-Hispanic White Sample

VARIABLES	(1) Currently Employed	(2) Personal Income	(3) Number of Times Fired	(4) Job Satisfaction	(5) Financial Stress	(6) Ever Married	(7) Children Ever Born	(8) Depression Index
Male	0.130***	14.71***	0.332***	0.0118	-0.0117	-0.0988***	-0.215***	-0.106***
	(0.0167)	(1.505)	(0.0457)	(0.0294)	(0.0141)	(0.0164)	(0.0397)	(0.0348)
Male*No Father	-0.0407	-1.970	0.199	-0.0205	-0.0564*	-0.103***	-0.333***	-0.0265
	(0.0406)	(2.308)	(0.121)	(0.0765)	(0.0317)	(0.0318)	(0.0762)	(0.0739)
Male*Other Father	0.0930**	1.785	-0.0424	0.0468	0.0328	-0.0493	-0.222**	-0.274**
	(0.0390)	(5.828)	(0.117)	(0.0960)	(0.0500)	(0.0444)	(0.0967)	(0.116)
No Father	-0.0208	-5.372***	0.0719**	-0.0679*	0.140***	0.00727	0.306***	0.158***
	(0.0253)	(1.237)	(0.0336)	(0.0391)	(0.0226)	(0.0230)	(0.0611)	(0.0493)
Other Father	-0.0624*	-3.496	0.103**	-0.205***	0.0633**	0.00555	0.231***	0.258***
	(0.0339)	(2.462)	(0.0502)	(0.0633)	(0.0311)	(0.0348)	(0.0787)	(0.0935)
Constant	0.401***	-43.14***	0.621*	3.489***	0.451***	-1.143***	-1.071***	0.152
	(0.141)	(11.30)	(0.355)	(0.253)	(0.124)	(0.165)	(0.332)	(0.302)
Observations	6,105	7,083	7,142	7,259	7,327	7,321	7,307	7,311
R-squared	0.046	0.062	0.026	0.005	0.038	0.073	0.109	0.025
Mean of dependent variable	0.788	36.060	0.467	3.909	0.216	0.560	0.847	0.010
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "No Father" and "Other Father" refer to living arrangements at Wave I. "Currently Employed" is a dummy for whether is currently working at Wave IV. "Personal Income" is the total personal earnings before tax in 2006/2007/2008. "Number of Times Fired" is the total number of times being fired since 2001. "Job Satisfaction" is on 5 point scale. "Financial Stress" is a dummy for whether has met liquidity constraints in the past year. "Ever Married" is a dummy for whether has been married up to Wave IV. "Children Ever Born" is the self-reported total number of child births up to Wave IV. "Depression Index" is a standardized index based on factor analysis of 11 variables, a subset of 19 Wave I variables. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave IV weights.

lower likelihood of being ever married for men. Gender differences also emerge for the number of reported children ever born: father absence and other father figures are significantly associated with more births for women and, since marriage probabilities remain unchanged, many of these are likely to be non-marital births. These positive effects are offset for men by the negative coefficients on the interaction terms, however, indicating that family background has no apparent effect on fertility for men. This reduction in marriage for men and increased fertility for women are generally consistent with Anderson (2017), who finds that adverse childhood environments are associated with increased sexual risk behaviors and greater likelihood of living in an unmarried couple for both men and women.

One possible mechanism for these effects on family dynamics is that non-traditional family structure in adolescence differentially changes the preferences and expectations about family formation and stability for boys and girls, and these gender differences persist into adulthood. An alternative explanation is that non-traditional family structure is associated with more negative marriage market characteristics, especially for men. Column (8), Table 2.1 provides some evidence that fails to support the former hypothesis. Although non-traditional family structure is negatively associated the Wave I reports regarding their chances of being married at age 25, the effects are not different across gender. Therefore, self-reports in adolescence do not indicate that the effects come from changes in marriage aspirations that differ between boys and girls. This is consistent with the findings of Kamp Dush et al. (2018), who attribute intergenerational correlations of partnering behaviors to the transmission of poor marriageable characteristics and relationship skills.[18]

Non-traditional family structure is associated with higher levels of depression in adult-

[18]If the gender difference in marital behavior reflects a more general increase in relationship instability for young men, then some part of the apparent gender gap in the fertility effects of father absence may be due to underreporting of children not born into committed relationships.

hood (Column 8). In addition, the association between step-father households and depression, which was positive for girls but not for boys in adolescence, appears to be a pattern that persists into adulthood.

The key takeaway from these analyses is that, though the absence of a biological father appears to have some persistent effects that differ by gender, particularly for family formation in adulthood, these effects do not fall readily into a vulnerable boys framework. There are few gendered effects on economic and labor market outcomes, but these tend to disadvantage women, and step-father families have a persistent association with depression in adulthood only for women.

2.4.4 School Quality and Neighborhood Effects

The "male vulnerability" hypothesis has been studied primarily in terms of adolescent responses to family disadvantage, but Autor et al. (2019) have also found that boys appear to be more sensitive than girls to variations in school quality and neighborhood characteristics in terms of test scores, absences, and suspensions, though these environmental factors explain only a small portion of the family SES contribution to this gap.

The estimates in Table 2.5 examine whether the short-run and long-run outcomes of male students in Add Health are more responsive to variations in school quality than are outcomes for female students. The school quality index is strongly associated with a lower probability of school suspension, higher educational aspirations, higher grades in adolescence, higher high school and college graduation rates, and lower fertility. For adolescent outcomes, some of the gender interaction effects are also significant. The gender gaps in the college attendance desires and expectations and in school suspensions are much smaller in high-quality schools. In other words, boys are indeed more responsive to school quality than girls in terms of educational aspirations and suspensions. However, as

Table 2.5: School Quality Effects, Add Health Non-Hispanic White Sample

	Adolescent Outcomes							Adult Outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
VARIABLES	School Problem Index	Ever Suspended from School	Depression Index	Wants to Attend College	Expects to Attend College	English Grade	Math Grade	College Graduation	High School Graduation	Currently Employed	Personal Income	Ever Married	Children Ever Born
Male	0.212***	0.162***	-0.269***	-0.214***	-0.286***	-0.399***	-0.0687*	-0.0901***	-0.0280***	0.118***	15.68***	-0.122***	-0.272***
	(0.0377)	(0.0157)	(0.0332)	(0.0400)	(0.0361)	(0.0375)	(0.0392)	(0.0192)	(0.00676)	(0.0163)	(2.018)	(0.0165)	(0.0471)
Male*School Quality Index	-0.0168	-0.0319**	-0.0150	0.0710*	0.0860***	0.0386	0.0158	0.0106	0.0105	-0.00109	0.554	-0.00398	0.0787**
	(0.0315)	(0.0154)	(0.0330)	(0.0357)	(0.0268)	(0.0359)	(0.0312)	(0.0167)	(0.00993)	(0.0170)	(2.006)	(0.0184)	(0.0334)
School Quality Index	-0.0123	-0.0291**	-0.0145	0.0560**	0.0856***	0.0501*	0.0787**	0.0507**	0.0177**	0.0179	1.531	0.00139	-0.108***
	(0.0271)	(0.0137)	(0.0264)	(0.0267)	(0.0250)	(0.0264)	(0.0361)	(0.0208)	(0.00803)	(0.0139)	(1.247)	(0.0201)	(0.0368)
Male*No Father	0.186**	0.0560*	0.0236	-0.177**	-0.0903	0.0312	-0.138	0.0401	-0.0335	-0.0478	-1.219	-0.0697*	-0.298***
	(0.0926)	(0.0305)	(0.0775)	(0.0729)	(0.0913)	(0.106)	(0.0888)	(0.0372)	(0.0217)	(0.0380)	(3.203)	(0.0361)	(0.0942)
Male*Other Father	0.0969	0.0421	-0.279*	-0.0208	-0.0597	-0.00570	0.0331	0.0350	-0.00407	0.0891*	4.980	-0.0160	-0.146
	(0.0962)	(0.0485)	(0.140)	(0.0969)	(0.106)	(0.112)	(0.136)	(0.0464)	(0.0347)	(0.0479)	(8.544)	(0.0523)	(0.109)
No Father	0.158**	0.117***	0.186***	-0.0439	-0.227***	-0.264***	-0.151**	-0.132***	-0.0380***	-0.0186	-6.052***	-0.0344	0.270***
	(0.0618)	(0.0217)	(0.0576)	(0.0635)	(0.0716)	(0.0624)	(0.0671)	(0.0258)	(0.0122)	(0.0301)	(1.676)	(0.0279)	(0.0734)
Other Father	0.171*	0.0563**	0.390***	0.00280	-0.125	-0.0948	-0.125	-0.127***	-0.0374*	-0.0981**	-3.616	-0.0187	0.193**
	(0.0869)	(0.0262)	(0.128)	(0.0683)	(0.0840)	(0.0750)	(0.114)	(0.0269)	(0.0206)	(0.0383)	(3.411)	(0.0419)	(0.0899)
Constant	-0.456	-0.345**	-1.031***	0.870***	-0.727**	2.516***	3.077***	0.0107	0.616***	0.483***	-26.51	-0.897***	-0.827**
	(0.422)	(0.162)	(0.299)	(0.315)	(0.305)	(0.403)	(0.594)	(0.173)	(0.110)	(0.128)	(16.87)	(0.182)	(0.381)
Observations	5,207	5,314	5,306	5,305	5,304	5,098	4,809	5,316	5,316	4,433	5,149	5,313	5,302
R-squared	0.039	0.137	0.057	0.103	0.168	0.085	0.030	0.163	0.094	0.042	0.054	0.055	0.114
Mean of dependent variable	0.034	0.224	0.053	-0.048	-0.026	2.882	2.746	0.376	0.943	0.803	37.741	0.590	0.880
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "School Quality Index" is a standardized index based on school administrator reports of average daily attendance, class size, % of new and experienced teachers, % of teachers with a Masters' degree, grade 12 dropout rates, % of students with achievement tests below and above grade level, and % of 12th graders enrolled in college next year. "No Father" and "Other Father" refer to living arrangements at Wave I. "School problems" is a standardized index based on factor analysis of the variables in Table B.5. "Depression" is a standardized index based on factor analysis of the variables in Table B.6. Grades are student-reported and range from 1=D or lower to 4=A. College desires/expectations are standardized measures based on a 0-4 scale. "Currently Employed" is a dummy for whether is currently working at Wave IV. "Personal Income" is the total personal earnings before tax in 2006/2007/2008. "Ever Married" is a dummy for whether has been married up to Wave IV. "Children Ever Born" is the self-reported total number of child births up to Wave IV. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. Models (1)-(7) are weighted by Wave I weights and models (8)-(13) are weighted by Wave IV weights.

with father absence, these differential effects do not appear to have implications for eventual educational attainment, including high school and college graduation, employment, income and marriage. The only exception among adult outcomes is fertility: though the number of births by Wave IV is significantly reduced by higher school quality for women, the reported fertility of males is much less responsive to school quality.

Table 2.6 investigates whether neighborhood effects are different across gender, for short-run and long-run outcomes. The proportion of highly-educated people in the neighborhood adolescents live in is strongly associated with fewer suspensions, higher educational aspirations, and higher math grades in adolescence, and with higher high school and college graduation rates, higher income, higher employment probabilities, lower probabilities of marriage, and fewer births in adulthood. In adolescence, the educational aspirations of boys are significantly more responsive to neighborhood quality than those of girls. There is also some evidence that the relative math grades of girls are higher in highly-educated neighborhoods. These results indicate that a highly-educated neighborhood boosts the aspirations for higher education for boys relative to girls, but improves the math performance of girls relative to boys. These differential neighborhood effects don't persist into adulthood for college graduation, employment, income and marriage, but there is some evidence that an educated neighborhood benefits boys more than girls in terms of high school graduation. In addition, the fertility decisions of men are less responsive to neighborhood quality than women's, as is the case with school quality effects.

Table B.9 presents results from models that include both school quality and neighborhood effects. The results are robust, implying that although school quality and neighborhood education levels are correlated, they have different patterns of effects. In general, school quality and neighborhood effects on some adolescent outcomes, especially educational aspirations, suspension and math grades, differ across genders, implying that boys

Table 2.6: Neighborhood Effects, Add Health Non-Hispanic White Sample

	Adolescent Outcomes							Adult Outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
VARIABLES	School Problem Index	Ever Suspended from School	Depression Index	Wants to Attend College	Expects to Attend College	English Grade	Math Grade	College Graduation	High School Graduation	Currently Employed	Personal Income	Ever Married	Children Ever Born
Male	0.209***	0.129***	-0.228***	-0.234***	-0.300***	-0.356***	-0.00300	-0.0664***	-0.0426***	0.160***	12.24***	-0.0856***	-0.408***
	(0.0759)	(0.0254)	(0.0662)	(0.0596)	(0.0566)	(0.0565)	(0.0703)	(0.0245)	(0.0159)	(0.0271)	(2.118)	(0.0301)	(0.0617)
Male*Educated Neighborhood	0.220	0.0204	-0.0379	0.395**	0.372**	-0.0117	-0.363*	-0.0827	0.0800*	-0.109	10.31	-0.0588	0.762***
	(0.270)	(0.0732)	(0.250)	(0.169)	(0.151)	(0.178)	(0.196)	(0.0845)	(0.0435)	(0.0942)	(8.431)	(0.0942)	(0.169)
Educated Neighborhood	0.0137	-0.183***	-0.0867	0.444***	0.705***	0.232	0.425**	0.625***	0.0487*	0.218***	16.87***	-0.323***	-1.420***
	(0.180)	(0.0617)	(0.184)	(0.108)	(0.109)	(0.180)	(0.191)	(0.0779)	(0.0286)	(0.0830)	(5.335)	(0.0809)	(0.152)
Male*No Father	0.178**	0.0807***	-0.00548	-0.150**	-0.105	0.00881	-0.0491	0.0269	-0.0300	-0.0477	-2.020	-0.0993***	-0.305***
	(0.0798)	(0.0280)	(0.0667)	(0.0659)	(0.0749)	(0.0791)	(0.0726)	(0.0286)	(0.0197)	(0.0402)	(2.250)	(0.0315)	(0.0756)
Male*Other Father	0.0433	0.0747**	-0.210*	-0.125	-0.117	-0.0137	0.0554	0.0172	0.0117	0.0869**	1.430	-0.0433	-0.177*
	(0.0908)	(0.0366)	(0.114)	(0.102)	(0.105)	(0.0974)	(0.118)	(0.0395)	(0.0270)	(0.0396)	(5.775)	(0.0432)	(0.0957)
No Father	0.176***	0.0989***	0.186***	2.74e-05	-0.155***	-0.236***	-0.183***	-0.134***	-0.0319**	-0.0185	-5.005***	0.00176	0.268***
	(0.0575)	(0.0194)	(0.0493)	(0.0484)	(0.0535)	(0.0541)	(0.0520)	(0.0212)	(0.0131)	(0.0254)	(1.226)	(0.0229)	(0.0630)
Other Father	0.147**	0.0326*	0.277***	0.0337	-0.0423	-0.0671	-0.129	-0.114***	-0.0326*	-0.0618*	-3.401	0.00339	0.203**
	(0.0703)	(0.0191)	(0.0973)	(0.0623)	(0.0781)	(0.0654)	(0.0889)	(0.0270)	(0.0175)	(0.0343)	(2.492)	(0.0351)	(0.0800)
Constant	-1.041***	-0.468***	-1.608***	1.362***	-0.00826	3.425***	3.833***	-0.235	0.611***	0.336**	-46.72***	-1.091***	-0.809**
	(0.348)	(0.142)	(0.253)	(0.254)	(0.266)	(0.332)	(0.376)	(0.161)	(0.0793)	(0.138)	(10.98)	(0.161)	(0.319)
Observations	7,142	7,262	7,248	7,246	7,245	7,008	6,702	7,264	7,264	6,049	7,025	7,258	7,244
R-squared	0.050	0.124	0.063	0.097	0.150	0.093	0.040	0.194	0.079	0.051	0.067	0.081	0.125
Mean of dependent variable	0.000	0.206	-0.002	0.002	0.001	2.918	2.790	0.369	0.935	0.788	36.100	0.560	0.844
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "Educated Neighborhood" is the proportion of aged 25+ individuals with college degree or more at census tract level. "No Father" and "Other Father" refer to living arrangements at Wave I. "School problems" is a standardized index based on factor analysis of the variables in Table B.5. "Depression" is a standardized index based on factor analysis of the variables in Table B.6. Grades are student-reported and range from 1=D or lower to 4=A. College desires/expectations are standardized measures based on a 0-4 scale. "Currently Employed" is a dummy for whether is currently working at Wave IV. "Personal Income" is the total personal earnings before tax in 2006/2007/2008. "Ever Married" is a dummy for whether has been married up to Wave IV. "Children Ever Born" is the self-reported total number of child births up to Wave IV. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. Models (1)-(7) are weighted by Wave I weights and models (8)-(13) are weighted by Wave IV weights.

are more responsive to advantages/disadvantages in adolescence. However, as with father absence, the differential effects vanish in adulthood, except for fertility decisions.[19]

2.4.5 Additional Results

Add Health Black Sample

The African-American sample in Add Health is much smaller than the non-Hispanic white sample (about 2700 vs. 7200), but the higher prevalence of non-traditional families in this population makes a parallel analysis of key outcomes on this subsample potentially informative. On some dimensions, the results reported in Table B.10 contrast sharply with those from the majority sub-sample. Young Black men are less likely to graduate from high school or college than young Black women (and by larger margins than in the white sample) and no-father households are still associated with less education, more school problems, and a higher probability of school suspension. However, in important departures from the white sample results, there are no significant gender or family structure effects on college aspirations, and no family structure effects on the depression index. There is only one significant gender/family structure interaction, and it is a surprising one. The gender gap in school suspensions is smaller for adolescents in no-father families, rather than larger. In general, school discipline rates are much higher for Black students, male and female, and the behavioral determinants appear to be very different as well. The differences between the Black and white samples on this dimension may be reflective of racial differences in the institutions of school discipline.

[19]As we did with father absence, we can test for non-differential selection on observables of boys and girls into these alternative measures of early disadvantage. Controlling for father absence, we find little evidence of such selection–one of 16 interaction terms in models of observed maternal and child characteristics and sample attrition is significant at the 10 percent level.

NLSY97 White Non-Hispanic Sample

In order to examine the external validity of our results, we replicate the main analysis using another representative longitudinal dataset, NLSY97. The results for adult outcomes are reported in Table 2.7. Most results are consistent with those from the Add Health sample. Father absence has no differential effect on high school and college graduation, personal income and job satisfaction for NLSY97 men and women. Step-father households, in one departure from the Add Health results, have less negative impacts on college graduation for men than for women, though they do have a marginally significant negative effect on the job satisfaction of men compared to women. Non-traditional family structures tend to increase the likelihood of being currently employed for men relative to women and to decrease the likelihood of being ever married and fertility more for men, as in Add Health. In addition, there is some evidence that other father figures increase depression in adulthood less for males relative to females, which is also consistent with the Add Health results. Table B.11 reports estimates using alternative measures of educational attainment, and these results also support the robustness of our Add Health findings. In general, the results using NLSY97 sample support a conclusion that non-traditional family structures have few differential impacts on educational attainment across gender and any significant effects tend to be in favor of males rather than females. This implies that the apparent excess vulnerability of adolescent boys in school-based behaviors doesn't persist into adulthood.

2.5 Conclusion

Using data on young cohorts of men and women from the National Longitudinal Study of Adolescent to Adult Health, we investigate the association between economic disadvantage in adolescence and relative outcomes for men and women, both in school

Table 2.7: Educational Attainment and Other Adult Outcomes, NLSY97 Non-Hispanic White Sample

VARIABLES	(1) College Graduation	(2) High School Graduation	(3) Currently Employed	(4) Personal Income	(5) Job Satisfaction	(6) Ever Married	(7) Children Ever Born	(8) Depression Index
Male	-0.113***	-0.0151**	0.0839***	15.61***	-0.0499	-0.110***	-0.284***	-0.230***
	(0.0205)	(0.00721)	(0.0149)	(1.972)	(0.0485)	(0.0211)	(0.0529)	(0.0479)
Male*No Father	0.0142	0.00722	0.0920***	0.121	-0.113	-0.0420	-0.371***	-0.0474
	(0.0391)	(0.0236)	(0.0324)	(3.652)	(0.106)	(0.0449)	(0.121)	(0.106)
Male*Other Father	0.0758*	0.0372	-0.00616	0.407	-0.212*	-0.117**	-0.307**	-0.242*
	(0.0410)	(0.0308)	(0.0456)	(4.221)	(0.119)	(0.0507)	(0.139)	(0.124)
No Father	-0.205***	-0.0524***	-0.0777***	-9.143***	-0.0256	-0.0297	0.338***	0.143*
	(0.0276)	(0.0155)	(0.0269)	(2.148)	(0.0753)	(0.0297)	(0.0887)	(0.0769)
Other Father	-0.250***	-0.0797***	-0.0591	-9.286***	0.116	0.0345	0.248**	0.326***
	(0.0304)	(0.0224)	(0.0365)	(2.718)	(0.0874)	(0.0344)	(0.109)	(0.0910)
Constant	-0.155*	0.794***	0.689***	-9.435	4.004***	0.339***	1.324***	0.0852
	(0.0916)	(0.0531)	(0.0834)	(9.518)	(0.251)	(0.103)	(0.270)	(0.249)
Observations	2,998	2,998	2,989	2,533	2,363	2,991	2,994	2,902
R-squared	0.224	0.059	0.043	0.091	0.007	0.036	0.095	0.032
Mean of dependent variable	0.400	0.949	0.853	49.933	4.129	0.647	1.268	0.025
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "No Father" and "Other Father" refer to living arrangements at baseline survey (1997). "Currently Employed" is a dummy for whether is currently working at round 17 survey. "Personal Income" is the total personal earnings before tax in 2014. "Job Satisfaction" is on 5 point scale. "Ever Married" is a dummy for whether has been married up to round 17 survey. "Children Ever Born" is the self-reported total number of child births up to round 17 survey. "Depression Index" is a standardized index based on factor analysis of 5 variables. Mother's characteristics include education and mother's age when the first child in the household and the respondent was/were born. All models include birth cohort. All models are weighted by round 17 sample weights.

and later in life. Girls appear more resilient to father absence when the outcomes are adolescent school problems, suspensions, and educational aspirations, while boys appear more resilient to father absence when we examine depression. Though these school-age outcomes are themselves associated with poor educational and labor market outcomes in adulthood, these gender gaps related to father absence do not result in differential college graduation rates, income or other adult economic outcomes. The principal exceptions are marriage and reported fertility: non-traditional family structures are associated with lower relative probabilities of marriage and number of children for men, compared to women. The pattern of results is similar when boy/girl vulnerability to poor school quality or less educated neighborhoods, instead of father absence, is examined. Additional results using another representative longitudinal dataset, the National Longitudinal Survey of Youth 1997, show very similar patterns as well. Null results are an important component of our findings but, with few exceptions, we can rule out substantial levels of excess male vulnerability in adult economic outcomes.

