Essays on instruction time, grades and parental investments in education

Collins, Matthew

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Collins, M. (2022). *Essays on instruction time, grades and parental investments in education*. [Doctoral Thesis (compilation), Lund University School of Economics and Management, LUSEM]. Lund University.

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Essays on instruction time, grades and parental investments in education

Matthew Collins

Lund Economic Studies
Number 229

LUND UNIVERSITY

Essays on instruction time, grades and parental investments in education

Essays on instruction time, grades and parental investments in education

by Matthew Collins

DOCTORAL DISSERTATION
By due permission of the School of Economics and Management, Lund University, Sweden.
To be defended at Holger Crafoords Ekonomicentrum EC3:210 on May 6, 2022 at 10am.

Thesis advisors: Therese Nilsson, Jan Bietenbeck
Faculty opponent: Helmut Rainer, University of Munich

Organization
LUND UNIVERSITY

Department of Economics
Box 7080
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2022-05-06

Sponsoring organization

Author(s)
Matthew Collins

Title and subtitle
Essays on instruction time, grades and parental investments in education

Abstract
This thesis consists of four self-contained papers in the economic of education. The first chapter examines the importance of instruction time for student achievement on international assessments. We successfully replicate the positive effect of weekly instruction time in the seminal paper by Lavy (*Economic Journal*, 125, F397-F424) in a narrow sense. Extending the analysis to other international assessments, we find effects that are consistently smaller in magnitude. We provide evidence that this discrepancy might be partly due to a different way of measuring instruction time in the data used in the original paper. Our results suggest that differences in instruction time are less important than previously thought for explaining international gaps in student achievement.

The second chapter identifies the causal effect of sibling gender on education and how this effect varies according to traditional inheritance customs. Using data from 27 sub-Saharan African countries, I find that boys who inherit their father's property experience no effect of sibling gender, while boys who do not inherit from their father experience a significant negative effect of having a brother. Having a brother has a small negative effect on the education of girls, regardless of inheritance customs. The effect of sibling gender converges after the introduction of laws guaranteeing that children inherit from their parents, suggesting that parents substitute between transferring property to their children and investing in their education.

The third chapter studies the effect of more informative feedback on student performance. Using data on the population of Swedish school children, we exploit a reform to the grading system in compulsory school which introduced a more granular grading scale and thus provided students with more informative feedback on their academic performance. Exploiting a difference-in-discontinuity research design, we find that students exposed to more informative grading were less likely to graduate from high school and from an academic high school track. The likelihood of a student graduating from a STEM high school track or enrolling in a STEM track in university also decreased as a result of more informative grading. These results appear to be driven by a negative shock to students' self-belief and increased stress levels.

The fourth chapter investigates the effect of university grade inflation on students' education and labour market outcomes. We exploit a series of reforms inducing grade inflation in English universities, using a staggered difference-in-differences strategy to identify the causal effect of grade inflation. Policies inducing grade inflation led to an increase in the proportion of students attaining first class honours and a decrease in the proportion obtaining the lowest final grades. We find that grade inflation reduces the likelihood of full-time employment but not the likelihood of being in any employment, while at the same time increasing the likelihood of students pursuing further studies six months after graduating. While we find no average effect on graduate salaries, we find that grade inflation led to a significant increase in the salary of graduates in the bottom decile of the earnings distribution and at the top five percentiles for male graduates.

Key words
instruction time; student achievement; PISA; TIMSS; sibling gender; patriliny; matriliny; educational attainment; grading; feedback; educational outcomes; natural experiment; HBSC; grade inflation; signalling; human capital

Classification system and/or index terms (if any)
JEL Classification: D13 I12 I20 I21 I23 I26 I28 J16 J24

Supplementary bibliographical information

Language
English

ISSN and key title
0460-0029 Lund Economic Studies no. 229

ISBN
978-91-8039-150-4 (print)
978-91-8039-149-8 (pdf)

Recipient's notes

Number of pages
273

Price

Security classification

Signature ______________________ Date 2022-03-07

Essays on instruction time, grades and parental investments in education

Matthew Collins

LUND UNIVERSITY

LUND ECONOMIC STUDIES NUMBER 229

School of Economics and Management, Department of Economics

ISBN: 978-91-8039-150-4 (print)
ISBN: 978-91-8039-149-8 (pdf)
ISSN: 0460-0029 Lund Economic Studies no. 229

Printed in Sweden by Media-Tryck, Lund University, Lund 2022

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

to Orla

Contents

Abstract

This thesis consists of four self-contained papers in the economic of education. The first chapter examines the importance of instruction time for student achievement on international assessments. We successfully replicate the positive effect of weekly instruction time in the seminal paper by Lavy (*Economic Journal*, 125, F397-F424) in a narrow sense. Extending the analysis to other international assessments, we find effects that are consistently smaller in magnitude. We provide evidence that this discrepancy might be partly due to a different way of measuring instruction time in the data used in the original paper. Our results suggest that differences in instruction time are less important than previously thought for explaining international gaps in student achievement.

The second chapter identifies the causal effect of sibling gender on education and how this effect varies according to traditional inheritance customs. Using data from 27 sub-Saharan African countries, I find that boys who inherit their father's property experience no effect of sibling gender, while boys who do not inherit from their father experience a significant negative effect of having a brother. Having a brother has a small negative effect on the education of girls, regardless of inheritance customs. The effect of sibling gender converges after the introduction of laws guaranteeing that children inherit from their parents, suggesting that parents substitute between transferring property to their children and investing in their education.

The third chapter studies the effect of more informative feedback on student performance. Using data on the population of Swedish school children, we exploit a reform to the grading system in compulsory school which introduced a more granular grading scale and thus provided students with more informative feedback on their academic performance. Exploiting a difference-in-discontinuity research design, we find that students exposed to more informative grading were less likely to graduate from high school and from an academic high school track. The likelihood of a student graduating from a STEM high school track or enrolling in a STEM track in university also decreased as a result of more informative grading. These results appear to be driven by a negative shock to students' self-belief and increased stress levels.

The fourth chapter investigates the effect of university grade inflation on students' education and labour market outcomes. We exploit a series of reforms inducing grade inflation in English universities, using a staggered difference-in-differences strategy to identify the causal effect of grade inflation. Policies inducing grade inflation led to an increase in the proportion of students attaining first class honours and a decrease in the proportion obtaining the lowest final grades. We find that grade inflation reduces the likelihood of full-time employment but not the likelihood of being in any employment, while at the same time increasing the likelihood of students pursuing further studies six months after graduating. While we find no average effect on graduate salaries, we find that grade inflation led to a significant increase in the salary of graduates in the bottom decile of the earnings distribution and at the top five percentiles for male graduates.

Keywords: instruction time; student achievement; PISA; TIMSS; sibling gender; patriliny; matriliny; educational attainment; grading; feedback; educational outcomes; natural experiment; HBSC; grade inflation; signalling; human capital

JEL Classifications: D13 I12 I20 I21 I23 I26 I28 J16 J24

Acknowledgements

This thesis is the result of a number of years of sometimes frustrating, sometimes difficult but almost always enjoyable and rewarding work. I am very lucky to have had the opportunity to study for a PhD and even more so to have had that opportunity in a country that values and rewards the work of PhD students. Although most of you will probably not read beyond this page, you should first know that this thesis, and the work that went into it, was helped by a large number of people.

I would like to first thank my supervisors, Therese Nilsson and Jan Bietenbeck. I have had the chance to co-author with both of you and benefited hugely as a result. Thanks for always being available for discussions, for the encouragement, patience and support you gave me and not least for helping to identify which of my research ideas were worth pursuing and which ones were not.

I have also benefited from feedback and insights from many other colleagues at the department. For this I would like to thank Alessandro, Alex, Ana, Elin, Gunes, Kaveh, Lina Maria, Mohammad, Petra, Petter and Roel. The PhD was a lot easier thanks to the excellent admin team so a big thanks goes to Anna, Azra, Jenny, Li, Marie, Peter and Ulf. Thanks also to the innebandy group, who welcomed me in to their games despite my more agricultural style of play.

The individual chapters of this thesis have undeniably improved thanks to the feedback of the external reviewers. For this, I am grateful to Natalie Bau, Nagore Iriberri and Georg Graetz.