These mixed results—gender-specific behavioral responses to family disadvantage among school children that do not result in gendered consequences for eventual educational attainment and other economic outcomes—suggest that previous findings of excess male vulnerability, while provocative and interesting, can be over-interpreted. Measures of problem behaviors in school seem to reflect gendered responses to disadvantage and they do not have clear implications for actual skill development in boys and girls or for eventual educational outcomes. Behavior in school is a consequence, not just of underlying skills and traits, but also of constraints and expectations that operate very differently for boys and girls due to gender norms in behavior on the part of parents, teachers, and the children themselves. Externalizing behavior that leads to problems in school is much more prevalent among boys, while internalizing behavior, which includes anxiety and depression, is a more common response to stress for girls, but is not included in the survey and administrative data used in prior studies. Most of the socio-behavioral outcomes examined in other studies, such as kindergarten readiness and school suspensions, are also related to externalizing behavior and so suggest greater male vulnerability to disadvantage. This analysis of Add Health data, though consistent with these earlier studies, finds no empirical support for the hypothesis that disadvantages in adolescence have contributed to the growing gender gap in college graduation, or to gender gaps in other adult economic outcomes.

Chapter 3

Short-run and Long-run Effects of Peers from Disrupted Families

3.1 Introduction

The family structures of Americans, as well as the family arrangements of American teenagers, have changed dramatically over the past 50 years. Although the majority of American children under 18 still live in families with two parents, this proportion decreased from 88% in 1960 to 69% in 2016.[1] Meanwhile, the proportion of children living in households with only a mother figure, which is the second most common family arrangement, tripled from 8% to 23%. Other non-traditional family structures, such as father-only households and no-parent households, also exhibit moderate increases over time.

Although the reasons for these dramatic changes in family structure are not clear, economists have agreed that a non-traditional family structure disadvantages children and is negatively associated with many important economic outcomes, including edu-

[1] https://www.census.gov/newsroom/press-releases/2016/cb16-192.html

cational attainment and future earnings. One important channel is through adolescent behavior–teenagers from disrupted families tend to exhibit more behavioral problems and mental health problems (Lei & Lundberg 2020). In fact, many of these effects may be interpreted as causal relationships.[2] Therefore, the rise of these disrupted families, defined as non-traditional family structures in which children are not living with both parents, has important implications for intergenerational mobility and inequality because of the intergenerational transmission of disadvantages (McLanahan et al. 2013).

More importantly, the disadvantages associated with non-traditional family structures may affect not only the teenager living in the household, but also her classmates or peers through classroom spillovers. A growing literature on peer effects in the classroom supports that possibility. For example, Carrell & Hoekstra (2010) show that having peers suffering from domestic violence decreases grades and increases misbehavior in the classroom. Carrell et al. (2018) further show that the effects are long lasting: having peers suffering from domestic violence decreases earnings in young adulthood. Bifulco et al. (2011) and Cools et al. (2019) show that having peers with highly-educated parents affects educational attainment. Billings & Hoekstra (2019) show that having peers with an arrested parent reduces achievement and increases antisocial behavior.

In addition, the gender composition of peers matters for cognitive outcomes (Lavy & Schlosser 2011), as well as college major choice, academic performance and labor market outcome (Anelli & Peri 2015). Moreover, Olivetti et al. (2020) show that the labor force participation of high school peers' mothers affects adult women's labor force participation. Therefore, it is reasonable to expect that the disadvantage of disrupted families also has spillover effects in the classroom. In that case, the mobility and inequality consequences

[2]Although there is a lack of fully convincing evidence in the literature, McLanahan et al. (2013) review dozens of studies using different strategies to identify causal impacts of father absence, including fixed effects models, natural experiments and instrumental variables, and find consistent evidence that father absence increases social-emotional problems, worsens mental health, decreases high school graduation and employment for children in the family.

of the rise of disrupted families could be even larger: teenagers suffer from disrupted families and also pass on some effects to their peers, who share similar backgrounds and characteristics.

Therefore, it is crucial to understand the effects of peers from disrupted families. Despite the growing literature on peer effects, to my knowledge, no previous study has investigated the peer effects of non-traditional family structures, especially in the long-run.[3] In this paper, I use rich longitudinal data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to examine the effects of peers from disrupted families on a broad set of adolescent outcomes, including behavior in school, mental health, grades and educational aspirations, and also extend analysis of the impacts to long-run outcomes, including educational attainment, employment, income, criminal behavior, marriage and fertility in young adulthood, for a recent cohort of young adults. The empirical strategy exploits idiosyncratic within-school cross-cohort variation in the composition of peers from different family structures, similar to previous studies (Bifulco et al. 2011, Olivetti et al. 2020, Cools et al. 2019).

Specifically, I focus on the peer group of all students within the same school, grade, gender, and race, which makes them more likely to become friends and have closer interactions compared to students in broader peer groups.[4] The extensive information on schoolmates and the longitudinal component of the Add Health data enable me to

[3]The focus of this paper, the effects of peers from disrupted families, is closely related to previous studies examining the effects of "disruptive peers" (Carrell & Hoekstra 2010, Carrell et al. 2018). Although both focus on the effects of family environment of peers, and some effects may go through similar channels, the concepts "peers from disrupted families" and "disruptive peers" are distinct concepts. The former is defined by whether living with both parents, while the latter is defined by exposure to domestic violence, according to Carrell & Hoekstra (2010) and Carrell et al. (2018). Kristoffersen et al. (2015) generalize the concept of "disruptive peers" and define peers to be "potentially disruptive" based on family backgrounds of children (with divorced parents, with parents convicted of crime, with a psychiatric diagnosis). Therefore, the concepts "peers from disrupted families" and "disruptive peers" are also closely related, as peers from disrupted families may be "potentially disruptive".

[4]Students are classified into one of four ethnicity groups: non-Hispanic white, African American, Hispanic, and other race/ethnicity. See Section 3.3 for more details on variable definitions.

examine the effects on both short-run and long-run outcomes.

I find that the effects of exposure to peers from disrupted families are pronounced among boys, but not for girls.[5] When boys are exposed to a one standard deviation higher proportion of peers from disrupted families, they exhibit 0.10 standard deviation more school problems in adolescence. The effects are in many domains, including self-reported trouble with students, trouble with teachers, trouble paying attention and trouble with homework. The primary channel of the effects is likely to be through peers' behavior. In young adulthood, these boys have 5.9 percentage points higher probability of ever having been arrested, 0.17 standard deviation more times being fired, and 6.2 percentage points higher probability of experiencing financial stress. The magnitude of these effects is sizable compared to other peer effects documented in the literature (Carrell & Hoekstra 2010, Olivetti et al. 2020, Cools et al. 2019). There is no evidence that cognitive skills, measured by academic performance in adolescence, are affected by peers from disrupted families. In addition, family outcomes such as marriage and fertility are unaffected by peers from disrupted families. These results together imply that boys are negatively affected by peers from disrupted families in terms of non-cognitive skills.

The main contribution of this paper is two-fold. First, this paper provides empirical evidence on a new type of long-lasting peer effects in the classroom and extends the literature on spillover effects of peers' family backgrounds and characteristics (Carrell & Hoekstra 2010, Carrell et al. 2018, Bifulco et al. 2011, Lavy & Schlosser 2011, Olivetti et al. 2020, Billings & Hoekstra 2019, Cools et al. 2019). Second, this paper focuses on a narrower peer group—students within the same school, grade, gender, and race—

[5]The findings are consistent with many previous studies documenting that peer effects may only be pronounced for one gender but not the other. For example, Carrell & Hoekstra (2010) find that the negative effects of disruptive peers are primarily driven by boys but not girls. Olivetti et al. (2020) find that the labor force participation of peers' mothers only affect girls but not boys in terms of long-run labor force participation. In addition, Cools et al. (2019) find that only girls but not boys are affected by "high-achieving peers".

and provides evidence that peers who share similar demographic characteristics may be more influential and are the main drivers of the peer effects, compared to peers defined by broader groups. The closest related previous study along this line is Billings et al. (2019), who examine the effects of exposure to neighborhood peers on crime and find trivial effects for same school-grade peers but pronounced effects for same school-grade-gender-race peers, illustrating that peer effects are likely to be stronger when investigating narrower peer groups who share similar demographic characteristics. In addition, the results of race-specific peer effects and peer effects heterogeneity in this paper also contribute to the literature on race-specific peer effects and race-specific social interactions (Hellerstein et al. 2011, Ananat et al. 2018, Davis et al. 2019, Fletcher et al. 2020) and highlight the importance of the racial dimension of peer effects.

The findings in this paper also have important implications for education and public policy. First, they suggest that interventions that can help alleviate the disadvantage of family disruption may have larger benefits than previously anticipated because of classroom spillovers. Second, they provide an additional reason to be concerned about the dramatic change in family structures in the United States. In other words, the intergenerational mobility and inequality consequences of the dramatic change in family structures may be larger than expected.

The remainder of the paper is organized as follows. Section 3.2 discusses the empirical strategy in this paper. Section 3.3 describes the data. Section 3.4 presents the main results of the analysis. Finally, Section 3.5 concludes.

3.2 Empirical Strategy

The main threat to estimating the causal effects of peer characteristics is that families or students may systematically sort across schools. Similar to previous studies (Bifulco et al. 2011, Olivetti et al. 2020), this paper exploits the within-school cross-cohort variation in peer composition to estimate the effects of peers from disrupted families. The specification can be written as:

$$Y_{igs} = \beta P_{igs} + \theta X_{igs} + \phi_s + \eta_g + \delta_s g + \varepsilon_{igs}$$

where Y_{igs} is the outcome of student i in grade g of school s. P_{igs} is the proportion of peers from disrupted families, which is the *leave-one-out* sample mean of the peer group. In contrast to previous studies using Add Health data, where peers are defined as within the same school-grade cell (Bifulco et al. 2011, Olivetti et al. 2020), or within the same school-grade-gender cell (Cools et al. 2019), the peer group of interest in this paper is defined as all students within the same school-grade-gender-race cell.[6] For example, for a white female student, all other white female students in the same grade of same school will be defined as same-gender same-race peers, the peer group of interest.

Compared to broader peer groups, students within the same school-grade-gender-race cell are presumably "closer" peers, and have higher probability of being close friends. Fruehwirth et al. (2019) examine the friend sorting patterns in the Add Health data and conclude that students are more likely to form friendships with other students of the same school, grade, race and gender. Therefore, these "closer" peers may have larger and more long-lasting effects that are more valuable to investigate, and peer group defined by this narrower definition is preferred to broader peer groups.

X_{igs} is a set of covariates that consists of three parts. First, I control for a set of family and individual characteristics, including age, race, standardized PVT score, parental education, family structure and maternal characteristics. Second, in order to address the

[6]Although under-explored by previous studies using Add Health data, defining students within the same school-grade-gender-race as the peer group of interest or focusing on peer effects by racial match is not uncommon in the literature (Hellerstein et al. 2011, Billings et al. 2019).

concern that other peer groups, i.e. students not of the same gender or the same race as the respondent, are also relevant and influential, I control for the proportion of peers from disrupted families for all other peers that are in the same school and grade but not of the same gender-race as the respondent. Third, in order to address the concern that the proportion of peers from disrupted families may be correlated with other influential peer characteristics, and the results may capture other aspects of peer effects pointed out by previous studies (Bifulco et al. 2011, Cools et al. 2019), I also control for other peer characteristics, including the gender composition of peers, the race/ethnicity composition of peers and education of peers' mothers.[7]

School fixed effects ϕ_s and grade fixed effects η_g are included in the specification, as well as a school-specific linear time-trend $\delta_s g$, in order to control for all the unobserved fixed differences across schools or across grades, and potential changes in school effects or school average characteristics over time. In addition, in order to address the concern that different race/ethnicity groups may have different sorting processes into schools (Fletcher et al. 2020), I also implement a more saturated and more conservative specification by allowing the school fixed effects and school time trends to vary by race. Note that there is a bias-precision trade-off between these two specifications: the more saturated specification with school-by-race fixed effects and school-by-race trends are less likely to suffer from bias due to potential race-specific sorting, but are also likely to get less precise estimates due to smaller residual variation for estimation.

Finally, standard errors are clustered at school level. In order to account for the type I error associated with multiple hypothesis testing, I also report the Romano-Wolf

[7] This group of variables is referred to as "other peer characteristics". Because the peer group of interest is the same-gender same-race peers within the same school and same grade, the proportion of peers' mothers having a college degree and the proportion of peers' mothers graduated from high school are also defined with respect to same-gender same-race peers within the same school-grade. Similarly, the gender composition is defined with-respect to same race peers within the same school-grade, and the race composition is defined with respect to same-gender peers within the same school-grade. Detailed variable definitions are listed in Table C.1.

adjusted p-values, following the Romano-Wolf multiple hypothesis correction procedure (Romano & Wolf 2016, Clarke et al. 2020).

The main identification assumption is a standard one in the literature (Bifulco et al. 2011, Olivetti et al. 2020, Cools et al. 2019): while students may sort across schools, they don't sort based on the cohort characteristics within the school. Therefore, the variation in peer compositions within the same school but across different cohorts can be viewed as quasi-random.

3.3 Data Description

The analysis is based on data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health has collected a rich array of longitudinal data on the social, economic, psychological and physical well-being of young men and women from adolescence through young adulthood.[8] The Add Health study began in 1994-95 with a nationally-representative school-based survey of more than 90,000 students in grades 7 through 12. The students were born between 1976 and 1984 and attended one of 132 schools in the sampling frame. In addition to oversamples of several ethnic groups and disabled students, the Add Health genetic sample includes sibling pairs living in the same household, including twins, half-siblings, and biologically unrelated siblings. About 20,000 respondents were followed in subsequent surveys, the last of which (Wave IV) was conducted in 2007-08 when the respondents were between 25 and 32 years of age.

The estimation strategy in this paper exploits the within-school cross-cohort variation

[8]This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

of peer composition from the population of students. All the information on school peers is obtained from the in-school survey, in order to recover information on all school peers, including those who were not interviewed in the in-home survey.[9]

The sample selection procedure is as follows: I start from the sample of over 18,000 individuals who completed Wave I in-home survey and have non-missing values of core variables, including sample weights, race/ethnicity, school ID and grade. Over 13,000 individuals are merged with peer composition variables constructed using in-school survey. I then drop individuals who have less than 8 same-gender same-race peers in the same school and grade, which is approximately the 10th percentile of the distribution, to avoid extreme values because of a too-small denominator. I also drop individuals who don't have any peers of opposite gender or of different race in the same school and same grade in the in-school survey, so that individuals in single-gender or single-race schools will be excluded from the analysis. There remains approximately 11,500 individuals. The demographics of the sample remains basically unchanged during the sample selection process, indicating that the analyzed sample is representative. Finally, I drop individuals who don't have a response for the Wave IV survey. The final sample for analysis contains 5,006 women and 4,170 men.[10]

3.3.1 Peer Compositions

The peer composition variables are generated using in-school survey data. I classify respondents into one of four race/ethnicity groups: non-Hispanic white, African American, Hispanic, and other race/ethnicity. The respondents report whether they live with

[9] The in-school survey was conducted on a single day between September 1994 and April 1995, and every student in attendance on the school's survey day was asked to complete the in-school questionnaire.

[10] The discrepancy in the male-female sample sizes is the result of consistently lower rates of both contact and response for male Add Health sample members. Attrition from the panel is higher for men and lower for white respondents. However, the attrition is independent of the proportion of same-gender same-race peers from disrupted families, and will not threaten the identification strategy.

biological mother, stepmother, foster mother, or adoptive mother, and the same for father figures.[11] An indicator of coming from a disrupted family is equal to 0 only if the respondent lives with both parents, and is equal to 1 otherwise.[12] Then based on the gender and race/ethnicity groups, I am able to construct the proportion of peers from disrupted families for same-gender same-race peers, as well as other peers that are not of the same gender-race.

As discussed in Section 3.2, I also construct other peer composition variables that could potentially affect the outcomes and can't be controlled by fixed effects and time trends. These variables include the proportion of female student among same-race peers, the proportion of African American student/Hispanic student/other race/ethnicity student among same-gender peers, the proportion of same-gender same-race peers with a high school graduate mother, and the proportion of same-gender same-race peers with a college graduate mother.

3.3.2 Adolescent Outcomes

The Wave I survey collected an array of child-reported variables, including health, personality, experiences in school, and relationships with parents, siblings, friends and others. I use self-reports of Math and English grades and school problems, including school suspensions, to generate academic and behavioral outcomes that are similar to those in previous studies and augment this with a standard depression scale and reports of educational aspiration. Detailed variable definitions are listed in Table C.1.

- *School Problems*: Students were asked about problems they experience in school,

[11] The in-school survey question only asks about whether the respondent is living with any of these mother/father figures, so I am not able to further distinguish whether the peers are living with biological parents or other parent figures.

[12] The results are similar when using an alternative measure of family disruption based on whether the respondent lives with a father figure.

including trouble getting along with teachers and other students, trouble getting homework done and trouble paying attention in class, how many times they have been absent without an excuse, and whether they have ever received an out-of-school suspension. Factor analysis was used to aggregate these measures into a standardized school problem index.

- *Depression*: Wave I respondents were asked how often during the past week they felt sad, lonely, depressed, blue, happy, or hopeful. These six items (plus 13 more) are the components of CES-D, a standard depression scale (Radloff 1977). Factor analysis indicated that a single factor is appropriate for these 19 items and was used to form a standardized depression index.

- *Grades and Aspirations*: Students reported their math and English grades in the most recent grading period. Educational aspirations in Wave I are based on student responses regarding how much they want to attend college, and how likely they think it is that they will attend college.

3.3.3 Adult Outcomes

The Wave IV survey collected an array of adult outcomes, including educational attainment, employment, income, family history and criminal behavior.

- *Educational Attainment*: Highest educational attainment is measured in the Wave IV survey of Add Health, collected when most respondents are between 25 and 32 years of age. Most, though not all, will have completed their final level of formal schooling at this point. I focus on the attainment of a 4-year college degree, but I also examine high school graduation as another educational attainment measure.

- *Employment and Income*: The respondents were asked about current employment status and employment history. I examine personal income, and define respondents as currently employed if they report working for pay at least 10 hours a week. I also consider two other aspects of employment histories: number of times fired and job satisfaction. A dummy variable for experiencing financial stress is equal to 1 if the respondents report that they have faced difficulties in paying bills in the past year. Detailed variable definitions are listed in Table C.1.

- *Marriage and Children Ever Born*: Marriage history and fertility are asked in the Wave IV survey of Add Health. A dummy variable of ever married is defined based on the reported marriage history. The respondents were also asked about number of times having been pregnant/having made a partner pregnant, and the live births resulting from these pregnancies.

- *Criminal Behavior*: One adult expression of the impulsivity that leads to adolescent school problems and disciplinary interventions is criminal behavior (Lundberg 2017*b*). I create a dummy variable for ever having been arrested on the basis of a self-reported arrest history.

3.3.4 Family and Individual Characteristics

The Wave I in-home survey collected an array of individual characteristics and family background variables that are included as controls in most specifications. Detailed variable definitions are listed in Table C.1.

- *Standardized PVT Score*: At the beginning of Wave I survey, respondents were asked to complete an abbreviated version of the Peabody Picture Vocabulary Test (PVT), which is a standard assessment of verbal ability used in the United States.[13]

[13] The PVT score is often considered as a proxy for ability and often included as control in previous studies using Add Health data (Olivetti et al. 2020, Cools et al. 2019).

The raw PVT score is again standardized in the sample to have a mean of 0 and standard deviation of 1. When the PVT score is missing, the score is set to be 0 (sample mean) and a dummy for missing PVT score is included.

- *Family Structure*: According to the family structure reported in the in-home survey, I generate dummy variables for whether the respondent is not living with any father (mother) figure at the baseline survey, and whether the respondent is living with step-father (mother) or other father (mother) figure. The omitted categories are that the respondent is living with biological father (mother).

- *Parental Education*: The educational attainment of the respondent's mother is divided into 4 categories: "less than high school", "high school degree", " some college", and "college degree". I omit the first category and generate dummy variables for the last three categories. When the educational attainment is missing, all three dummies are set to 0 and a dummy for missing mother's education is included. Father's education variables are constructed in the same way.

- *Other Maternal Characteristics*: I also include indicators for whether the biological mother is foreign-born and young (under age 22) at the birth. Whenever the value is missing, the dummy is set to be 0 and an additional dummy for missing value is included.

3.4 Results

There are two things needed for the identification strategy to be valid. First, there has to be sufficient variation in the peer composition measure after controlling for fixed effects and school-specific trends to obtain precise estimates. Second, to be viewed as quasi-random variation, peer composition should be uncorrelated with pre-determined student characteristics.

Table 3.1 presents evidence for the former point. Following Bifulco et al. (2011), I examine the extent of variation in peer composition that is left after removing race, cohort and school fixed effects, as well as school trends. As the table shows, for male respondents, on average 29% of the same-gender same-race peers in the same grade and school are from disrupted families, with a standard deviation of 14.2%. Although removing fixed effects and school trends reduces variation, the residual variation is 6.9%, which accounts for about 48.6% of the overall variation. In the more saturated specification that controls for grade fixed effects, school-by-race fixed effects and school-by-race trends, the residual variation is 5.4%, which accounts for 38.0% of the overall variation. The pattern is very similar for female respondents. Therefore, evidence suggests that there is sufficient residual variation.

Table 3.1: Raw and Residual Variations in Proportion of Peers from Disrupted Families

% Peers from disrupted families: same gender-race	Mean	Std. dev.	Min	Max	Obs
Male					
Raw cohort variable	0.289	0.142	0.000	1.000	4170
Residuals: net of race, grade and school fixed effects and school trends	0.000	0.069	-0.463	0.420	4170
Residuals: net of grade and school-by-race fixed effects and school-by-race trends	0.000	0.054	-0.359	0.566	4170
Female					
Raw cohort variable	0.289	0.152	0.000	0.909	5006
Residuals: net of race, grade and school fixed effects and school trends	0.000	0.072	-0.382	0.526	5006
Residuals: net of grade and school-by-race fixed effects and school-by-race trends	0.000	0.058	-0.400	0.457	5006

Notes: The table reports descriptive statistics for the proportion of same-gender same-race peers within the same school-grade from disrupted families.

Table 3.2 presents an array of balancing tests, which supports the latter point. Each column represents a regression of the proportion of same-gender same-race peers from disrupted families on a set of family and individual characteristics, controlling for fixed effects and school trends, and I perform a F-test for the joint hypothesis that the coefficients of all characteristics are equal to 0. Note that there is a mechanical negative correlation between peers' and own characteristics when using the leave-one-out construction (Guryan et al. 2009, Cools et al. 2019), so I also control for whether the respondent's own family is disrupted to correct for the mechanical correlation.[14]

Columns (1) and (2) present the results for males. Column (1) presents the specification controlling for race, school, grade fixed effects and school trends, and column (2) presents the more saturated specification controlling for school-by-race fixed effects, grade fixed effects and school-by-race trends.[15] There is no evidence that peer composition is correlated with family and individual characteristics, and the joint F-test is not rejected, in either specification. Columns (3) and (4) present the results for females. Similarly, the joint F-test is not rejected in either specification as well. Generally speaking, the pre-determined characteristics are relatively balanced across different peer compositions, and the identification assumption is likely to hold under both specifications.[16]

[14]Note that the respondent's own family disruption status is included to correct for the mechanical correlation and is not included in the joint F-test. Special thanks for an anonymous referee for suggestions on the analysis.

[15]The Stata command *reghdfe* drops singleton observations when performing estimation with high-dimension fixed effects, so the effective number of observation is smaller in the specification with school-by-race fixed effects, grade fixed effects and school-by-race trends. See Correia (2017) for more details.

[16]Attrition is also unlikely to invalidate the results because there is no evidence that attrition is systematically different across different peer compositions (not reported).

Table 3.2: Balancing Tests

	% Peers from disrupted families: same gender-race			
	Male subsample		Female subsample	
	(1)	(2)	(3)	(4)
Mother college	-0.0041	-0.0036	-0.0025	0.0018
	(0.0030)	(0.0025)	(0.0032)	(0.0028)
Father college	-0.0003	-0.0009	-0.0042	-0.0038
	(0.0038)	(0.0036)	(0.0038)	(0.0032)
Foreign mother	0.0114	0.0098	0.0041	0.0084*
	(0.0093)	(0.0081)	(0.0074)	(0.0049)
Young mother	0.0029	0.0044	-0.0048	-0.0055**
	(0.0040)	(0.0035)	(0.0032)	(0.0027)
Log family income	-0.0008	-0.0007	-0.0003	-0.0005
	(0.0010)	(0.0007)	(0.0009)	(0.0007)
Age at baseline	0.0005	0.0015	0.0016	0.0003
	(0.0022)	(0.0022)	(0.0022)	(0.0021)
ZPVT score	0.0004	0.0012	0.0016	0.0013
	(0.0016)	(0.0016)	(0.0016)	(0.0013)
Disrupted family	0.0010	-0.0024	-0.0057	-0.0086***
	(0.0034)	(0.0032)	(0.0039)	(0.0032)
Number of observations	4,170	4,127	5,004	4,959
F-test statistics	1.08	1.37	0.88	1.43
F-test p-value	0.38	0.22	0.53	0.20
Race FE	Yes	No	Yes	No
School FE	Yes	No	Yes	No
Grade FE	Yes	Yes	Yes	Yes
School time trend	Yes	No	Yes	No
School-by-race FE	No	Yes	No	Yes
School-by-race time trend	No	Yes	No	Yes

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The reported coefficient is the regression coefficient of % Peers from disrupted families: same gender-race on the seven family and individual characteristics (mother college, father college, foreign mother, young mother, log family income, age, ZPVT score), controlling for an indicator of own family disruption status. Race fixed effects, school fixed effects, grade fixed effects and school time trends are controlled in columns (1) and (3). Grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in columns (2) and (4). The F-test statistics and F-test p-value are test statistics of F-tests for the joint null hypothesis that the coefficients of the seven family and individual characteristics (own family disruption indicator not included) are all equal to zero. All models are weighted by Wave I weights.

In addition, I also perform a Monte Carlo simulation, following Lavy & Schlosser (2011) and Cools et al. (2019), to examine whether the variation in peers from disrupted families is similar to what would be expected by chance. For each student i in the sample, I generate the family disruption status of i using a binomial distribution function with p equal to the fraction of students of the same gender and same race as i in the same school who come from disrupted families. Then I compute the within-school standard deviation, using the residuals of the proportion of same-gender same-race peers from disrupted families on race, school, grade fixed effects and school trends. The process is repeated 500 times to obtain an empirical 90% confidence interval for the standard deviation of each school.[17]

The results show that the variation in the peer composition is indeed "as-good-as-random". For males, 96% of schools have a standard deviation falling within the empirical 90% confidence interval. For females, 88% of schools have a standard deviation falling within the empirical 90% confidence interval. In addition, I also perform similar analysis using the more saturated specification that controls for school-by-race fixed effects, grade fixed effects and school-by-race trends. The proportions of schools having a standard deviation falling within the empirical 90% confidence interval are 94% for males and 89% for females. Therefore, these simulation results further support the validity of the identification assumption.

3.4.1 Adolescent Outcomes

In Table 3.3, I focus on male subsample and school problem index as the outcome to illustrate how the estimates change with the inclusion of additional controls and the robustness of the results. In column (1), I control for family and individual characteristics, race, school, grade fixed effects and school trends, but do not control for the proportion of not same gender-race peers from disrupted families or other peer characteristics. The result shows that a higher proportion of same-gender same-race peers from disrupted

[17] The analysis is performed only for schools with at least 3 grades. For those with fewer grades, the variation in the main variable is absorbed by the school fixed effect and its time trend.

families significantly increases the school problem index for boys. In column (2), I include the proportion of not same gender-race peers from disrupted families. The effect of same-gender same-race peers from disrupted families is stable and pronounced, and other peers that are not of the same gender-race show little effect on school problem index, suggesting that same-gender same-race peers is indeed the most influential peer group. In addition, I include other peer characteristics in column (3). Again the effect of same-gender same-race peers from disrupted families is mostly unchanged, suggesting that the results are unlikely to be capturing other aspects of peer effects. Note that school problems decreases with the proportion of female peers, consistent with previous findings on gender peer effects (Lavy & Schlosser 2011). Other aspects of peer characteristics do not show strong effects on school problems of boys.

Finally, in column (4), I allow the school fixed effects and school trends to vary by race, in order to address the concerns that different race/ethnicity groups may have different sorting processes into schools. The effects are less precisely estimated under this more conservative and more saturated specification. Note that there is a bias-precision trade-off between the specifications in columns (3) and (4): the more saturated specification in column (4) leads to smaller residual variation (as shown in Table 3.1) and slightly smaller effective number of observations, and the estimates may become less precise. In order to minimize the chance of having biased results, I present the more conservative specification (same as column (4)) as the baseline specification throughout the paper. And note that when the analysis is done separately by race, these two specifications become the same.

Table 3.3: Effects on School Problem Index, Male Subsample

	(1)	(2)	(3)	(4)
Variables		School problem index		
% Peers from disrupted families: same gender-race	1.012***	1.021***	1.046***	0.731*
	(0.323)	(0.323)	(0.320)	(0.387)
% Peers from disrupted families: not same gender-race		0.262	0.199	0.284
		(0.617)	(0.611)	(0.670)
% Peers female			-1.145***	-1.471***
			(0.381)	(0.495)
% Peers African American			-0.165	-0.232
			(0.775)	(0.975)
% Peers Hispanic			-0.639	-0.123
			(0.923)	(0.936)
% Peers other			0.291	0.538
			(0.621)	(0.637)
% Peers with college-educated mother			0.544*	0.352
			(0.323)	(0.408)
% Peers with high school-educated mother			-0.414	-0.562
			(0.296)	(0.438)
Observations	4,160	4,160	4,160	4,118
R-squared	0.157	0.157	0.161	0.227
Family and individual characteristics	Yes	Yes	Yes	Yes
Race FE	Yes	Yes	Yes	No
School FE	Yes	Yes	Yes	No
Grade FE	Yes	Yes	Yes	Yes
School time trend	Yes	Yes	Yes	No
School-by-race FE	No	No	No	Yes
School-by-race time trend	No	No	No	Yes

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Family and individual characteristics" include age, ZPVT score, parental education, family structure variables and maternal characteristics variables. Race fixed effects, school fixed effects, grade fixed effects and school time trends are controlled in columns (1)-(3). Grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in column (4). All models are weighted by Wave I weights.

Table 3.4 reports the results for the components of the school problem index for the male subsample. Panel A presents the results for the baseline specification. There is an increase in many dimensions of school problems, including trouble with teachers, trouble paying attention, trouble with homework and trouble with students, though most of the estimates are imprecisely estimated and have marginal statistical significance. Most of these effects are more precisely estimated and are significant at 5% level when using the less conservative specification. There is some evidence that more not same gender-race peers from disrupted families may increase trouble with students, but also decrease the chance of absences.

In Panels B and C, I present results of the effects of peers from disrupted families based on alternative peer group definitions (same-gender and not same-gender peers in Panel B, same-race and not same-race peers in Panel C), in order to show clear comparisons and illustrate the necessity of focusing on narrower peer groups. Results show similar positive effects on school problem index for boys when focusing on same-gender peers or same-race peers, although less precisely estimated. This is because the effects of broader peer groups are combinations of the effects of narrower peer groups. For example, the effect of (the proportion of) same-gender peers is approximately equal to the weighted average of the effect of same-gender same-race peers and the effect of same-gender other-race peers, with the relative size of population serving as weights. Indeed, these results further illustrate that the effects on school problems are driven by same-gender same-race peers, and same-gender same-race peers are the most relevant peer group. When looking at the components of the school problem index, most effects are driven by same-gender same-race peers. The exception is that all peer groups seem to have an impact on trouble with students, suggesting that having more peers from disrupted families increases the chance of having trouble with other students, regardless of the group of these peers.

In addition, in Panel D, I relax the assumption that the effect is linear in the proportion of peers from disrupted families by binning the proportion of same-gender same-race peers from disrupted families into quartiles. The rest of the specification remains the same as the baseline specification. For school problems, especially trouble with teachers, trouble paying attention and trouble with students, there are large and significant increases in the third and fourth quartiles, relatively to the first quartile. For trouble with students, there is even an increase in the second quartile. These results suggest that trouble with students is monotonically increasing in the proportion of peers from disrupted families, while the effects on other school problems are more pronounced when the share of peers from disrupted families is relatively large. Therefore, there is some evidence that these effects could indeed be non-linear, with the largest effects when peers from disrupted families are highly concentrated.

Table 3.5 reports the estimation results for all adolescent outcomes for the male subsample. Panel A presents the results for the baseline specification. A higher proportion of same-gender same-race peers from disrupted families increases the school problem index for boys, but it doesn't show significant effects on other outcomes including depression, grades, and educational aspirations. The effect on school problem index is significant at 10% level and has a Romano-Wolf adjusted p-value of 0.14.[18] In contrast, when using the less conservative specification, the effect on school problem index is more precisely estimated and robust to Romano-Wolf multiple hypothesis correction (Romano-Wolf adjusted p-value of less than 0.01).[19] There is some evidence in Table 3.5 that peers from disrupted families among other peers not of the same gender-race may also have some impacts on adolescent outcomes, including exacerbated depression and increased English grades. These results further illustrate that peer effects may be complex and heterogeneous across groups. Panels B and C present the results using alternative peer group definitions. Similar to Table 3.4, these results further illustrate the necessity of focusing on narrower peer groups.

The magnitude of the effect of same-gender same-race peers from disrupted families on school problems is relatively large: increasing the proportion of same-gender same-

[18] The Romano-Wolf hypothesis correction procedure is a correction of multiple hypothesis testing that uses a step-down approach that orders the tests by the p-value and reduces the number of competing tests as one moves from the highest level of significant down, and uses a resampling approach to adjust the correction for correlation between the tests. Throughout the paper, the resampling is clustered at school level, the same level as standard error clusters, to account for correlations between students within the same school. Special thanks for an anonymous referee for suggestions on this approach.

[19] Note that the multiple hypothesis correction here is performed only within sample (for boys). When also adjusting for the fact that the tests are performed separately for boys and girls, and doing a simple Bonferroni style correction, the p-value should be doubled (now significance level at 2%).

Table 3.4: School Problem Index and Components, Male Subsample

Variables	(1) School problem index	(2) Absences	(3) Ever suspended from school	(4) Trouble with teachers	(5) Trouble paying attention	(6) Trouble with homework	(7) Trouble with students
Panel A: baseline specification							
% Peers from disrupted families: same gender-race	0.731*	-1.547	0.00278	0.416	0.656*	0.532	0.770*
	(0.387)	(1.880)	(0.172)	(0.304)	(0.384)	(0.426)	(0.407)
% Peers from disrupted families: not same gender-race	0.284	-6.397**	0.0447	-0.632	0.128	0.535	1.398**
	(0.670)	(3.087)	(0.298)	(0.803)	(0.641)	(0.793)	(0.683)
Panel B: peer groups by gender							
% Peers from disrupted families: same gender	0.996	-6.236*	0.00745	0.708	0.746	0.772	1.278***
	(0.601)	(3.367)	(0.228)	(0.545)	(0.676)	(0.720)	(0.484)
% Peers from disrupted families: not same gender	0.412	-2.912	0.0609	-0.496	0.371	0.489	1.193*
	(0.610)	(2.751)	(0.290)	(0.709)	(0.559)	(0.695)	(0.626)
Panel C: peer groups by race							
% Peers from disrupted families: same race	0.683	-3.516*	0.142	0.0127	0.668	0.339	1.341***
	(0.465)	(2.063)	(0.192)	(0.420)	(0.458)	(0.575)	(0.485)
% Peers from disrupted families: not same race	0.587	-5.980**	0.0305	0.557	0.121	0.715	0.837*
	(0.476)	(2.733)	(0.206)	(0.466)	(0.396)	(0.444)	(0.490)
Panel D: non-linear specification							
% Peers from disrupted families: same gender-race, quartile 2	0.0492	-0.125	0.0373	-0.00144	0.0729	-0.0902	0.172**
	(0.105)	(0.479)	(0.0441)	(0.0902)	(0.105)	(0.123)	(0.0865)
% Peers from disrupted families: same gender-race, quartile 3	0.254**	0.325	0.0447	0.102	0.293***	0.0380	0.293**
	(0.105)	(0.450)	(0.0493)	(0.0852)	(0.107)	(0.124)	(0.116)
% Peers from disrupted families: same gender-race, quartile 4	0.248*	-0.837	0.0381	0.275**	0.151	0.122	0.284**
	(0.132)	(0.572)	(0.0639)	(0.112)	(0.148)	(0.153)	(0.136)
% Peers from disrupted families: not same gender-race	0.159	-6.828**	-0.0397	-0.575	-0.106	0.752	1.046
	(0.696)	(3.166)	(0.318)	(0.849)	(0.662)	(0.761)	(0.726)
Observations	4,118	4,118	4,125	4,125	4,125	4,124	4,124
Mean of dependent variable	0.136	2.038	0.316	0.905	1.290	1.277	0.879
Family and individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "Family and individual characteristics" include age, ZPVT score, parental education, family structure variables and maternal characteristics variables. "Other peer characteristics" include % Peers female, % Peers African American, % Peers Hispanic, % Peers other, % Peers with college-educated mother, and % Peers with high school-educated mother. Grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in all models. All models are weighted by Wave I weights.

Table 3.5: Adolescent Outcomes, Male Subsample

Variables	(1) School problem index	(2) Depression index	(3) English grade	(4) Math grade	(5) Wants to attend college	(6) Expects to attend college
Panel A: baseline specification						
% Peers from disrupted families: same gender-race	0.731*	0.362	-0.148	0.421	-0.189	-0.0604
	(0.387)	(0.359)	(0.452)	(0.443)	(0.416)	(0.411)
% Peers from disrupted families: not same gender-race	0.284	1.205*	1.372*	0.0162	0.0166	-0.147
	(0.670)	(0.629)	(0.824)	(0.650)	(0.878)	(0.710)
R-W adjusted p-value	0.1287	0.5743	0.8515	0.5743	0.8515	0.8515
Panel B: peer groups by gender						
% Peers from disrupted families: same gender	0.996	0.845**	0.238	0.275	0.668	0.572
	(0.601)	(0.419)	(0.837)	(0.573)	(0.718)	(0.529)
% Peers from disrupted families: not same gender	0.412	0.902*	0.476	-0.0270	-0.763	-0.854
	(0.610)	(0.542)	(0.733)	(0.521)	(0.796)	(0.614)
Panel C: peer groups by race						
% Peers from disrupted families: same race	0.683	0.482	0.0140	0.783	-0.196	-0.438
	(0.465)	(0.429)	(0.667)	(0.534)	(0.589)	(0.583)
% Peers from disrupted families: not same race	0.587	0.929**	0.855*	-0.0898	0.792	0.680
	(0.476)	(0.435)	(0.497)	(0.416)	(0.581)	(0.469)
Observations	4,118	4,114	4,034	3,883	4,116	4,114
Mean of dependent variable	0.136	-0.168	2.677	2.636	-0.086	-0.114
Family and individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Other peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. “Family and individual characteristics” include age, ZPVT score, parental education, family structure variables and maternal characteristics variables. “Other peer characteristics” include % Peers female, % Peers African American, % Peers Hispanic, % Peers other, % Peers with college-educated mother, and % Peers with high school-educated mother. Grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in all models. All models are weighted by Wave I weights.

race peers from disrupted families by one standard deviation increases the school problem index by 0.10, or equivalently 0.10 standard deviation. In contrast, living in a father-absent household is only associated with a 0.19 increase in the school problem index, and living in a mother absent household has even smaller and insignificant effects. In other words, a one standard deviation increase in father absence leads to a 0.08 standard deviation increase in school problem index. Though not directly comparable, these results suggest that family backgrounds of peers matter non-trivially compared to the family background of the boy himself, in terms of school problems. Disrupted families could increase the externalizing behavior of boys, in the form of school problems, and the results suggest that boys may suffer even more from the classroom environment than his own disrupted family.[20]

The fact that the effects of peers' family structure are roughly of the same magnitude as the effects of one's own family structure also sheds light on the potential channel of the effects. Although not directly testable, the most likely channel of the effects of peers' family structure is through peers' behavior, which is highly correlated with peers' family structure.[21] Exposure to more peers from disrupted families indicates exposure to more behavioral problems of peers in school, which has negative impacts on non-cognitive skills for boys. In contrast, channels other than peers' behavior are unlikely to fully explain the effects of peers' family structure, especially when the magnitude is as large as the effects of one's own family structure.

Table C.2 reports the results for the female subsample. The proportion of same-

[20]When looking at components of the school problem index, again most of the effects are non-trivial compared to the effects of own family structure such as father absence. This is intuitive because school problem is a result of interactions between both students themselves and their peers. For measures such as "trouble with students", non-traditional family structure shows little effects, while peer composition shows large effects, implying that this form of school problem is mostly driven by the behavior of peers.