The PhD programme in Lund has been truly a fantastic experience. A large part of this is due to Tommy and Erik's great work as directors of the programme, but even more so it is due to having excellent colleagues. My time in the PhD would not have been the same without my friend, office-mate and co-author, Jonas. I joke that I benefited from the relative rank effect, rather than the overall peer-effect of sharing an office with you for five years but there is no doubt that it was more of the latter. We were part of an excellent cohort, along with Demid, Marco, Ovi, Sandra and Zahra. Hopefully we can arrange some more cohort dinners before we get separated. Outside of my own cohort, spending so much time in EC was all the better thanks to some of my many other great colleagues. A special thanks goes to Devon, Linn, Sara, Erik and Hampus for the chats about everything from econometrics and machine learning to gossip and afterwork plans. A big thanks also goes to the long list of all my other PhD colleagues year the past few years: Jörgen, Anna, Sara, Hjördis, John, Kristoffer, Thomas, Pol, Polina, Sanna, Yana, Danial, Hoang, Adrian, Charlotta, Marcus, Philipp, Albert, Emelie, Prakriti, Steve, Yousef, August, Filipe, Kajsa, Ludvig, Negar, Teppo, David, Iker, James, Natalie and Shayan. You had to endure listening to me discuss my favourite documentaries, even if my choices didn't always tickle your fancy. This is a great place to work and you all contributed that over the last few years.

The last few months of the PhD were, however, dominated by the dreaded job market. For their help in preparing for the market and for being there each step of the way, I thank Aurélien, Luca, Marco, Marta, Ovi and Simon.

I am very happy that I got to participate in the econometrics game, let alone to go and win the whole thing. Thanks to Devon, Natalie and Ester for being great teammates and making two of the most intense and stressful days of my whole PhD surprisingly fun. Plus, Yana and Hampus for encouraging (pressuring) us to participate in the first place and for helping us to prepare.

I thank everyone at the ESRI, who introduced me to the world of research in the first place and made me believe a career as an academic was worth pursuing. Thanks to Pete and what was then known as the PRICE lab, to John and everyone in the energy group and to everyone else who made my two years there enormously fun, especially Brian, Bryan, Martin, Mel, Seraphim and Yota.

I would never have been in this position without the privilege and opportunities that I grew up with and for that I thank my parents. Finally, my utmost gratitude goes to Orla. As grateful as I am to have had the chance to do a PhD in the first place, I am ever more grateful that you were alongside me the whole time. This thesis is dedicated to you.

Lund, March 2022
Matthew

Introduction

Introduction

Instruction Time, Grades and Parental Investments in Education

This thesis consists of four self-contained papers covering different topics in the economics of education. The first chapter examines the importance of in-class instruction time for student achievement in international tests. As much research has shown that differences in educational achievement are important drivers of cross-country differences in economic growth, it is of importance to understand the inputs that can determine achievement.

The second chapter examines the effect of having a brother rather than a sister on education outcomes for sub-Saharan Africa and how that effect varies according to the inheritance customs of different ethnic groups. In many developing countries, boys continue to receive more education than girls, yet in sub-Saharan Africa, there exist diverging trends in the gender gap in education across countries. Identifying the cultural factors that drive these gender gaps within families can inform our understanding of how successful certain policies to improve education will be in different regions.

The third and fourth chapters deal with different aspects of grading in education. The third chapter explores whether more informative grades, by means of providing grading feedback on a more granular scale, can improve education outcomes for children in lower secondary school. The fourth chapter examines whether grade inflation at the university level affects the education outcomes of students and the labour market outcomes of graduates. Feedback is a powerful tool that can be used to motivate students and grading in particular can provide a signal of a student's productivity for employers. Despite the fact that feedback in the form of grades is widespread throughout all stages of education, there remain large gaps in our understanding of how the grading system can affect individuals' economic outcomes.