[21]It is not possible to directly identify the effects of peers' behavior on one's own behavior, which is referred to as "endogenous peer effects" (Bifulco et al. 2011). The lack of identification for "endogenous peer effects", or known as "the reflection problem", is first pointed out by Manski (1993). See Bifulco et al. (2011) for more discussions on "contextual peer effects" and "endogenous peer effects".

gender same-race peers from disrupted families shows no effects on any of the adolescent outcomes. The point estimates are mostly statistically insignificant, and some estimates show reduction in school problems and depression, as well as increase in educational aspirations. Generally speaking, girls are mostly unaffected by same-gender same-race peers from disrupted families. And if there are any effects, they are unlikely to be negative for girls.[22]

3.4.2 Adult Outcomes

Table 3.6 presents the estimation results for adult outcomes for the male subsample. Panel A presents the results for the baseline specification. Columns (1)-(2) report the results for educational outcomes. There is no evidence that educational attainment, measured by college graduation rate and high school graduation rate, is affected by same-gender same-race peers from disrupted families.[23] Columns (3)-(10) present the estimation results for other adult outcomes. A higher proportion of same-gender same-race peers from disrupted families doesn't affect income or family outcomes. However, peers from disrupted families increase the number of times that the respondent has been fired, the probability of having ever been arrested, and the probability of experiencing financial stress in the past year. When adjusted for multiple hypothesis testing, the Romano-Wolf adjusted p-values are 0.01 for the number of times being fired, 0.03 for the probability of experiencing financial stress, and 0.06 for the probability of having ever

[22] As there is no evidence of girls being affected by peers from disrupted families in schools, only results for the male subsample are reported throughout the rest of the paper. In fact, there is also no evidence that girls are affected by peers from disrupted families in the long-run.

[23] There is suggestive evidence that human capital accumulation for boys may still be impeded by peers from disrupted families on the "intensive margin", through the quality of education they receive in post-secondary institutions (measured by the academic selectivity of the post-secondary institution that students attend, which is based on the national ranking in terms of the median SAT score of entering students). However, the sample with data on post-secondary institution quality has a too small sample size to support the saturated specification with school-by race fixed effects and school-by-race trends, so the results are not reported.

been arrested. These results indicate that the composite null hypothesis that same-gender same-race peers from disrupted families have no long-run effects for boys is rejected. There is no evidence that not same gender-race peers from disrupted families have any impacts on the long-run outcomes for boys.

Therefore, these results suggest that peers from disrupted families in school can not only increase the school problems that boys experience in school, but also have long-lasting effects on them. Boys exposed to higher proportion of peers from disrupted families tend to have less stable jobs, have a higher chance of being arrested and experiencing financial stress. These results imply that boys may be negatively affected by peers from disrupted families in terms of non-cognitive skills. Again these effects seem to be non-trivial compared to the effects of own non-traditional family structure, further showing the importance of classroom environments and spillover effects of disrupted families within classrooms.

The magnitude of these effects is sizable compared to other peer effects documented in the literature. Cools et al. (2019) find that increasing the proportion of male peers with college-educated parents by one standard deviation decreases the probability of obtaining a bachelor's degree for girls by 2.2-4.5 percentage points (0.05-0.09 standard deviation). Carrell & Hoekstra (2010) find that increasing the proportion of peers suffering from domestic violence by one standard deviation decreases the academic grades by 0.69 percentage points, or equivalently 0.025 standard deviation. Olivetti et al. (2020) find that increasing the proportion of peers with working mothers increases the labor force participation of girls by 5.6-7 percentage points, or equivalently 0.13-0.16 standard deviation. In contrast, I find that increasing the proportion of peers from disrupted families by one standard deviation will increase the school problem index by 0.10 (0.10 standard deviation), increase the probability of having ever been arrested by 5.9 percentage points (0.12 standard deviation), increase the probability of experiencing financial

Table 3.6: Adult Outcomes, Male Subsample

Variables	(1) College graduation	(2) High school graduation	(3) Personal income	(4) Financial stress	(5) Currently employed	(6) Number of times fired	(7) Job satisfaction	(8) Ever arrested	(9) Ever married	(10) Children ever born
Panel A: baseline specification										
% Peers from disrupted families: same gender-race	0.0133	-0.127	14.58	0.434**	0.200	2.202***	-0.0557	0.413**	0.132	-0.0538
	(0.171)	(0.116)	(28.07)	(0.180)	(0.142)	(0.664)	(0.307)	(0.202)	(0.150)	(0.398)
% Peers from disrupted families: not same gender-race	0.523	0.175	27.25	0.155	0.317	0.320	0.229	-0.523	-0.396	0.0442
	(0.340)	(0.158)	(35.07)	(0.342)	(0.283)	(1.070)	(0.641)	(0.369)	(0.255)	(0.653)
R-W adjusted p-value	0.9802	0.7525	0.9505	0.0297	0.4851	0.0099	0.9802	0.0594	0.7822	0.9802
Panel B: peer groups by gender										
% Peers from disrupted families: same gender	0.0195	0.0618	17.14	0.365	0.0694	1.666**	0.156	-0.230	0.201	0.114
	(0.290)	(0.148)	(29.68)	(0.234)	(0.258)	(0.758)	(0.512)	(0.276)	(0.177)	(0.545)
% Peers from disrupted families: not same gender	0.450	0.0102	36.65	0.193	0.567*	1.541	0.103	0.0964	-0.417*	0.128
	(0.288)	(0.146)	(33.29)	(0.297)	(0.286)	(1.606)	(0.465)	(0.297)	(0.250)	(0.547)
Panel C: peer groups by race										
% Peers from disrupted families: same race	0.115	0.0548	33.06	0.479**	0.437**	2.200**	-0.201	0.336	-0.0282	0.0734
	(0.220)	(0.151)	(37.89)	(0.199)	(0.190)	(0.878)	(0.459)	(0.255)	(0.196)	(0.445)
% Peers from disrupted families: not same race	0.177	0.124	-7.426	0.196	-0.163	-0.153	0.277	-0.309	0.0209	0.0284
	(0.158)	(0.110)	(24.33)	(0.192)	(0.166)	(0.624)	(0.464)	(0.224)	(0.209)	(0.433)
Observations	4,127	4,127	3,948	4,127	3,320	4,009	4,072	4,114	4,122	4,109
Mean of dependent variable	0.318	0.924	44.150	0.206	0.857	0.633	3.865	0.390	0.460	0.703
Family and individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Family and individual characteristics" include age, ZPVT score, parental education, family structure variables and maternal characteristics variables. "Other peer characteristics" include % Peers female, % Peers African American, % Peers Hispanic, % Peers other, % Peers with college-educated mother, and % Peers with high school-educated mother. Grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in all models. All models are weighted by Wave IV weights.

stress by 6.2 percentage points (0.15 standard deviation), and increase the number of times being fired by 0.31 (0.17 standard deviation).[24] These estimates are pronounced and of similar magnitude of the estimates in the literature.

Panels B and C present the results using alternative peer group definitions. Again, most long-run effects are driven by same-gender same-race peers. Focusing on broader peer groups defined by gender, as shown in Panel B, may miss important dimensions of peer effects on the arrest probability and the probability of suffering from financial stress. These results further indicate that peer effects may be heterogeneous across groups, and it is necessary to focus on narrower peer groups. In addition, same-gender same-race peers are the most relevant peer group that have long-lasting effects on boys.

3.4.3 Heterogeneous Effects

In this section, I present some additional evidence on heterogeneous effects of exposure to peers from disrupted families. In Table 3.7, I split the male sample into subsamples based on family and individual characteristics. I only report the coefficient of the proportion of same-gender same-race peers from disrupted families, while the specifications are the same as the baseline. In panel A, I split the sample based on race/ethnicity. The results suggest that the increased school problems, financial stress and arrest probability is mainly driven by white males, while minority males are not affected, and even exhibit higher labor force participation rate as young adults. For the number of times fired, both groups show large and positive effects. These results suggest that white boys are more negatively affected by peers from disrupted families, while minority boys are less vulner-

[24] When the increase in the proportion of peers from disrupted families is measured by an one residual standard deviation (net of grade and school-by-race fixed effects and school-by-race trends), these estimates translate into 0.04 units (0.04 standard deviation) increase in the school problem index, 2.2 percentage points (0.05 standard deviation) increase in the arrest probability, 2.3 percentage points (0.06 standard deviation) increase in the financial stress probability, and 0.12 units (0.07 standard deviation) increase in the number of times being fired.

able. These results are consistent with Cross (2020), who shows that African American youths are less vulnerable to family disruption compared to white youths, in terms of educational success.

In panel B, I split the sample by the family structure of the boy. If the boy lives in a household with both parents, he is classified as coming from "non-disrupted family".[25] Otherwise he is classified as coming from "disrupted family". The school problem effects are mainly driven by boys from non-disrupted families, probably because they are more vulnerable to peers from disrupted families. While for boys already living in disrupted families, they may be more used to this type of disadvantage and less vulnerable to peers from disrupted families. However, boys living in disrupted families exhibits larger effects on the probability of experiencing financial stress.

One may then want to see how race/ethnicity and own family structure interact in treatment effects. In panel C, I combine these two criteria and split the sample into four subsamples.[26] Though the long-run outcomes are less precisely estimated, the results are consistent with panel A and panel B, that white boys and boys not from disrupted families are more vulnerable groups, especially for white boys that are not coming from disrupted families, the interaction of these two criteria. Note that there is some evidence that minority boys from non-disrupted families may show beneficiary effects of same-gender same-race peers from disrupted families in terms of college graduation and current employment. Possible explanations for the beneficiary effects include better self-cognition and self-perception of the relative position among close peers, and differences in teachers' responses. These results further illustrate the necessity of focusing on narrower peer groups, as the peer effects may be fundamentally different across subgroups.

[25] In order to be consistent with the family disruption measure constructed for peers, I combine "living with biological mother/father" with "living with other mother/father", and the boy is classified as coming from "non-disrupted family" if he lives with both parent figures of any kinds.

[26] The sample sizes are much smaller in each subsample, so the point estimates are less precise.

Table 3.7: Heterogeneous Effects by Individual Characteristics, Male Subsample

Variables	(1) School problem index	(2) Trouble with teachers	(3) Trouble paying attention	(4) Trouble with homework	(5) Trouble with students	(6) College graduation	(7) High school graduation	(8) Personal income	(9) Financial stress	(10) Currently employed	(11) Number of times fired	(12) Job satisfaction	(13) Ever arrested
Panel A													
Subsample:	1.144**	0.569	1.213**	0.825	0.945*	-0.196	-0.127	27.21	0.704***	-0.121	1.435***	-0.417	0.597*
white	(0.522)	(0.382)	(0.500)	(0.632)	(0.490)	(0.254)	(0.152)	(42.33)	(0.260)	(0.175)	(0.484)	(0.385)	(0.305)
Subsample:	-0.587	-0.192	-0.661	-0.335	-0.0523	0.248	-0.189	0.624	-0.0598	0.702***	3.259*	0.353	0.278
not white	(0.554)	(0.574)	(0.664)	(0.507)	(0.845)	(0.201)	(0.219)	(19.30)	(0.232)	(0.222)	(1.870)	(0.532)	(0.269)
Panel B													
Subsample:	0.906*	0.224	0.724	0.595	1.057**	-0.0992	-0.0181	20.06	0.188	0.429**	1.610***	-0.413	0.430
non-disrupted family	(0.535)	(0.368)	(0.522)	(0.639)	(0.495)	(0.260)	(0.0760)	(54.80)	(0.268)	(0.182)	(0.575)	(0.407)	(0.275)
Subsample:	0.721	0.844	0.823	1.003	0.688	0.162	-0.134	50.63	0.934**	0.219	1.240	0.147	0.536
disrupted family	(0.702)	(0.861)	(0.740)	(0.763)	(0.697)	(0.250)	(0.297)	(33.18)	(0.357)	(0.348)	(0.966)	(0.618)	(0.362)
Panel C													
Subsample:	1.423**	0.446	1.561***	0.903	1.009**	-0.297	-0.0491	17.87	0.336	0.162	1.593**	-0.424	0.581
white+non-disrupted family	(0.610)	(0.430)	(0.557)	(0.763)	(0.506)	(0.334)	(0.105)	(65.04)	(0.335)	(0.213)	(0.698)	(0.481)	(0.357)
Subsample:	0.694	0.676	0.316	1.645	1.117	0.0281	-0.326	92.30	1.568***	-0.0980	1.169	-0.500	0.548
white+disrupted family	(1.386)	(1.332)	(1.217)	(1.402)	(1.215)	(0.363)	(0.464)	(75.16)	(0.507)	(0.315)	(1.203)	(0.887)	(0.613)
Subsample:	-1.482	-0.845	-1.871	-1.142	0.700	0.507*	0.132	9.999	-0.377	1.484***	0.523	0.619	0.348
not white+non-disrupted family	(0.906)	(0.764)	(1.163)	(0.748)	(1.254)	(0.291)	(0.146)	(31.79)	(0.250)	(0.280)	(0.725)	(0.687)	(0.353)
Subsample:	0.898	0.417	0.795	1.569	0.712	0.346	0.00801	24.98	0.0215	0.427	1.810	0.396	0.688*
not white+disrupted family	(1.003)	(1.265)	(1.045)	(1.159)	(1.285)	(0.355)	(0.538)	(19.71)	(0.619)	(0.588)	(1.902)	(1.282)	(0.386)

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Columns (1)-(5) are weighted by Wave I weights and columns (6)-(13) are weighted by Wave IV weights. The reported coefficient is the regression coefficient of each outcome on % Peers from disrupted families: same gender-race. % Peers from disrupted families: not same gender-race, family and individual characteristics, other peer characteristics, grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in all columns.

To conclude, these results imply that effects of peers from disrupted families are heterogeneous across different types of students. White boys and boys not from disrupted families are more vulnerable to their peers.[27]

3.5 Conclusion

In this paper, I use data on young cohorts of men and women from the National Longitudinal Study of Adolescent to Adult Health, to investigate the short-run and long-run effects of being exposed to peers from disrupted families. I find that the effects of peers from disrupted families are trivial for girls but pronounced for boys. When boys are exposed to more peers from disrupted families, they exhibit more misbehavior in school, in terms of trouble with students, trouble with teachers, trouble paying attention and trouble with homework. More importantly, these effects persist for years afterwards and into young adulthood. These boys are more likely to have ever been arrested, are fired from jobs more frequently, and have a higher probability of experiencing financial stress. These effects are especially pronounced for white boys and boys not from disrupted families. These results imply that boys are negatively affected by peers from disrupted families, in terms of non-cognitive skills, through classroom and school interactions. In contrast, I do not find any evidence that cognitive skills, measured by academic performance, are affected through classroom spillovers. Family outcomes, such as marriage and fertility, are unaffected as well.

These results further illustrate the importance of peer attributes in determining both short-run and long-run outcomes for teenagers. They suggest that overcoming the disadvantage of family disruption not only improves the outcomes for children within the

[27] In another analysis splitting the sample based on school characteristics, the results show that schools with over 50% of white students and small schools exhibit stronger peer effects within classrooms (not reported).

family, but also has important positive spillovers. These estimates are lower bounds for the total spillover effects, as children may interact not only within schools, but also in neighborhoods or through other social networks, where similar effects may also exist. As children are likely to share similar backgrounds in these social networks, overcoming the disadvantage of family disruption could help improve intergenerational mobility and reduce inequality, at a even larger magnitude than anticipated. On the other side, we should be more concerned about the dramatic increase in family disruption in the United States, because of the additional intergenerational mobility and inequality consequences through classroom spillovers.

Appendix A

Appendix for "Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes"

A.1 Theoretical Results on Testing Rational Selection Hypothesis

Let $Z_i = \mathbf{1}(S_i < c)$ denote the treatment assignment for individual i with running variable S_i, which is equal to 1 if the running variable is below the cutoff c. Let the treatment take-up status $D_i = Z_i \times D_i(1) + (1 - Z_i) \times D_i(0)$ denote whether individual i chooses to retake. Specifically, $D_i(1)$ and $D_i(0)$ denote the potential treatment value when the individual is assigned to be below and above the cutoff, respectively. The monotonicity condition (Imbens & Angrist 1994, Cattaneo, Idrobo & Titiunik 2018) requires that $D_i(1) \geq D_i(0)$, which rules out the possibility that the individual only chooses to retake when scoring above the cutoff ($D_i(1) = 0$ and $D_i(0) = 1$). Then the individuals can be classified into three types: "always-retakers" who always choose to retake regardless of being above or below the cutoff ($D_i(1) = D_i(0) = 1$), "compliers" who only choose to retake if being below the cutoff ($D_i(1) = 1$ and $D_i(0) = 0$), and "never-retakers" who never choose to retake ($D_i(1) = D_i(0) = 0$). Note that the first-stage RD estimand $\lim_{s\uparrow c} E(D_i|S_i = c) - \lim_{s\downarrow c} E(D_i|S_i = c)$, which is estimated by the coefficient β in Equation 1.3.1, identifies the proportion of "compliers" at the cutoff under continuity assumptions (Cattaneo, Idrobo & Titiunik 2018):

$$\lim_{s\uparrow c} E(D_i|S_i = c) - \lim_{s\downarrow c} E(D_i|S_i = c) = E(D_i(1) - D_i(0)|S_i = c).$$

Let the outcome $Y_i = D_i \times Y_i(1) + (1 - D_i) \times Y_i(0)$ denote the initial and final outcomes such as standardized test score. Note that the 1 and 0 in the function $Y_i(.)$ refer to the treatment take-up status $D_i = 1$ and $D_i = 0$, not the treatment assignment status Z_i. In addition, for initial outcomes that are realized before the cutoff is determined, such as standardized test score, $Y_i^I(1) = Y_i^I(0)$ holds by construction. Then, the 2SLS RD estimand identifies the local average treatment effect (LATE) of retake on the outcome for the "compliers" at the cutoff under continuity assumptions (Cattaneo, Idrobo & Titiunik 2018):

$$\frac{\lim_{s\uparrow c} E(Y_i|S_i = c) - \lim_{s\downarrow c} E(Y_i|S_i = c)}{\lim_{s\uparrow c} E(D_i|S_i = c) - \lim_{s\downarrow c} E(D_i|S_i = c)} = E(Y_i(1) - Y_i(0)|S_i = c, D_i(1) = 1, D_i(0) = 0).$$

Note that $Y_i^F(0) - Y_i^I(0) = 0$ holds by construction, which means that the final outcomes and the initial outcomes would be the same if not retaking the exam. Therefore, the coefficient β_{IV} in Equation 1.3.4 estimates the LATE of retake on the improvement of the outcome for the "compliers" at the cutoff, $E(Y_i^F(1) - Y_i^I(1)|S_i = c, D_i(1) = 1, D_i(0) = 0)$.

One important feature of our setting is that we can also estimate the LATE of retake on the improvement of the outcome for the "always-retakers" as well. Specifically, we have

$$E(Y_i^F - Y_i^I|S_i = c, D_i = 1) = E(Y_i^F(1) - Y_i^I(1)|S_i = c, D_i(1) = D_i(0) = 1).$$

Therefore, we can estimate the mean of the improvement of the outcome for individuals who choose to retake and scoring exactly at the cutoff to recover the LATE for the "always-retaker" at the cutoff. The intuition is that the individuals who choose to retake when scoring above the cutoff must be "always-retakers" by construction. If the rational selection hypothesis is true, then the LATE for the "always-retakers" at the cutoff should be higher than the LATE for the "compliers" at the cutoff, because the former group has strong motivation to retake regardless of the cutoff, and should be the group with the strongest incentive to retake.

To formalize this intuition, consider the following simple model where individual i chooses to retake if

$$R_i(S_i) - \theta_i > 0,$$

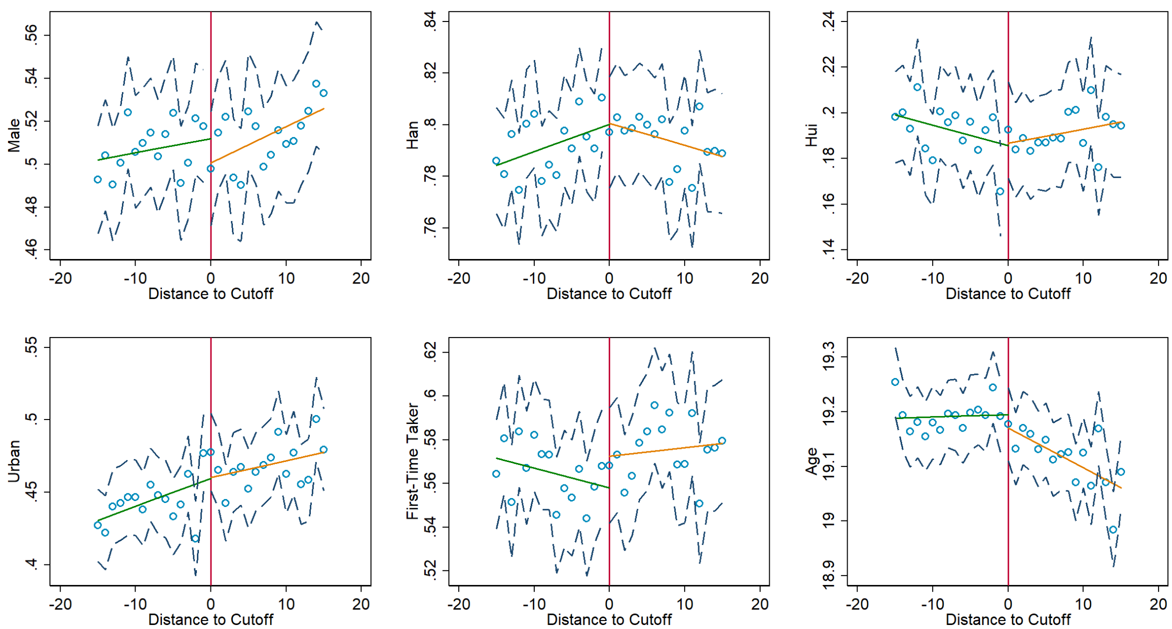
that the pecuniary return to retake R_i for individual i with test score is higher than the pecuniary cost of retake θ_i. Let $R_i(S_i) = r_i + \mathbf{1}(S_i < c) \times \mu_i$, that the return to retake is equal to a random variable r_i drawn from the distribution $F(.)$ if the test score is announced to be above or equal to the cutoff, and is equal to r_i plus an additional pecuniary benefit term $\mu_i > 0$ if the test score is announced to be below the cutoff. The term μ_i reflects the potential benefit of the possibility to move up in the admission tiers in the next take as the test score is very close to but below the cutoff in this year. In fact, the term μ_i is the source of the discontinuity in retake probability at the cutoff. For simplicity, let θ_i and μ_i to be constant.

Under this simple model with rational selection, individuals are divided into different types based on their return. For individuals with test score $S_i = c$, they are "always-retakers" if $r_i > \theta$, as they choose to retake regardless of whether the cutoff is below their test scores or not. Individuals with $\theta - \mu < r_i \leq \theta$ are "compliers", as their returns are not sufficiently high for them to retake if their scores are already above the cutoff, but would be sufficiently high for them to retake if their scores are below the cutoff. Individuals with $r_i \leq \theta - \mu$ are "never-retakers" that have low returns even if their scores are below the cutoff. Therefore, at any score S_i, the expected return for "always-retakers" should always be higher than the expected return for "compliers":

$$E(r_i|r_i > \theta) > \theta > E(r_i|\theta - \mu < r_i \leq \theta).$$

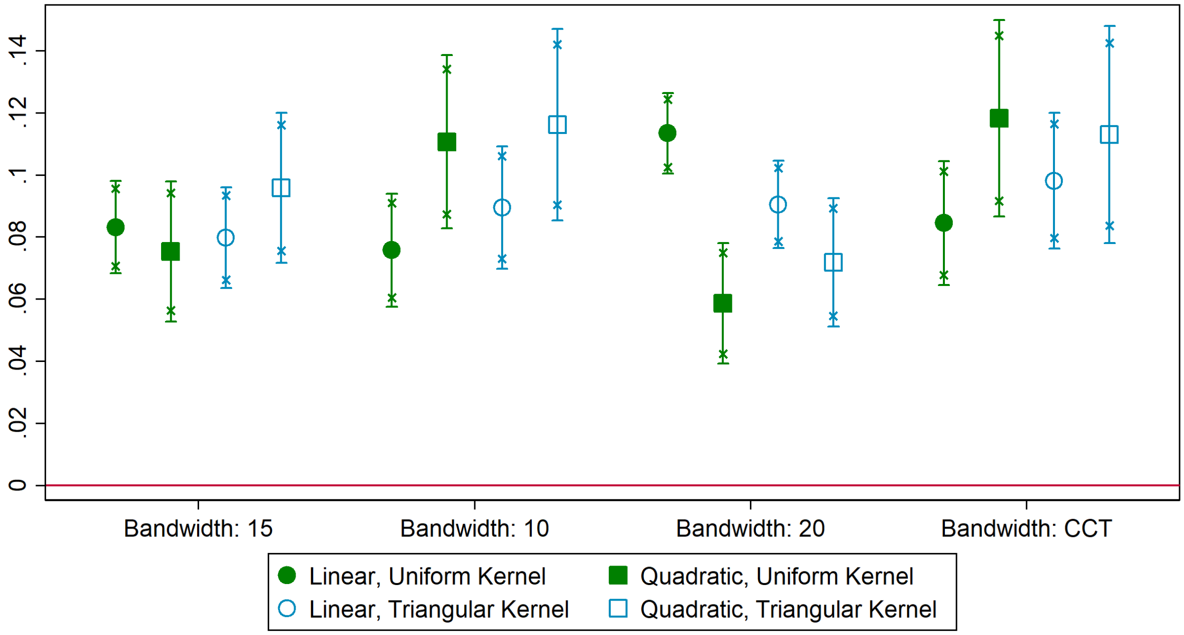
A.2 Appendix Figures and Tables

Figure A.1: Covariates Balancedness



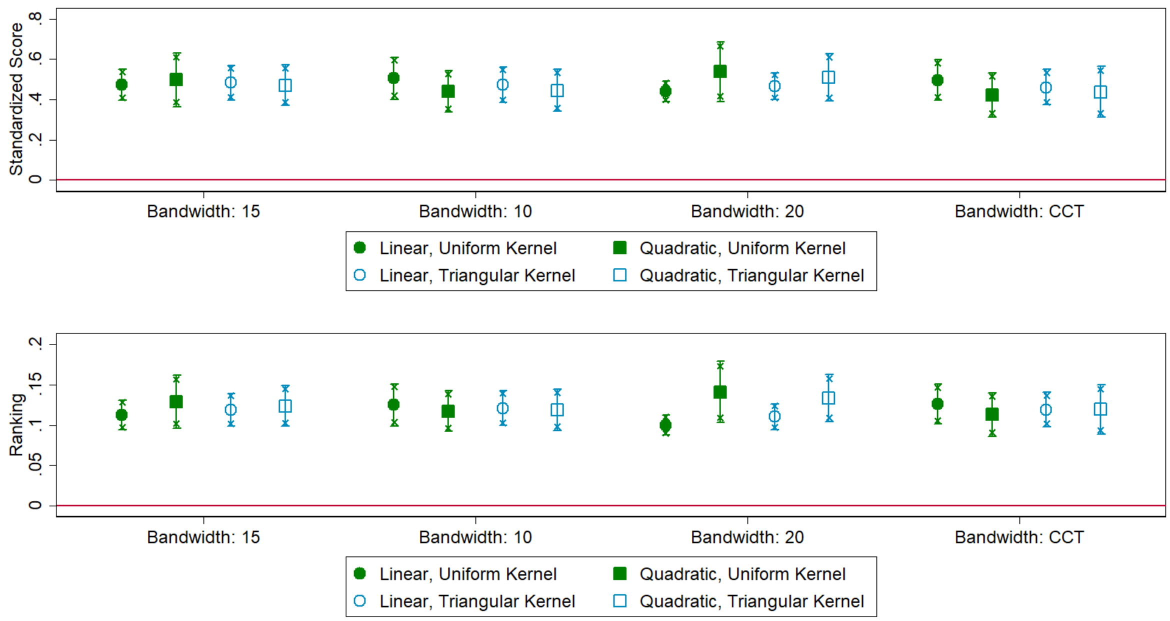
Notes: This figure plots the individual characteristics (gender, Han ethnicity, Hui ethnicity, hukou status, being a first-time taker, and age) against the distance to the tier-2 cutoff score. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure A.2: Robustness: Effects of Below Tier-2 University Cutoff on Retake Probability



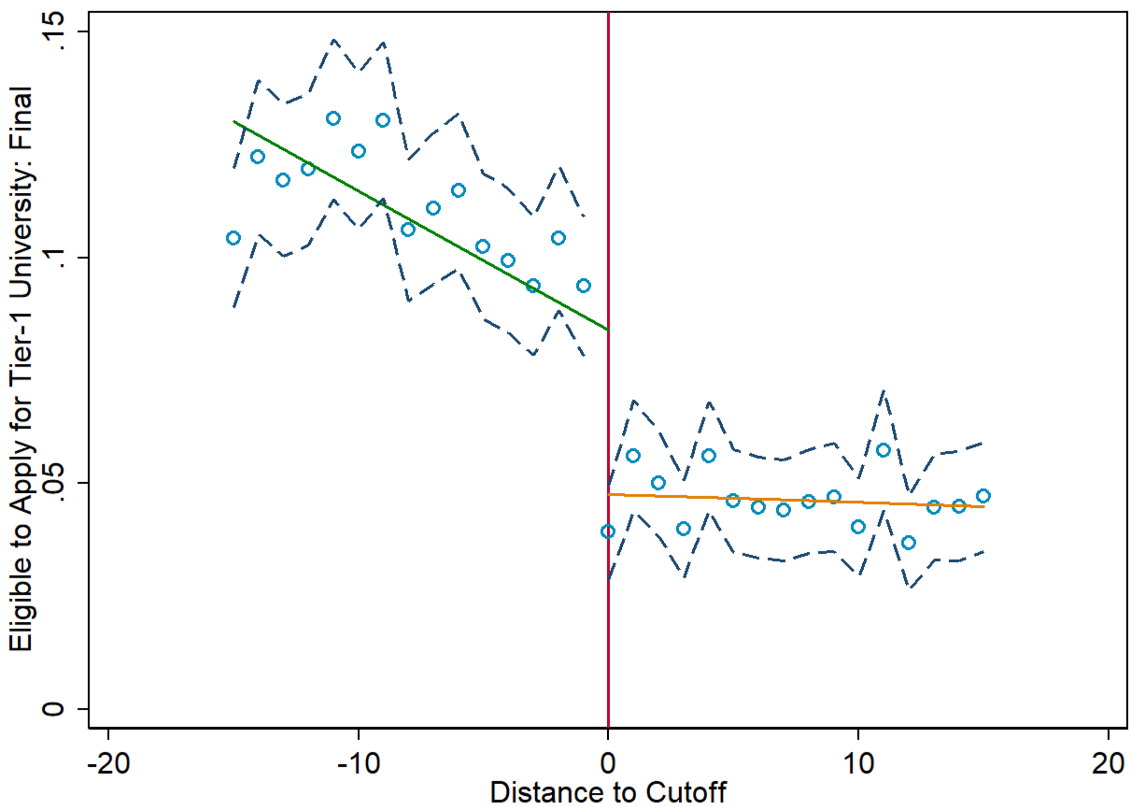
Notes: This figure plots the estimated effects of below the tier-2 cutoff on retake probability using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The parametric function of the running variable and its interaction with the indicator of below the cutoff are controlled in all regressions. Gender, ethnicity, hukou status, whether the individual is a first-time taker, age and year-by-track fixed effects are controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Figure A.3: Robustness: Effects of Retake on Exam Outcomes



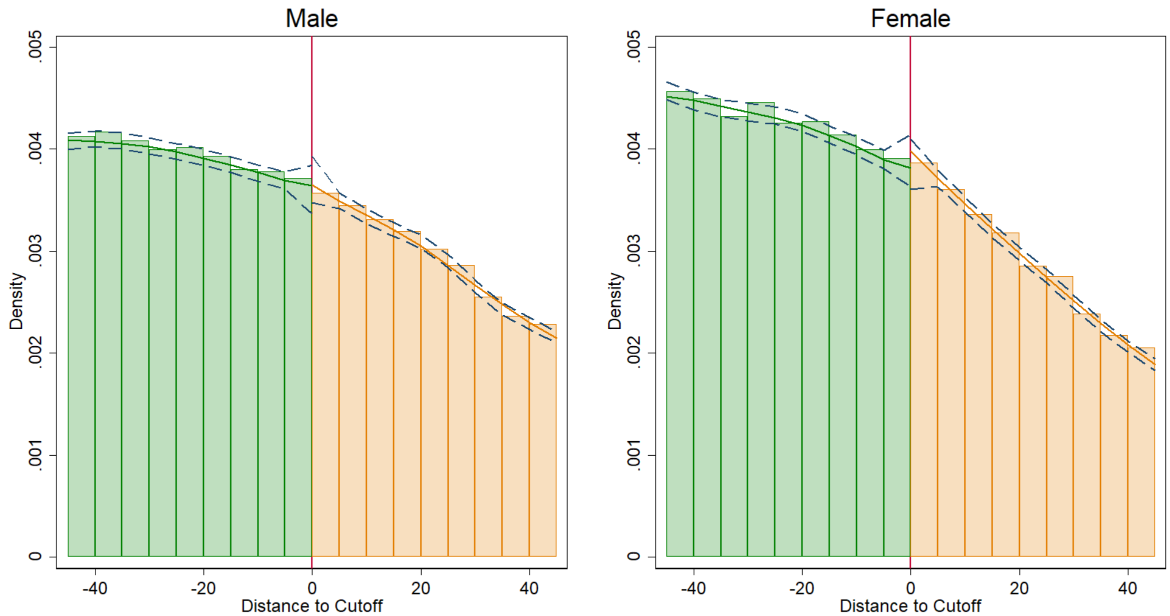
Notes: This figure plots the estimated effects of retaking the NCEE on exam outcomes using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The indicator of below the cutoff is used as an instrument for the indicator of retaking the NCEE in the 2SLS specification, with the differences between the final outcomes and the initial outcomes (standardized score, ranking within the year-track) as the dependent variables. The parametric function of the running variable and its interaction with the indicator of below the cutoff are controlled in all regressions. Individual characteristics (gender, ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects are controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Figure A.4: Eligibility to Apply for Tier-1 University in Terms of Final Outcome vs. Distance to Tier-2 University Cutoff Score



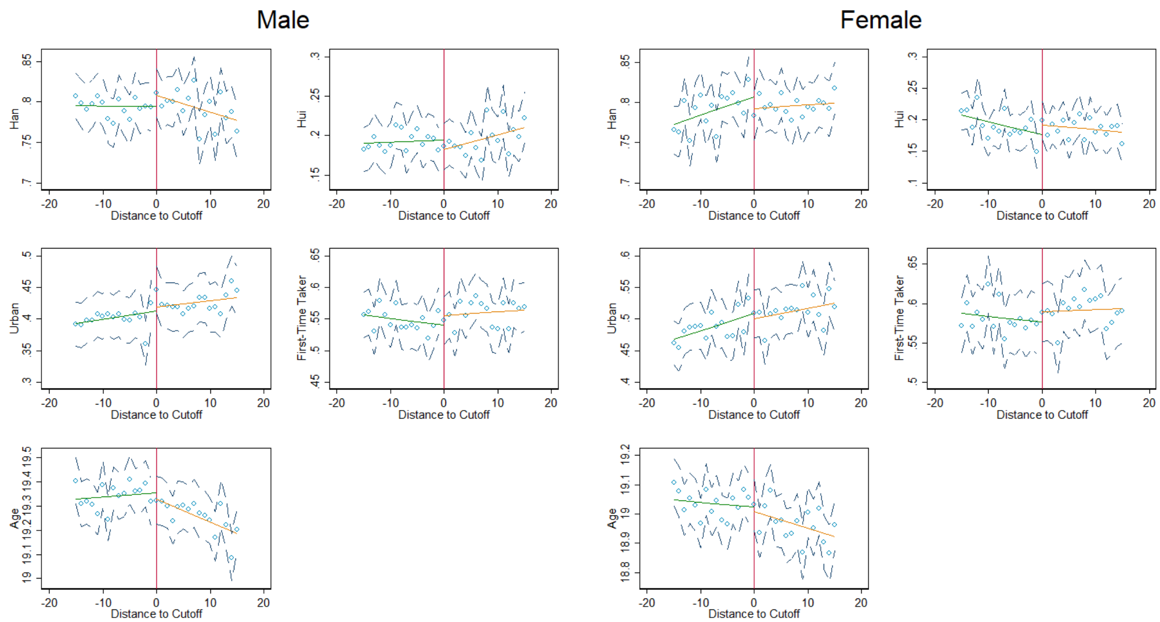
Notes: This figure plots the probability of being eligible to apply for tier-1 universities in terms of the final outcome, which is equal to the dependent variable in the current year if the individual does not retake the NCEE in the next year, and is equal to the dependent variable in the next year if the individual retakes the NCEE in the next year, against the distance to the tier-2 cutoff score. The individual is defined as eligible to apply for tier-1 universities if the score of the individual is above or equal to the tier-1 cutoff in the corresponding year-track. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure A.5: Test of Running Variable Density Smoothness around Tier-2 Cutoff by Gender



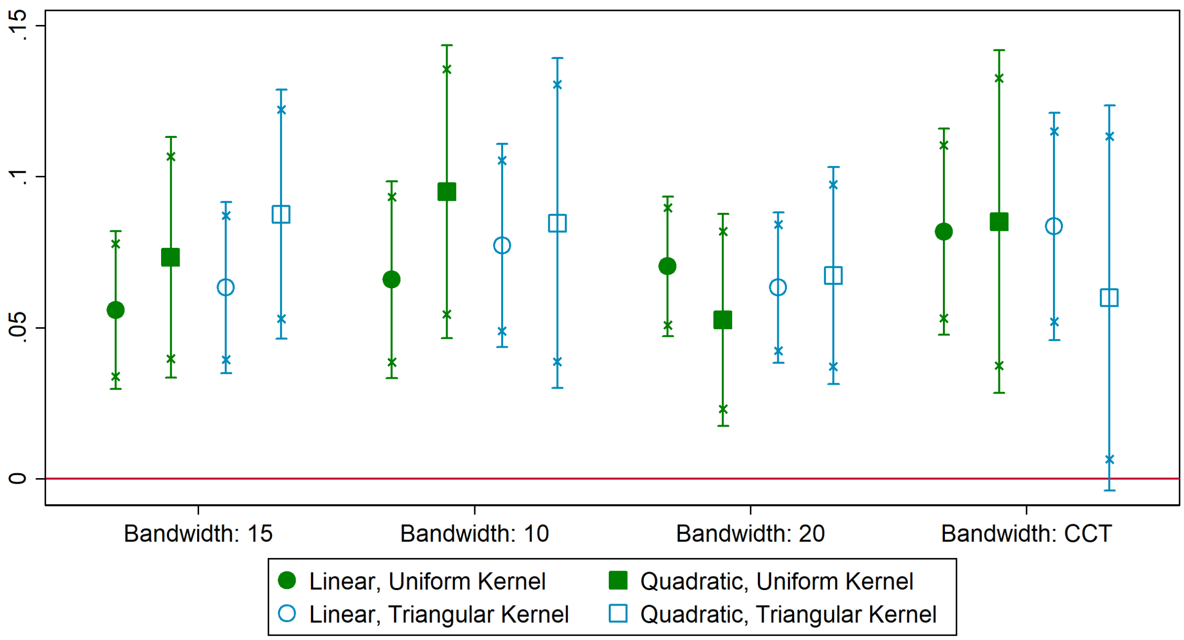
Notes: This figure plots the density of the running variable (the distance to the tier-2 cutoff score) following the manipulation testing procedure in Cattaneo, Jansson & Ma (2018) for males and females separately. The bars in this figure represent the density distribution of the running variable over 5-point bins. The straight lines represent the estimated density to the left and to the right of the cutoff using the local polynomial density estimators proposed in Cattaneo et al. (2020). The dashed lines represent the lower and upper bounds of the 95% confidence interval for the estimated density.

Figure A.6: Covariates Balancedness by Gender



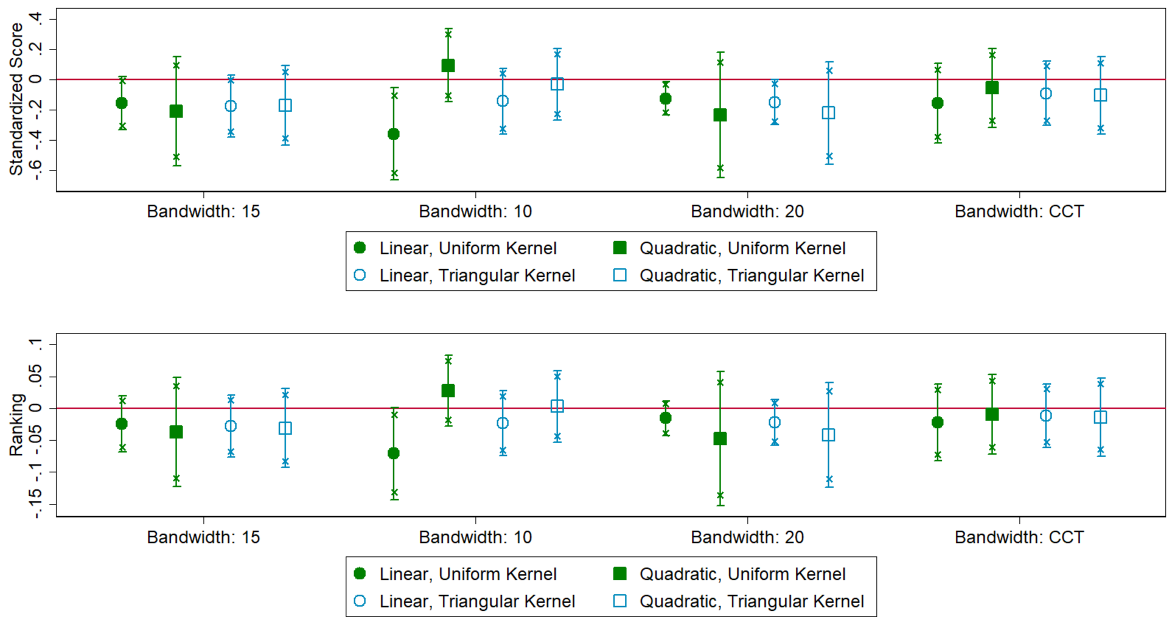
Notes: This figure plots the individual characteristics (Han ethnicity, Hui ethnicity, hukou status, being a first-time taker, and age) against the distance to the tier-2 cutoff score for males and females separately. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure A.7: Robustness: Gender Differences in the Effects of Below Cutoff on Retake Probability



Notes: This figure plots the estimated gender differences in the effects of below the tier-2 cutoff on retake probability using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The parametric function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in all regressions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects, are also controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Figure A.8: Robustness: Gender Differences in the Effects of Retake on Exam Outcomes



Notes: This figure plots the estimated gender differences in the effects of retaking the NCEE on exam outcomes using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The indicator of below the cutoff and its interaction with male dummy are used as instruments for the indicator of retaking the NCEE and its interaction with male dummy in the 2SLS specification, with the differences between the final outcomes and the initial outcomes (standardized score, ranking within the year-track) as the dependent variables. The parametric function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in all regressions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects, are also controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Table A.1: Balancing Tests

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Male	Han	Hui	Urban	First-Time Taker	Age
Panel A: Linear Control						
Below Cutoff	0.0146	0.0002	-0.0030	-0.0052	-0.0175*	0.0295
	(0.0092)	(0.0075)	(0.0073)	(0.0099)	(0.0103)	(0.0270)
Panel B: Quadratic Control						
Below Cutoff	0.0095	0.0056	-0.0086	-0.0159	-0.0100	0.0759**
	(0.0142)	(0.0116)	(0.0116)	(0.0154)	(0.0141)	(0.0373)
Observations	41,477	41,477	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15	15	15
Year-Track FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The sample consists of observations within the 15-point bandwidth around the cutoff. Panel A controls for a linear function of the running variable and its interaction with the indicator of below the cutoff. Panel B controls for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table A.2: Effects of Below Tier-2 University Cutoff on Retake Probability, Alternative Clustering

	(1)	(2)	(3)	(4)
Variables		Retake		
Panel A: Baseline, Two-way Clustering by Individual Identifier and High School-Year				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0081)	(0.0076)	(0.0122)	(0.0115)
Panel B: Two-way Clustering by Individual Identifier and High School				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0079)	(0.0076)	(0.0121)	(0.0119)
Panel C: Two-way Clustering by Running Variable and Individual Identifier				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0089)	(0.0081)	(0.0148)	(0.0124)
Panel D: Two-way Clustering by Running Variable and High School-Year				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0090)	(0.0081)	(0.0142)	(0.0119)
Observations	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15
Interaction Controls	Linear	Linear	Quadratic	Quadratic
Individual Characteristics	No	Yes	No	Yes
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (2) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (3) and (4) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Columns (1) and (3) do not control for individual characteristics. Columns (2) and (4) control for a set of individual characteristics, including gender, ethnicity, hukou status, whether the individual is a first-time taker, and age. Year-by-track fixed effects are controlled in all columns. In Panel A, standard errors are two-way clustered at the individual identifier level and the high school-year level. In Panel B, standard errors are two-way clustered at the individual identifier level and the high school level. In Panel C, standard errors are two-way clustered at the running variable level and the individual identifier level. In Panel D, standard errors are two-way clustered at the running variable level and the high school-year level.

Table A.3: Summary Statistics of Exam Outcomes

	(1)	(2)	(3)	(4)
Sample	Full	[-15,15]	[-15,0)	[0,15]
Observations	362,592	41,477	21,123	20,354
		Mean (S.D.)		
Standardized Score: Initial	0.00	0.77	0.69	0.86
	(1.00)	(0.21)	(0.19)	(0.19)
Ranking: Initial	0.50	0.77	0.74	0.80
	(0.29)	(0.06)	(0.06)	(0.05)
Standardized Score: Final	0.13	0.86	0.82	0.90
	(1.03)	(0.30)	(0.33)	(0.25)
Ranking: Final	0.54	0.79	0.77	0.80
	(0.30)	(0.08)	(0.09)	(0.06)

Notes: This table shows the summary statistics of the exam outcomes. Initial outcomes are the variables in the current year. Final outcomes are the final payoffs of the outcomes, and are equal to the variables in the current year if the individual does not retake the NCEE in the next year, and are equal to the variables in the next year if the individual retakes the NCEE in the next year. Column (1) is for the full sample. Column (2) is for the sample within the 15-point bandwidth around the tier-2 cutoff. Column (3) is for the sample in column (2) that is below the cutoff. Column (4) is for the sample in column (2) that is above or equal to the cutoff.

Table A.4: Predicting the NCEE Retakes

	(1)	(2)	(3)	(4)
Sample	Full	Full	[-15,15]	[-15,15]
Variables	Retake			
Male	0.0207***	0.0273***	0.0620***	0.0782***
	(0.0019)	(0.0019)	(0.0039)	(0.0037)
Han		0.1131***		0.1756***
		(0.0060)		(0.0125)
Hui		0.0415***		0.0100
		(0.0080)		(0.0132)
Urban		-0.1264***		-0.0599***
		(0.0055)		(0.0056)
First-Time Taker		0.1403***		0.1843***
		(0.0050)		(0.0065)
Age		-0.0248***		-0.0306***
		(0.0014)		(0.0018)
Standardized Score		-0.0161***		-1.1963***
		(0.0041)		(0.0282)
Observations	362,592	362,572	41,477	41,477
R-squared	0.004	0.055	0.013	0.207
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (2) use the full sample. Columns (3) and (4) use the sample within the 15-point bandwidth around the cutoff. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table A.5: Balancing Tests by Gender

	(1)	(2)	(3)	(4)	(5)
Variables	Han	Hui	Urban	First-Time Taker	Age
Panel A: Male, Linear Control					
Below Cutoff	-0.0181*	0.0132	-0.0174	-0.0169	0.0324
	(0.0105)	(0.0105)	(0.0134)	(0.0137)	(0.0381)
Panel B: Male, Quadratic Control					
Below Cutoff	-0.0034	-0.0037	-0.0429**	-0.0069	0.0775
	(0.0168)	(0.0166)	(0.0206)	(0.0204)	(0.0579)
Observations	21,162	21,162	21,162	21,162	21,162
Panel C: Female, Linear Control					
Below Cutoff	0.0200*	-0.0204**	0.0093	-0.0164	0.0164
	(0.0109)	(0.0102)	(0.0143)	(0.0147)	(0.0339)
Panel D: Female, Quadratic Control					
Below Cutoff	0.0163	-0.0149	0.0129	-0.0135	0.0700
	(0.0161)	(0.0154)	(0.0219)	(0.0209)	(0.0478)
Observations	20,315	20,315	20,315	20,315	20,315
Bandwidth	15	15	15	15	15
Year-Track FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The sample consists of observations within the 15-point bandwidth around the cutoff. Panels A and B are for the male sample. Panels C and D are for the female sample. Panels A and C control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Panels B and D control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table A.6: Gender Differences in the Effects of Below Tier-2 University Cutoff on Retake Probability, Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Retake					
Male*Below Cutoff	0.0492***	0.0548***	0.0558***	0.0724***	0.0720***	0.0732***
	(0.0144)	(0.0133)	(0.0133)	(0.0218)	(0.0203)	(0.0203)
Below Cutoff	0.0545***	0.0553***	0.0549***	0.0360**	0.0384***	0.0377***
	(0.0098)	(0.0093)	(0.0093)	(0.0145)	(0.0138)	(0.0137)
Male	0.0453***	0.0584***		0.0334***	0.0465***	
	(0.0071)	(0.0072)		(0.0093)	(0.0097)	
Observations	41,477	41,477	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic
Individual Characteristics	No	Without Gender Interaction	With Gender Interaction	No	Without Gender Interaction	With Gender Interaction
Year-Track FE	Yes	Yes	By Gender	Yes	Yes	By Gender

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1)-(3) control for a linear function of the running variable and its interaction with the indicator of below the cutoff, as well as their gender interactions. Columns (4)-(6) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff, as well as their gender interactions. Columns (1) and (4) control for year-by-track fixed effects. Columns (2) and (5) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects. Columns (3) and (6) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Appendix B

Appendix for "Vulnerable Boys: Short-term and Long-term Gender Differences in the Impacts of Adolescent Disadvantage"

B.1 Data Appendix: Description of NLSY97 Sample

NLSY97 is a representative longitudinal study with surveys from 1997 (round 1) to 2015-2016 (round 17). The cohort was born between 1980 and 1984, with respondents aged between 12 and 18 at the time of the first interview and between 30 and 36 at round 17. 8,984 individuals were initially interviewed in round 1. Nearly 80 percent (7,103) of the round 1 sample were interviewed in round 17. Consistent with Add Health sample, we uses subsamples of 1,486 non-Hispanic white women and 1,515 non-Hispanic white men who lived with their biological mother in round 1 survey. Table B.3 shows summary statistics of important variables in both Add Health sample and NLSY97 sample. For most of the characteristics, these two samples are comparable.

B.1.1 Adult Outcomes

The round 17 interview collected an array of adult outcomes, including educational attainment, marriage, number of births, employment, income and depression.

- *Educational Attainment*: There are two survey questions on educational attainment. The first one is the highest grade completed as of the survey date, from 1st grade to 12th grade, and from 1st year college to 8th year college or more. The second one is the highest degree received as of the survey date. In order to construct the educational attainment variable to be directly comparable to Add Health sample, the educational attainment is divided into 4 categories: "less than high school", "high school degree", "some college", and "college degree or more".[1]

  As shown in Table B.3, although the survey questions on educational attainment are different in two samples, the proportions of respondents with high school degree or more, and some college or more, are very similar in two samples. Though the proportion of respondents with college degree or more is higher in NLSY97 sample, for both female and male. In general, these two samples are comparable.

- *Employment and Income*: A dummy variable of currently employed is defined based on whether the respondent reports to receive any income from a job such as wages,

[1] The educational attainment is defined to be "less than high school" if the respondent reports the highest degree to be "none". The educational attainment is defined to be "high school degree" if the respondent reports the highest degree to be GED or high school diploma, and the highest grade completed to be less or equal to 12th grade. The educational attainment is defined to be "some college" if the respondent reports the highest degree to be Associate/Junior college, or that the respondent reports the highest degree to be GED or high school diploma but the highest grade completed to be more or equal to 1st year college. The educational attainment is defined to be "college degree or more" if the respondent reports the highest degree to be Bachelor's degree, Master's degree, PhD or Professional degree (DDS, JD, MD).

salary, commissions, or tips. Personal income is measured by self-reported before-tax job income in 2014, in thousand of dollars. Job satisfaction for the primary current job is measured on a 5 point scale.

- *Marriage and Children Ever Born*: A dummy variable of ever married is defined based on the reported marriage history. The respondents were also asked about number of biological children born and residing in the household/born but not residing in the household as of the survey date. The children ever born variable is the sum of these two reported numbers.

- *Depression*: Respondents were asked how often during the past month they felt nervous, calm and peaceful, down or blue, happy, and depressed. Factor analysis indicated that a single factor is appropriate for these 5 items and was used to form a standardized depression index.

B.1.2 Disadvantage: Father Absence

During the round 1 survey, 93% of the NLSY97 white non-Hispanic respondents were living with their biological mother. Of this group, 14% were living with a step-father or other father figure, and 21% were living with no father figure at all. As shown in Table B.3, these two samples are comparable, while the proportion of respondents living with other father rather than biological father is higher in NLSY97 sample, for both female and male.

B.1.3 Maternal Characteristics

Maternal characteristics are included as control variables in most regressions. In the NLSY97 sample, there is only one survey question, the mother's highest grade completed, on biological mother's educational attainment. In order to be consistent with the categorical definitions in Add Health sample, the maternal educational attainment is defined to be "less than high school" if the highest grade completed is less or equal to 11th grade, "high school degree" if the highest grade completed is equal to 12th grade, "some college" if the highest grade completed is between 1st year college and 3rd year college, "college degree or more" if the highest grade completed is between 4th year college and 8th year college. As shown in Table B.3, NLSY97 sample has smaller proportion of mother with high school degree, but larger proportion of mother with some college education. This is because the survey questions on educational attainment are different in two samples, making it hard to distinguish between "high school degree" and "some college" in the NLSY97 sample. However, the proportion of mother with college degree or more is very similar in these two samples.

B.2 Appendix Tables

Table B.1: Sample Selection: Add Health Sample

	(1) Full Sample	(2) With Complete Data	(3) White Non-Hispanic Subsample	(4) Living with Biological Mother	(5) With Wave IV Data
Mother's Education: High School Degree	0.40	0.41	0.44	0.44	0.44
	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)
Mother's Education: Some College	0.18	0.18	0.19	0.20	0.20
	(0.38)	(0.39)	(0.39)	(0.40)	(0.40)
Mother's Education: College Degree or More	0.24	0.23	0.25	0.26	0.26
	(0.42)	(0.42)	(0.43)	(0.44)	(0.44)
Young Mother	0.11	0.12	0.11	0.12	0.15
	(0.31)	(0.33)	(0.32)	(0.33)	(0.36)
Foreign Mother	0.15	0.12	0.04	0.04	0.03
	(0.35)	(0.33)	(0.19)	(0.19)	(0.18)
No Father	0.30	0.31	0.23	0.23	0.22
	(0.46)	(0.46)	(0.42)	(0.42)	(0.42)
Other Father	0.10	0.10	0.10	0.10	0.10
	(0.30)	(0.30)	(0.31)	(0.30)	(0.30)
Family Income at Baseline Survey	45.73	45.73	52.53	53.56	53.50
	(51.62)	(51.03)	(54.67)	(55.58)	(54.98)
Low Birth Weight	0.10	0.10	0.08	0.08	0.08
	(0.31)	(0.30)	(0.28)	(0.27)	(0.27)
Male	0.49	0.49	0.49	0.49	0.47
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age at Baseline Survey	15.10	15.11	15.04	14.98	14.98
	(1.75)	(1.74)	(1.73)	(1.72)	(1.72)
Number of Observations	20745	17320	10227	8916	7327

Notes: This table reports summary statistics for the Add Health sample. The first column is the full Add Health Wave I sample. The second column drops the observations without race/ethnicity data, or without Wave I sample weights. The third column further drops the observations that are not White Non-Hispanic. The fourth column keeps only those living with biological mother at Wave I, and with non-missing maternal characteristics. The fifth column keeps the observations that are still in the sample at Wave IV survey. The sample shown in column (5) is the main sample investigated in this paper.

Table B.2: Descriptive Statistics, Add Health Non-Hispanic White Sample

	(1) Female	(2) Male	(3) Differences
Family Background			
No Father	0.236	0.205	0.031
	(0.425)	(0.404)	(0.010)
No Father Recently	0.114	0.110	0.004
	(0.318)	(0.313)	(0.007)
No Father Always	0.122	0.095	0.026
	(0.327)	(0.293)	(0.007)
Other Father	0.100	0.103	-0.003
	(0.300)	(0.304)	(0.007)
Mother High School	0.433	0.438	-0.005
	(0.496)	(0.496)	(0.012)
Mother Some College	0.194	0.207	-0.013
	(0.396)	(0.405)	(0.009)
Mother College Graduate	0.264	0.260	0.003
	(0.441)	(0.439)	(0.010)
Young Mother	0.152	0.149	0.004
	(0.359)	(0.356)	(0.008)
Foreign Mother	0.032	0.034	-0.003
	(0.175)	(0.182)	(0.004)
Adolescent Outcomes (Wave I)			
School Problems Index	-0.144	0.160	-0.303
	(0.941)	(1.039)	(0.023)
Depression Index	0.122	-0.136	0.258
	(1.086)	(0.874)	(0.023)
Ever Suspended from School	0.130	0.292	-0.162
	(0.336)	(0.455)	(0.009)
English Grade	3.095	2.720	0.375
	(0.911)	(0.972)	(0.022)
Math Grade	2.841	2.731	0.111
	(1.022)	(1.046)	(0.025)
Want to Attend College	0.107	-0.120	0.227
	(0.883)	(1.105)	(0.023)
Expect to Attend College	0.129	-0.144	0.273
	(0.911)	(1.073)	(0.023)
Expect to Be Married by Age 25	0.069	-0.078	0.147
	(0.998)	(0.997)	(0.023)
Adult Outcomes (Wave IV)			
High School Graduate	0.948	0.920	0.028
	(0.222)	(0.271)	(0.006)
College Graduate	0.405	0.326	0.078
	(0.491)	(0.469)	(0.011)
Currently Employed	0.723	0.863	-0.140
	(0.448)	(0.344)	(0.010)
Personal Income	28.709	44.308	-15.599
	(30.138)	(47.273)	(0.932)
Number of Times Fired	0.312	0.640	-0.328
	(0.770)	(1.625)	(0.030)
Job Satisfaction	3.900	3.920	-0.020
	(0.937)	(0.927)	(0.022)
Financial Stress	0.232	0.199	0.033
	(0.422)	(0.399)	(0.010)
Ever Married	0.609	0.506	0.103
	(0.488)	(0.500)	(0.012)
Children Ever Born	0.996	0.677	0.320
	(1.122)	(0.988)	(0.025)
Depression Index	0.076	-0.064	0.140
	(1.120)	(1.004)	(0.025)
Number of Observations	3868	3459	7327

Notes: Standard errors in parentheses.

Table B.3: Summary Statistics: Add Health and NLSY97 Non-Hispanic White Sample

	(1)	(2)	(3)	(4)
	Add Health Sample		NLSY97 Sample	
	Female	Male	Female	Male
High School Graduation	0.940	0.916	0.952	0.947
	(0.238)	(0.278)	(0.214)	(0.225)
At Least Some College	0.738	0.638	0.715	0.640
	(0.440)	(0.481)	(0.452)	(0.480)
College Graduation	0.388	0.320	0.443	0.358
	(0.487)	(0.466)	(0.497)	(0.480)
Mother's Education: High School Degree	0.447	0.458	0.337	0.357
	(0.497)	(0.498)	(0.473)	(0.479)
Mother's Education: Some College	0.194	0.197	0.288	0.256
	(0.396)	(0.397)	(0.453)	(0.436)
Mother's Education: College Degree or More	0.247	0.248	0.249	0.273
	(0.431)	(0.432)	(0.433)	(0.446)
No Father at Baseline Survey	0.235	0.209	0.220	0.186
	(0.424)	(0.407)	(0.415)	(0.389)
Other Father at Baseline Survey	0.098	0.099	0.145	0.138
	(0.297)	(0.299)	(0.352)	(0.345)
Family Income at Baseline Survey	43.117	43.403	47.979	46.134
	(47.923)	(49.048)	(47.815)	(45.894)
Age at Baseline Survey	14.727	14.938	14.289	14.295
	(1.751)	(1.829)	(1.491)	(1.501)
Number of Observations	3868	3459	1486	1515

Notes: The Add Health sample is weighted by Wave I sample weights. The NLSY97 sample is weighted by 1997 sample weights. The baseline survey for Add Health sample is in 1994-1995, while the baseline survey for NLSY97 sample is in 1997. The family income is measured in thousands of dollar.

Table B.4: Identification Tests, Add Health Non-Hispanic White Sample

VARIABLES	(1) Mother High School	(2) Mother Some College	(3) Mother College	(4) Foreign Mother	(5) Young Mother	(6) Family Income	(7) Low Birth Weight	(8) Attrition
Male	-0.00417	0.0172	0.00748	-0.00109	0.00314	-0.400	-0.0133	0.0521***
	(0.0199)	(0.0139)	(0.0153)	(0.00576)	(0.0122)	(1.380)	(0.00948)	(0.0115)
Male*No Father	0.0416	-0.0450	-0.0113	-0.00134	0.00282	3.782	-0.00803	-0.0234
	(0.0399)	(0.0316)	(0.0266)	(0.0125)	(0.0316)	(2.478)	(0.0187)	(0.0252)
Male*Other Father	0.0524	-0.0280	-0.0474	0.0132	0.0160	-4.550	-0.00924	0.0297
	(0.0535)	(0.0498)	(0.0414)	(0.0164)	(0.0488)	(3.858)	(0.0236)	(0.0371)
No Father	-0.0453*	0.0604***	-0.0701***	-0.00297	0.102***	-22.59***	0.0188	0.0436**
	(0.0268)	(0.0211)	(0.0217)	(0.00789)	(0.0220)	(1.882)	(0.0136)	(0.0203)
Other Father	0.00965	0.0637*	-0.0845***	-0.0154*	0.202***	-2.993	0.0133	0.0118
	(0.0377)	(0.0352)	(0.0271)	(0.00921)	(0.0314)	(2.919)	(0.0199)	(0.0207)
Constant	0.457***	0.174***	0.270***	0.0333***	0.119***	48.25***	0.0613***	0.142***
	(0.0218)	(0.0112)	(0.0205)	(0.00441)	(0.00964)	(1.985)	(0.00785)	(0.00917)
Observations	7,327	7,327	7,327	7,327	7,327	7,327	6,690	8,916
R-squared	0.002	0.003	0.009	0.000	0.035	0.033	0.002	0.006

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at Wave I. "Mother High School", "Mother Some College", and "Mother College" are dummies that equal to 1 when mother's education falls in the category, omitting the category of "Mother Less Than High School". "Foreign Mother" is a dummy for foreign-born mother. "Young Mother" is a dummy for mother with age under 22. "Family Income" is annual income from all sources in thousands of 1994 dollars. "Low Birth Weight" is less than 88 oz. All regressions are weighted by Wave I sample weights.

Table B.5: Adolescent Outcomes: School Problems Index and Components

VARIABLES	(1) School Problems Index	(2) Absences	(3) Ever Suspended from School	(4) Trouble with Teachers	(5) Trouble paying Attention	(6) Trouble with Homework	(7) Trouble with Students
Male	0.263***	0.191	0.136***	0.246***	0.121***	0.196***	0.154***
	(0.0384)	(0.229)	(0.0140)	(0.0376)	(0.0369)	(0.0450)	(0.0341)
Male*No Father	0.179**	1.017	0.0741***	0.0779	0.179**	0.139	-0.0190
	(0.0787)	(0.638)	(0.0263)	(0.0688)	(0.0790)	(0.0842)	(0.0653)
Male*Other Father	0.0466	0.226	0.0682*	-0.0158	0.144	-0.0212	-0.139
	(0.0884)	(0.443)	(0.0368)	(0.0920)	(0.0972)	(0.119)	(0.0976)
No Father	0.175***	0.856**	0.102***	0.0956*	0.107*	0.114**	0.0931
	(0.0569)	(0.404)	(0.0191)	(0.0500)	(0.0555)	(0.0537)	(0.0581)
Other Father	0.146**	0.242	0.0366*	0.0395	0.0598	0.192**	0.141**
	(0.0702)	(0.355)	(0.0191)	(0.0658)	(0.0674)	(0.0932)	(0.0638)
Constant	-0.999***	-12.87***	-0.505***	2.116***	-0.375	-0.420	2.411***
	(0.353)	(2.082)	(0.143)	(0.268)	(0.326)	(0.343)	(0.302)
Observations	7,203	7,205	7,325	7,210	7,209	7,209	7,210
R-squared	0.050	0.049	0.120	0.034	0.030	0.029	0.022
Mean of dependent variable	0.000	1.686	0.207	0.862	1.313	1.220	0.883
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "No Father" and "Other Father" refer to living arrangements at Wave I. "School problems" is a standardized index based on factor analysis of the other variables in this table. "Absences" is student-reported absences without excuse in past year, "trouble" variables from student-reported experiences from 0=never to 4=every day. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave I weights.

Table B.6: Adolescent Outcomes: Depression Index and Components

VARIABLES	(1) Depression Index	(2) Hopeful about Future	(3) Can't Shake Blues	(4) Depressed	(5) Happy	(6) Lonely	(7) Sad
Male	-0.240***	0.0416	-0.144***	-0.163***	-0.0193	-0.133***	-0.179***
	(0.0295)	(0.0292)	(0.0194)	(0.0223)	(0.0321)	(0.0219)	(0.0214)
Male*No Father	-0.000464	-0.0576	-0.0291	-0.0965**	-0.0187	0.0296	0.00848
	(0.0663)	(0.0701)	(0.0529)	(0.0433)	(0.0672)	(0.0458)	(0.0427)
Male*Other Father	-0.205*	-0.0273	-0.189**	-0.151**	-0.0132	-0.00819	-0.134*
	(0.115)	(0.0997)	(0.0795)	(0.0699)	(0.0869)	(0.0699)	(0.0781)
No Father	0.184***	-0.0304	0.127***	0.167***	-0.0668*	0.0571*	0.0669**
	(0.0489)	(0.0437)	(0.0313)	(0.0356)	(0.0390)	(0.0299)	(0.0335)
Other Father	0.272***	-0.0169	0.213***	0.178***	-0.0586	0.104*	0.139**
	(0.0969)	(0.0697)	(0.0754)	(0.0658)	(0.0636)	(0.0531)	(0.0602)
Constant	-1.651***	0.915***	-0.616***	-0.595***	3.080***	-0.806***	-0.181
	(0.260)	(0.260)	(0.156)	(0.189)	(0.261)	(0.157)	(0.167)
Observations	7,311	7,320	7,323	7,326	7,327	7,326	7,326
R-squared	0.063	0.022	0.045	0.054	0.021	0.030	0.038
Mean of dependent variable	0.000	1.870	0.364	0.488	2.183	0.423	0.537
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at Wave I. "Depression Index" is the CES-D depression scale (standardized) based on 19 items, including the other variables in this table. Each item is based on responses to "How often have you felt this way during the past week?" ranging from 0=never or rarely to 3=most/all of the time. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave I weights.

Table B.7: Mechanisms: Adolescent Self-Reports about Relationship with Parents

VARIABLES	(1) Parents Care about Me	(2) Family Has Fun	(3) Mother Warm and Loving	(4) Satisfied with Relationship with Mother
Male	-0.00162	0.0527	0.0875***	0.156***
	(0.0161)	(0.0352)	(0.0236)	(0.0277)
Male*No Father	0.0102	-0.0531	0.0378	0.0278
	(0.0356)	(0.0754)	(0.0515)	(0.0560)
Male*Other Father	-0.0116	0.0123	0.104	0.0429
	(0.0479)	(0.0859)	(0.0726)	(0.0752)
No Father	-0.0666***	-0.148***	-0.147***	-0.126***
	(0.0248)	(0.0434)	(0.0416)	(0.0473)
Other Father	-0.0429	-0.231***	-0.0999*	-0.0176
	(0.0373)	(0.0734)	(0.0600)	(0.0646)
Constant	5.383***	6.132***	5.491***	6.303***
	(0.106)	(0.252)	(0.176)	(0.205)
Observations	7,306	7,301	7,319	7,318
R-squared	0.013	0.038	0.026	0.036
Mean of dependent variable	4.82	3.72	4.42	4.29
Mother's characteristics	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at Wave I. Dependent variables are measured on a 1-5 scale (column 1 and 2) or a 0-4 scale (column 3 and 4) from "strongly disagree" to "strongly agree". Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave I weights.

Table B.8: Alternative Measures of Educational Attainment, Add Health Non-Hispanic White Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Years of Completed Education					Categorical Educational Attainment: Ordered Logit				
VARIABLES	Full Sample		Mother High School	Mother Some College	Mother College Grad	Full Sample		Mother High School	Mother Some College	Mother College Grad
Male	-0.556***	-0.614***	-0.490***	-0.478**	-0.870***	-0.412***	-0.503***	-0.430***	-0.375**	-0.634***
	(0.0897)	(0.0974)	(0.119)	(0.229)	(0.169)	(0.0668)	(0.0769)	(0.0942)	(0.175)	(0.130)
Male*No Father		-0.0671	0.00157	-0.470	-0.126		-0.150	-0.134	-0.450	-0.0975
		(0.166)	(0.196)	(0.432)	(0.402)		(0.137)	(0.161)	(0.358)	(0.310)
Male*Other Father		0.301	0.297	-0.188	0.657		0.203	0.0972	0.0791	0.543
		(0.212)	(0.280)	(0.419)	(0.436)		(0.169)	(0.240)	(0.357)	(0.336)
No Father		-0.743***	-0.715***	-0.643***	-1.058***		-0.539***	-0.491***	-0.539***	-0.824***
		(0.113)	(0.136)	(0.239)	(0.304)		(0.0900)	(0.108)	(0.181)	(0.226)
Other Father		-0.802***	-0.659***	-0.529*	-1.423***		-0.556***	-0.357**	-0.510**	-1.053***
		(0.130)	(0.194)	(0.303)	(0.294)		(0.110)	(0.174)	(0.249)	(0.232)
Constant	11.85***	9.962***	11.43***	10.97***	13.06***					
	(1.139)	(0.897)	(1.133)	(1.565)	(1.217)					
Observations	7,327	7,327	3,932	1,468	1,922	7,327	7,327	3,932	1,468	1,922
R-squared	0.014	0.199	0.049	0.056	0.076					
Mean of dependent variable	14.14		13.36	14.22	15.66	2.54		2.08	2.63	3.43
Mother's characteristics	NO	YES	YES	YES	YES	NO	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at Wave I. "Years of education" range from 10 to 20 years. "Categorical Educational Attainment" is a discrete measure ranging from 0=less than high school to 5=post-graduate degree. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. All models are weighted by Wave IV weights.

Table B.9: School Quality Effects and Neighborhood Effects, Add Health Non-Hispanic White Sample

	Adolescent Outcomes							Adult Outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
VARIABLES	School Problem Index	Ever Suspended from School	Depression Index	Wants to Attend College	Expects to Attend College	English Grade	Math Grade	College Graduation	High School Graduation	Currently Employed	Personal Income	Ever Married	Children Ever Born
Male	0.491***	0.154**	-0.172	-0.461***	-0.481***	-0.505***	-0.177	-0.0601	-0.0546**	0.152**	7.583	-0.175**	-0.589***
	(0.110)	(0.0654)	(0.118)	(0.160)	(0.127)	(0.130)	(0.115)	(0.0595)	(0.0231)	(0.0671)	(6.894)	(0.0663)	(0.132)
Male*Educated Neighborhood	0.0185	0.132	-0.315	0.240	0.500**	-0.0120	-0.621*	-0.166	0.0753	-0.0443	17.73	-0.134	0.586***
	(0.285)	(0.102)	(0.324)	(0.245)	(0.220)	(0.248)	(0.317)	(0.149)	(0.0553)	(0.117)	(16.77)	(0.134)	(0.216)
Male*School Quality Index	-0.0136	-0.0362**	-0.00430	0.0536	0.0543*	0.0329	0.0344	0.0153	0.00818	-0.000227	-0.391	-0.00246	0.0528
	(0.0329)	(0.0159)	(0.0377)	(0.0350)	(0.0290)	(0.0346)	(0.0325)	(0.0179)	(0.00982)	(0.0193)	(2.324)	(0.0189)	(0.0339)
School Quality Index	-0.0212	-0.0238*	-0.0196	0.0434	0.0704***	0.0482*	0.0651*	0.0324*	0.0175**	0.00912	0.826	0.00745	-0.0672**
	(0.0263)	(0.0133)	(0.0288)	(0.0265)	(0.0251)	(0.0265)	(0.0361)	(0.0169)	(0.00818)	(0.0145)	(1.115)	(0.0165)	(0.0287)
Educated Neighborhood	0.0825	-0.196**	0.182	0.550***	0.656***	0.150	0.359	0.469***	0.0150	0.174*	16.59*	0.0149	-0.886***
	(0.260)	(0.0909)	(0.306)	(0.168)	(0.157)	(0.224)	(0.416)	(0.127)	(0.0433)	(0.0924)	(8.397)	(0.110)	(0.182)
Male*No Father	0.173*	0.0597*	0.0159	-0.177**	-0.0952	0.0141	-0.156*	0.0207	-0.0350	-0.0581	-2.280	-0.0584	-0.259***
	(0.0974)	(0.0330)	(0.0781)	(0.0778)	(0.0917)	(0.108)	(0.0935)	(0.0363)	(0.0214)	(0.0396)	(3.234)	(0.0353)	(0.0977)
Male*Other Father	0.0697	0.0470	-0.299**	-0.0299	-0.0677	-0.00936	0.0357	0.0251	-0.00757	0.0766	4.711	-0.0147	-0.109
	(0.0981)	(0.0481)	(0.141)	(0.0937)	(0.109)	(0.115)	(0.138)	(0.0443)	(0.0346)	(0.0501)	(8.460)	(0.0521)	(0.110)
Male*Median Houshold Income	-9.33e-06**	-7.28e-07	-5.14e-07	6.50e-06	2.84e-06	3.44e-06	8.46e-06**	4.54e-07	2.90e-07	-5.35e-07	0.000148	2.57e-06	5.37e-06
	(3.88e-06)	(2.05e-06)	(4.12e-06)	(4.93e-06)	(4.06e-06)	(4.19e-06)	(4.08e-06)	(2.27e-06)	(6.75e-07)	(2.00e-06)	(0.000229)	(2.14e-06)	(4.21e-06)
No Father	0.173**	0.116***	0.198***	-0.0449	-0.220***	-0.260***	-0.140**	-0.116***	-0.0365***	-0.0173	-5.320***	-0.0427	0.231***
	(0.0657)	(0.0220)	(0.0576)	(0.0647)	(0.0727)	(0.0636)	(0.0693)	(0.0248)	(0.0125)	(0.0317)	(1.702)	(0.0276)	(0.0744)
Other Father	0.195**	0.0535**	0.405***	0.00823	-0.125	-0.0944	-0.122	-0.116***	-0.0332	-0.0881**	-2.976	-0.0277	0.154*
	(0.0862)	(0.0258)	(0.128)	(0.0706)	(0.0878)	(0.0796)	(0.120)	(0.0280)	(0.0209)	(0.0395)	(3.487)	(0.0429)	(0.0898)
Median Household Income	1.03e-05**	1.97e-06	4.22e-07	-4.26e-06	-2.38e-06	-7.04e-06	-7.77e-07	4.68e-06**	-2.35e-07	2.95e-06*	0.000428***	-1.05e-05***	-1.73e-05***
	(4.70e-06)	(1.87e-06)	(3.70e-06)	(3.51e-06)	(3.35e-06)	(4.38e-06)	(5.34e-06)	(1.92e-06)	(7.12e-07)	(1.71e-06)	(0.000137)	(1.71e-06)	(3.54e-06)
Constant	-0.815*	-0.396**	-1.047***	0.861***	-0.769**	2.688***	2.961***	-0.188	0.629***	0.364**	-41.65***	-0.595***	-0.169
	(0.474)	(0.174)	(0.318)	(0.312)	(0.318)	(0.462)	(0.665)	(0.195)	(0.108)	(0.139)	(15.71)	(0.174)	(0.421)
Observations	5,160	5,265	5,257	5,256	5,255	5,052	4,770	5,267	5,267	4,391	5,102	5,264	5,253
R-squared	0.042	0.139	0.057	0.109	0.177	0.087	0.032	0.181	0.094	0.050	0.067	0.077	0.134
Mean of dependent variable	0.035	0.223	0.051	-0.046	-0.024	2.883	2.748	0.377	0.943	0.803	37.795	0.590	0.878
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. "School Quality Index" is a standardized index based on school administrator reports of average daily attendance, class size, % of new and experienced teachers, % of teachers with a Masters' degree, grade 12 dropout rates, % of students with achievement tests below and above grade level, and % of 12th graders enrolled in college next year. "Educated Neighborhood" is the proportion of aged 25+ individuals with college degree or more at census tract level. "No Father" and "Other Father" refer to living arrangements at Wave I. "Median Household Income" is the median household income at county level, and is included as an additional control variable. "School problems" is a standardized index based on factor analysis of the variables in Table B.5. "Depression" is a standardized index based on factor analysis of the variables in Table B.6. Grades are student-reported and range from 1=D or lower to 4=A. College desires/expectations are standardized measures based on a 0-4 scale. "Currently Employed" is a dummy for whether is currently working at Wave IV. "Personal Income" is the total personal earnings before tax in 2006/2007/2008. "Ever Married" is a dummy for whether has been married up to Wave IV. "Children Ever Born" is the self-reported total number of child births up to Wave IV. Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. Models (1)-(7) are weighted by Wave I weights and models (8)-(13) are weighted by Wave IV weights.

Table B.10: Educational Attainment and Adolescent Outcomes, Add Health Black Sample

VARIABLES	(1) School Problem Index	(2) Ever Suspended from School	(3) Depression Index	(4) Expects to Attend College	(5) Wants to Attend College	(6) English Grade	(7) Math Grade	(8) College Graduation	(9) High School Graduation
Male	0.205**	0.240***	-0.345***	-0.112	-0.0794	-0.194**	-0.156**	-0.134***	-0.0593***
	(0.0887)	(0.0352)	(0.0985)	(0.0941)	(0.0747)	(0.0791)	(0.0749)	(0.0327)	(0.0195)
Male*No Father	-0.0809	-0.130***	-0.0445	-0.125	-0.0627	-0.0932	0.0731	0.0386	-0.0108
	(0.114)	(0.0479)	(0.127)	(0.117)	(0.0995)	(0.0935)	(0.106)	(0.0436)	(0.0308)
Male*Other Father	0.0194	-0.0816	0.155	-0.219	-0.266	-0.263	-0.264	-0.0210	-0.00263
	(0.206)	(0.0893)	(0.210)	(0.237)	(0.213)	(0.187)	(0.186)	(0.0551)	(0.0631)
No Father	0.161**	0.152***	0.0402	-0.0152	-0.0133	-0.0991	-0.113*	-0.0954***	-0.0659***
	(0.0727)	(0.0317)	(0.0976)	(0.0825)	(0.0672)	(0.0790)	(0.0668)	(0.0280)	(0.0193)
Other Father	0.176	0.163***	0.125	-0.148	-0.0974	-0.0869	-0.00640	-0.0969*	-0.0516
	(0.131)	(0.0562)	(0.139)	(0.174)	(0.131)	(0.125)	(0.128)	(0.0516)	(0.0356)
Constant	-0.651	0.316	-1.069**	1.234**	2.046***	2.525***	2.972***	0.280	0.515*
	(0.617)	(0.255)	(0.478)	(0.479)	(0.501)	(0.457)	(0.605)	(0.235)	(0.261)
Observations	2,658	2,700	2,692	2,687	2,692	2,616	2,545	2,704	2,704
R-squared	0.025	0.082	0.065	0.074	0.048	0.049	0.023	0.169	0.068
Mean of dependent variable	0.000	0.400	0.000	0.000	0.000	2.703	2.491	0.295	0.918
Mother's characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. Standard errors clustered by school. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "No Father" and "Other Father" refer to living arrangements at Wave I. "School problems" is a standardized index based on factor analysis of the variables in Table B.5. "Depression" is a standardized index based on factor analysis of the variables in Table B.6. Grades are student-reported and range from 1=D or lower to 4=A. College desires/expectations are standardized measures based on a 0-4 scale.Mother's characteristics include education and dummies for foreign-born and young mother (under 22). All models include birth cohort. Models (1)-(7) are weighted by Wave I weights and models (8)-(9) are weighted by Wave IV weights.

Table B.11: Alternative Specifications for Educational Attainment, NLSY97 Non-Hispanic White Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	College Graduation		High School Graduation		Number of Grades Completed		Categorical Educational Attainment: Ordered Logit	
Male	-0.0852***	-0.113***	-0.00506	-0.0151**	-0.565***	-0.693***	-0.339***	-0.584***
	(0.0180)	(0.0205)	(0.00803)	(0.00721)	(0.111)	(0.116)	(0.0676)	(0.0904)
Male*No Father		0.0142		0.00722		0.119		0.218
		(0.0391)		(0.0236)		(0.258)		(0.183)
Male*Other Father		0.0758*		0.0372		-0.0159		0.363*
		(0.0410)		(0.0308)		(0.285)		(0.202)
No Father		-0.205***		-0.0524***		-1.426***		-1.100***
		(0.0276)		(0.0155)		(0.180)		(0.128)
Other Father		-0.250***		-0.0797***		-1.263***		-1.133***
		(0.0304)		(0.0224)		(0.219)		(0.152)
Constant	0.433***	-0.155*	0.937***	0.794***	14.85***	10.70***		
	(0.0881)	(0.0916)	(0.0386)	(0.0531)	(0.536)	(0.587)		
Observations	2,998	2,998	2,998	2,998	2,981	2,981	2,998	2,998
R-squared	0.008	0.224	0.000	0.059	0.009	0.250		
Mean of dependent variable	0.400		0.949		14.54		3.025	
Mother's characteristics	NO	YES	NO	YES	NO	YES	NO	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "No Father" and "Other Father" refer to living arrangements at baseline survey (1997). "Number of Grades Completed" ranges from 6th grades to 8th year college (20th grades). "Categorical Educational Attainment" is a discrete measure ranging from 1=less than high school to 4=college or higher degree. Mother's characteristics include education and mother's age when the first child in the household and the respondent was/were born. All models include birth cohort. All models are weighted by round 17 sample weights.

Appendix C

Appendix for "Short-run and Long-run Effects of Peers from Disrupted Families"

C.1 Appendix Tables

Table C.1: Data Description

Variables	Definition
Adolescent outcomes (wave I survey)	
Absences	categorical variable (0-3) on number of times they have been absent without an excuse
Ever suspended from school	dummy variable on whether they have received an out-of-school suspension
Trouble with teachers	categorical variable (0-4) on how often they have trouble getting along with teachers
Trouble paying attention	categorical variable (0-4) on how often they have trouble paying attention in school
Trouble with homework	categorical variable (0-4) on how often they have trouble getting homework done
Trouble with students	categorical variable (0-4) on how often they have trouble getting along with other students
School problem index	standardized index from factor analysis of the above six variables
Depression index	standardized index from factor analysis of 19 items from the CES-D depression scale
English grade	categorical variable (1-4) on the grade in English or language arts
Math grade	categorical variable (1-4) on the grade in mathematics
Wants to attend college	categorical variable (1-5) on how much they want to go to college
Expects to attend college	categorical variable (1-5) on how likely is it that they will go to college
Adult outcomes (wave IV survey)	
College graduation	dummy variable on whether they have completed college
High school graduation	dummy variable on whether they have completed high school
Personal income	self-reported before-tax personal earnings
Financial stress	dummy variable equal to 1 if there is a positive response to at least one of: "without phone service because you didn't have enough money", "didn't pay the full amount of the rent or mortgage because you didn't have enough money", "were evicted from your house or apartment for not paying the rent or mortgage", "didn't pay the full amount of a gas, electricity, or oil bill because you didn't have enough money", "had the service turned off by the gas or electric company, or the oil company wouldn't deliver, because payments were not made", and "worried whether food would run out before you would get money to buy more"
Currently employed	dummy variable equal to 1 if working for pay at least 10 hours a week
Number of times fired	number of times they have been fired, let go or laid off from a job from 2001 to the previous year
Job satisfaction	categorical variable (1-5) on how satisfied they are with this job
Ever arrested	dummy variable on whether they have ever been arrested
Ever married	dummy variable on whether they have ever been married
Children ever born	number of live births from the pregnancies of them or their partner
Peer compositions (in-school survey)	
% Peers from disrupted families: same gender-race	proportion of same-gender same-race peers within the same school-grade not living with both parents
% Peers from disrupted families: not same gender-race	proportion of peers that are not of the same gender-race but within the same school-grade not living with both parents
% Peers from disrupted families: same gender	proportion of same gender peers within the same school-grade not living with both parents
% Peers from disrupted families: not same gender	proportion of peers that are not of the same gender but within the same school-grade not living with both parents
% Peers from disrupted families: same race	proportion of same race peers within the same school-grade not living with both parents
% Peers from disrupted families: not same race	proportion of peers that are not of the same race but within the same school-grade not living with both parents
% Peers female	proportion of female students within the same school-grade-race
% Peers African-American	proportion of African-American students within the same school-grade-gender
% Peers Hispanic	proportion of Hispanic students within the same school-grade-gender
% Peers other	proportion of other race/ethnicity students within the same school-grade-gender
% Peers with college-educated mother	proportion of students whose mother has a college degree within the same school-grade-gender-race
% Peers with high school-educated mother	proportion of students whose mother has a high school degree within the same school-grade-gender-race
Family and individual characteristics (wave I survey)	
No father	dummy variable equal to 1 if not living with any father figure at Wave I survey
Other father	dummy variable equal to 1 if living with non-biological father figure at Wave I survey
No mother	dummy variable equal to 1 if not living with any mother figure at Wave I survey
Other mother	dummy variable equal to 1 if living with non-biological mother figure at Wave I survey
Father high school graduate	dummy variable equal to 1 if father's educational attainment is classified as high school graduate
Father some college	dummy variable equal to 1 if father's educational attainment is classified as some college
Father college graduate	dummy variable equal to 1 if father's educational attainment is classified as at least college graduate
Mother high school graduate	dummy variable equal to 1 if mother's educational attainment is classified as high school graduate
Mother some college	dummy variable equal to 1 if mother's educational attainment is classified as some college
Mother college graduate	dummy variable equal to 1 if mother's educational attainment is classified as at least college graduate
Young mother	dummy variable equal to 1 if mother gave birth at age under 22
Foreign mother	dummy variable equal to 1 if mother was foreign-born
ZPVT score	standardized score of Add Health Peabody Picture Vocabulary Test

Table C.2: Adolescent Outcomes, Female Subsample

Variables	(1) School problem index	(2) Depression index	(3) English grade	(4) Math grade	(5) Wants to attend college	(6) Expects to attend college
% Peers from disrupted families: same gender-race	-0.140	-0.531	-0.399	-0.0360	0.106	-0.202
	(0.305)	(0.366)	(0.288)	(0.378)	(0.293)	(0.297)
% Peers from disrupted families: not same gender-race	-0.172	0.812	-0.950	-1.557*	-0.156	0.0987
	(0.579)	(0.646)	(0.623)	(0.814)	(0.675)	(0.523)
Observations	4,950	4,942	4,896	4,643	4,953	4,949
Mean of dependent variable	-0.113	0.140	3.004	2.716	0.072	0.094
R-W adjusted p-value	0.9406	0.2475	0.2673	0.9406	0.9406	0.8317
Family and individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Other peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. "Family and individual characteristics" include age, ZPVT score, parental education, family structure variables and maternal characteristics variables. "Other peer characteristics" include % Peers female, % Peers African American, % Peers Hispanic, % Peers other, % Peers with college-educated mother, and % Peers with high school-educated mother. Grade fixed effects, school-by-race fixed effects and school-by-race time trends are controlled in all models. All models are weighted by Wave I weights.

Bibliography

Alan, S., Boneva, T. & Ertac, S. (2019), 'Ever failed, try again, succeed better: Results from a randomized educational intervention on grit', *The Quarterly Journal of Economics* **134**(3), 1121–1162.

Almond, D. & Edlund, L. (2007), 'Trivers–willard at birth and one year: evidence from us natality data 1983–2001', *Proceedings of the Royal Society B: Biological Sciences* **274**(1624), 2491–2496.

Amato, P. R. & Gilbreth, J. G. (1999), 'Nonresident fathers and children's well-being: A meta-analysis', *Journal of Marriage and the Family* pp. 557–573.

Ananat, E., Shihe, F. & Ross, S. L. (2018), 'Race-specific urban wage premia and the black-white wage gap', *Journal of Urban Economics* **108**, 141–153.

Anderson, K. G. (2017), 'Adverse childhood environment: Relationship with sexual risk behaviors and marital status in a large american sample', *Evolutionary Psychology* **15**(2), 1474704917710115.

Anelli, M. & Peri, G. (2015), 'Peers' composition effects in the short and in the long run: College major, college performance and income'.

Astorne-Figari, C. & Speer, J. D. (2019), 'Are changes of major major changes? the roles of grades, gender, and preferences in college major switching', *Economics of Education Review* **70**, 75–93.

Autor, D., Figlio, D., Karbownik, K., Roth, J. & Wasserman, M. (2019), 'Family disadvantage and the gender gap in behavioral and educational outcomes', *American Economic Journal: Applied Economics* **11**(3), 338–81.

Autor, D. & Wasserman, M. (2013), 'Wayward sons: The emerging gender gap in labor markets and education', *Third Way Report* .

Bailey, M. J. & Dynarski, S. M. (2011), 'Inequality in postsecondary education'. In G.J. Duncan and R.J. Murnane (eds.), Whither Opportunity? Rising Inequality, Schools, and Childrens Life Chances. (Russell Sage: New York, New York, September 2011).

Becker, G. S., Hubbard, W. H. & Murphy, K. M. (2010), 'Explaining the worldwide boom in higher education of women', *Journal of Human Capital* **4**(3), 203–241.

Bedard, K. & Witman, A. (2020), 'Family structure and the gender gap in adhd', *Review of Economics of the Household* **18**(4), 1101–1129.

Berlin, N. & Dargnies, M.-P. (2016), 'Gender differences in reactions to feedback and willingness to compete', *Journal of Economic Behavior & Organization* **130**, 320–336.

Bertrand, M. & Pan, J. (2013), 'The trouble with boys: Social influences and the gender gap in disruptive behavior', *American economic journal: applied economics* **5**(1), 32–64.

Beyer, S. (1998), 'Gender differences in causal attributions by college students of performance on course examinations', *Current psychology* **17**(4), 346–358.

Bifulco, R., Fletcher, J. M. & Ross, S. L. (2011), 'The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health', *American Economic Journal: Economic Policy* **3**(1), 25–53.

Billings, S. B., Deming, D. J. & Ross, S. L. (2019), 'Partners in crime', *American Economic Journal: Applied Economics* **11**(1), 126–50.

Billings, S. B. & Hoekstra, M. (2019), Schools, neighborhoods, and the long-run effect of crime-prone peers, Technical report, National Bureau of Economic Research.

Borghans, L., Duckworth, A. L., Heckman, J. J. & Ter Weel, B. (2008), 'The economics and psychology of personality traits', *Journal of human Resources* **43**(4), 972–1059.

Boring, A. & Brown, J. (2016), Gender, competition and choices in higher education, Technical report, Tech. rep. Working Paper.

Brenøe, A. A. & Lundberg, S. (2018), 'Gender gaps in the effects of childhood family environment: Do they persist into adulthood?', *European Economic Review* **109**, 42–62.

Buchmann, C., DiPrete, T. A. & McDaniel, A. (2008), 'Gender inequalities in education', *Annu. Rev. Sociol* **34**, 319–337.

Buser, T., Niederle, M. & Oosterbeek, H. (2014), 'Gender, competitiveness, and career choices', *The quarterly journal of economics* **129**(3), 1409–1447.

Buser, T., Peter, N. & Wolter, S. C. (2017), 'Gender, competitiveness, and study choices in high school: Evidence from switzerland', *American economic review* **107**(5), 125–30.

Buser, T. & Yuan, H. (2019), 'Do women give up competing more easily? evidence from the lab and the dutch math olympiad', *American Economic Journal: Applied Economics* **11**(3), 225–52.

Cai, X., Lu, Y., Pan, J. & Zhong, S. (2019), 'Gender gap under pressure: Evidence from china's national college entrance examination', *Review of Economics and Statistics* **101**(2), 249–263.

Calonico, S., Cattaneo, M. D., Farrell, M. H. & Titiunik, R. (2017), 'rdrobust: Software for regression-discontinuity designs', *The Stata Journal* **17**(2), 372–404.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014), 'Robust nonparametric confidence intervals for regression-discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2015), 'Optimal data-driven regression discontinuity plots', *Journal of the American Statistical Association* **110**(512), 1753–1769.

Carrell, S. E., Hoekstra, M. & Kuka, E. (2018), 'The long-run effects of disruptive peers', *American Economic Review* **108**(11), 3377–3415.

Carrell, S. E. & Hoekstra, M. L. (2010), 'Externalities in the classroom: How children exposed to domestic violence affect everyone's kids', *American Economic Journal: Applied Economics* **2**(1), 211–28.

Cattaneo, M. D., Idrobo, N. & Titiunik, R. (2018), 'A practical introduction to regression discontinuity designs: Volume ii'.

Cattaneo, M. D., Jansson, M. & Ma, X. (2018), 'Manipulation testing based on density discontinuity', *The Stata Journal* **18**(1), 234–261.

Cattaneo, M. D., Jansson, M. & Ma, X. (2020), 'Simple local polynomial density estimators', *Journal of the American Statistical Association* **115**(531), 1449–1455.

Chetty, R. & Hendren, N. (2018), 'The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects', *The Quarterly Journal of Economics* **133**(3), 1107–1162.

Chetty, R., Hendren, N., Lin, F., Majerovitz, J. & Scuderi, B. (2016), 'Childhood environment and gender gaps in adulthood', *American Economic Review* **106**(5), 282–88.

Clarke, D., Romano, J. P. & Wolf, M. (2020), 'The romano–wolf multiple-hypothesis correction in stata', *The Stata Journal* **20**(4), 812–843.

Cools, A., Fernández, R. & Patacchini, E. (2019), Girls, boys, and high achievers, Technical report, National Bureau of Economic Research.

Correia, S. (2017), 'Linear models with high-dimensional fixed effects: An efficient and feasible estimator', *Working Paper* .

Cross, C. J. (2020), 'Racial/ethnic differences in the association between family structure and children's education', *Journal of Marriage and Family* **82**(2), 691–712.

Dahl, G. B. & Moretti, E. (2008), 'The demand for sons', *The review of economic studies* **75**(4), 1085–1120.

Davis, D. R., Dingel, J. I., Monras, J. & Morales, E. (2019), 'How segregated is urban consumption?', *Journal of Political Economy* **127**(4), 1684–1738.

Dee, T. S. (2007), 'Teachers and the gender gaps in student achievement', *Journal of Human resources* **42**(3), 528–554.

Delaney, J. & Devereux, P. J. (2021), 'Gender and educational achievement: Stylized facts and causal evidence'.

DiPrete, T. A. & Buchmann, C. (2013), *The rise of women: The growing gender gap in education and what it means for American schools*, Russell Sage Foundation.

DiPrete, T. A. & Jennings, J. L. (2012), 'Social and behavioral skills and the gender gap in early educational achievement', *Social Science Research* **41**(1), 1–15.

Dougherty, C. (2005), 'Why are the returns to schooling higher for women than for men?', *Journal of Human Resources* **40**(4), 969–988.

Duckworth, A. L. & Seligman, M. E. (2006), 'Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores.', *Journal of educational psychology* **98**(1), 198.

Dweck, C. S., Davidson, W., Nelson, S. & Enna, B. (1978), 'Sex differences in learned helplessness: Ii. the contingencies of evaluative feedback in the classroom and iii. an experimental analysis.', *Developmental psychology* **14**(3), 268.

Ellison, G. & Swanson, A. (2018), Dynamics of the gender gap in high math achievement, Technical report, National Bureau of Economic Research.

Fahle, E. M. & Reardon, S. F. (2018), ''Education'. in 'State of the union: The poverty and inequality report'', *Special issue, Pathways Magazine* .

Fan, X., Fang, H. & Markussen, S. (2015), Mothers' employment and children's educational gender gap, Technical report, National Bureau of Economic Research.

Fang, C., Zhang, E. & Zhang, J. (2021), 'Do women give up competing more easily? evidence from speedcubers', *Economics Letters* **205**, 109943.

Fletcher, J. M., Ross, S. L. & Zhang, Y. (2020), 'The consequences of friendships: Evidence on the effect of social relationships in school on academic achievement', *Journal of Urban Economics* **116**, 103241.

Flory, J. A., Leibbrandt, A. & List, J. A. (2015), 'Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions', *The Review of Economic Studies* **82**(1), 122–155.

Fortin, N. M., Oreopoulos, P. & Phipps, S. (2015), 'Leaving boys behind gender disparities in high academic achievement', *Journal of Human Resources* **50**(3), 549–579.

Fruehwirth, J. C., Iyer, S. & Zhang, A. (2019), 'Religion and depression in adolescence', *Journal of Political Economy* **127**(3), 1178–1209.

Goldin, C., Katz, L. F. & Kuziemko, I. (2006), 'The homecoming of american college women: The reversal of the college gender gap', *Journal of Economic perspectives* **20**(4), 133–156.

Golsteyn, B. H., Grönqvist, H. & Lindahl, L. (2014), 'Adolescent time preferences predict lifetime outcomes', *The Economic Journal* **124**(580), F739–F761.

Goodman, J., Gurantz, O. & Smith, J. (2020), 'Take two! sat retaking and college enrollment gaps', *American Economic Journal: Economic Policy* **12**(2), 115–58.

Gould, E. D., Lavy, V. & Paserman, M. D. (2011), 'Sixty years after the magic carpet ride: The long-run effect of the early childhood environment on social and economic outcomes', *The Review of Economic Studies* **78**(3), 938–973.

Guryan, J., Kroft, K. & Notowidigdo, M. J. (2009), 'Peer effects in the workplace: Evidence from random groupings in professional golf tournaments', *American Economic Journal: Applied Economics* **1**(4), 34–68.

Ha, W., Kang, L. & Song, Y. (2020), 'College matching mechanisms and matching stability: Evidence from a natural experiment in china', *Journal of Economic Behavior & Organization* **175**, 206–226.

Hamoudi, A. & Nobles, J. (2014), 'Do daughters really cause divorce? stress, pregnancy, and family composition', *Demography* **51**(4), 1423–1449.

Heckman, J. J., Stixrud, J. & Urzua, S. (2006), 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor economics* **24**(3), 411–482.

Hellerstein, J. K., McInerney, M. & Neumark, D. (2011), 'Neighbors and coworkers: The importance of residential labor market networks', *Journal of Labor Economics* **29**(4), 659–695.

Hetherington, E. M. & Kelly, J. (2002), *For better or for worse: Divorce reconsidered*, WW Norton & Company.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475.

Jacob, B. A. (2002), 'Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education', *Economics of Education review* **21**(6), 589–598.

Jia, R. & Li, H. (2021), 'Just above the exam cutoff score: Elite college admission and wages in china', *Journal of Public Economics* **196**, 104371.

Kamp Dush, C. M., Arocho, R., Mernitz, S. & Bartholomew, K. (2018), 'The intergenerational transmission of partnering', *PloS one* **13**(11), e0205732.

Keller, M. C., Nesse, R. M. & Hofferth, S. (2001), 'The trivers–willard hypothesis of parental investment: no effect in the contemporary united states', *Evolution and Human Behavior* **22**(5), 343–360.

Kleibergen, F. & Paap, R. (2006), 'Generalized reduced rank tests using the singular value decomposition', *Journal of econometrics* **133**(1), 97–126.

Kolesár, M. & Rothe, C. (2018), 'Inference in regression discontinuity designs with a discrete running variable', *American Economic Review* **108**(8), 2277–2304.

Krishna, K., Lychagin, S. & Frisancho, V. (2018), 'Retaking in high stakes exams: is less more?', *International Economic Review* **59**(2), 449–477.

Kristoffersen, J. H. G., Krægpøth, M. V., Nielsen, H. S. & Simonsen, M. (2015), 'Disruptive school peers and student outcomes', *Economics of Education Review* **45**, 1–13.

Landaud, F. & Maurin, E. (2020), 'Aim high and persevere! competitive pressure and access gaps in top science graduate programs'.

Lavy, V. & Schlosser, A. (2011), 'Mechanisms and impacts of gender peer effects at school', *American Economic Journal: Applied Economics* **3**(2), 1–33.

Leadbeater, B. J., Kuperminc, G. P., Blatt, S. J. & Hertzog, C. (1999), 'A multivariate model of gender differences in adolescents' internalizing and externalizing problems.', *Developmental psychology* **35**(5), 1268.

Lee, D. S. & Card, D. (2008), 'Regression discontinuity inference with specification error', *Journal of Econometrics* **142**(2), 655–674.

Lei, Z. & Lundberg, S. (2020), 'Vulnerable boys: Short-term and long-term gender differences in the impacts of adolescent disadvantage.', *Journal of Economic Behavior & Organization* **178**, 424–448.

Lopoo, L. M. & DeLeire, T. (2014), 'Family structure and the economic wellbeing of children in youth and adulthood', *Social Science Research* **43**, 30–44.

Lundberg, S. (2005), 'Sons, daughters, and parental behaviour', *Oxford Review of Economic Policy* **21**(3), 340–356.

Lundberg, S. (2017*a*), 'Father absence and the educational gender gap'.

Lundberg, S. (2017*b*), *'Non-cognitive skills as human capital'. in 'Education, skills, and technical change: Implications for Future US GDP Growth'*, University of Chicago Press.

Lundberg, S. & Rose, E. (2003), 'Child gender and the transition to marriage', *Demography* **40**(2), 333–349.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *The review of economic studies* **60**(3), 531–542.

McLanahan, S. & Sandefur, G. D. (1994), *Growing up with a single parent: What hurts, what helps*, Harvard University Press.

McLanahan, S., Tach, L. & Schneider, D. (2013), 'The causal effects of father absence', *Annual review of sociology* **39**, 399–427.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S. et al. (2011), 'A gradient of childhood self-control predicts health, wealth, and public safety', *Proceedings of the national Academy of Sciences* **108**(7), 2693–2698.

Niederle, M. & Vesterlund, L. (2007), 'Do women shy away from competition? do men compete too much?', *The quarterly journal of economics* **122**(3), 1067–1101.

Norberg, K. (2004), 'Partnership status and the human sex ratio at birth', *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**(1555), 2403–2410.

OECD (2015), 'The abc of gender equality in education: Aptitude, behaviour, confidence.'.

Olivetti, C., Patacchini, E. & Zenou, Y. (2020), 'Mothers, peers, and gender-role identity', *Journal of the European Economic Association* **18**(1), 266–301.

Owens, J. (2016), 'Early childhood behavior problems and the gender gap in educational attainment in the united states', *Sociology of education* **89**(3), 236–258.

Prevoo, T. & Ter Weel, B. (2015), 'The effect of family disruption on childrens personality development: Evidence from british longitudinal data', *De economist* **163**(1), 61–93.

Radloff, L. S. (1977), 'The ces-d scale: A self-report depression scale for research in the general population', *Applied psychological measurement* **1**(3), 385–401.

Reuben, E., Wiswall, M. & Zafar, B. (2017), 'Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender', *The Economic Journal* **127**(604), 2153–2186.

Romano, J. P. & Wolf, M. (2016), 'Efficient computation of adjusted p-values for resampling-based stepdown multiple testing', *Statistics & Probability Letters* **113**, 38–40.

Ryckman, D. B. & Peckham, P. D. (1987), 'Gender differences in attributions for success and failure', *The Journal of Early Adolescence* **7**(1), 47–63.

Salisbury, J., Rees, G. & Gorard, S. (1999), 'Accounting for the differential attainment of boys and girls at school', *School Leadership & Management* **19**(4), 403–426.

Saygin, P. O. (2016), 'Gender differences in preferences for taking risk in college applications', *Economics of Education Review* **52**, 120–133.

Segal, C. (2013), 'Misbehavior, education, and labor market outcomes', *Journal of the European Economic Association* **11**(4), 743–779.

Slade, A. N., Beller, A. H. & Powers, E. T. (2017), 'Family structure and young adult health outcomes', *Review of Economics of the Household* **15**(1), 175–197.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D. & Trautmann, S. T. (2013), 'Impatience and uncertainty: Experimental decisions predict adolescents' field behavior', *American Economic Review* **103**(1), 510–31.

Trivers, R. L. & Willard, D. E. (1973), 'Natural selection of parental ability to vary the sex ratio of offspring', *Science* **179**(4068), 90–92.

U.S. Census Bureau (2016), 'Educational attainment in the united states: 2015.'. http://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html.

Wang, F. & Wu, H. (2018), 'Returns to education in rural and urban china: An empirical study', *Asian Journal of Social Science Studies* **3**(4), 18.

Wasserman, M. (2020), 'Gender differences in politician persistence'.

Woessmann, L. et al. (2015), 'An international look at the single-parent family', *Education Next* **15**(2), 42–49.

Zhang, Y., Ding, Y., Li, H., Li, W., Ma, Y. & Ye, X. (2019), 'Retaking college entrance exam: Evidence from a regression discontinuity design', *Global Education* **48**(7), 111–128. (in Chinese).