A second common theme across all four chapters is the methodology. All chapters aim to provide results that can inform policy. Doing so requires identifying causal relationships. For the research findings to have meaningful implications for policy requires identifying the effect of a specific treatment or intervention and not some spurious underlying relationship. Ideally, to identify the effect of some treatment on an outcome of interest, researchers would use what is called a randomised control trial, where individuals are randomly assigned to either receive or not receive some treatment. As this is often unfeasible when we want to study real-world behaviour and outcomes, research in economics has grown to identify so-called natural experiments.

In recent times, research in applied microeconomics has improved greatly thanks to better data and improvements in research design, specifically in regard to how we understand natural experiments. This has been referred to as the 'credibility revolution' and to underline how important this development has been to research in economics more generally, it is for their contributions to the development of applied microeconomic methods that Joshua Angrist, David Card and Guido Imbens were awarded the most recent Nobel Prize in Economics. By using the methods developed by these and others, we ae able to identify such causal effects more credibly than in the past.

The remainder of this introduction provides a short summary of each chapter, outlining the motivation behind each chapter, their data, empirical strategy and results, before outlining their main contributions.

Summary and contributions of the Thesis

Paper I: New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments

Student achievement on international assessments differs widely across countries and research shows that these achievement gaps are important drivers of cross-country differences in economic growth (Hanushek and Woessmann, 2012). This has spurred interest in the question of what explains international variation in student achievement, with one line of research focusing on the importance of instruction time.

In this paper, which is co-authored with Jan Bietenbeck, we examine the importance of instruction time for student achievement on international assessments.

Identifying the effect of instruction time can be challenging. Comparing students who have different amounts of instruction time will likely be confounded by other unobserved factors that are correlated with instruction time. For example, students who receive more instruction time may do so because they are enrolled in schools with greater financial resources, which can be used to improve education in more ways than just increasing instruction time. To avoid this issue, we use a student fixed effects identification strategy. This means we rely on variation in instruction time across subjects for students taking exams in more than one subject in PISA, assuming that the ability of students, their own characteristics and the school environment are the same for each subject except for differences in instruction time.

We first successfully replicate the results by Lavy (2015) in a narrow sense, using data from the Programme for International Student Assessment (PISA) in 2006. We then show that the effect of instruction time is also positive but smaller in data from five further waves of PISA: whereas Lavy (2015) estimates that a one-hour increase in weekly instruction time raises achievement by 0.058 standard deviations, the estimates for the other waves range from 0.014 to 0.031 standard deviations. Using data from six waves of the Trends in International Mathematics and Science Study (TIMSS), another international assessment of student competencies, we similarly find smaller effects ranging from 0.015 to 0.037 standard deviations. While we are unable to fully explain this discrepancy in results, we provide evidence that the original estimate might be larger partly due to a different way of measuring instruction time in PISA 2006.

The main contribution of this paper is to add to the growing literature on the effect of instruction time on student achievement. Several other related studies estimate the impact of instruction time on achievement using data from international studies and from individual countries. We provide comparable international evidence from many different datasets. By doing so, we show that differences in instruction time are less important than previously thought for explaining international gaps in student achievement.

Paper II: Sibling Gender, Inheritance Customs and Educational Attainment: Evidence from Matrilineal and Patrilineal Societies

Inequalities in the allocation of resources within households can occur because of differences in cognitive and health endowments but also due to other factors such as gender, expected inheritances, labour market prospects and cultural practices. According to classical models of human capital investment, parents will invest more in the education of children who they believe have a higher marginal return to schooling (Becker, 1981). If parents invest in their children as predicted, this will serve to exacerbate inequalities. Children cannot choose their endowments or early-life circumstances, and the family and society into which a child is born has ramifications for the allocation of resources among siblings.

In this paper, I study the effect of sibling gender on education and how that effect varies according to the inheritance customs of different ethnic groups. The gender composition of one's siblings is not random and stems from parental preferences over the number of children of each gender they wish to have. I circumvent this problem by exploiting the fact that for first-born children who have a second-born sibling, the gender of that sibling is as good as random. This allows me to identify the causal effect of having a second-born brother, relative to a second-born sister, on the education of the first-born child. In sub-Saharan Africa, there exist many different inheritance customs among different ethnic groups, allowing me to identify variation in the effect of sibling gender according to whether inheritance customs allow inheritances to be passed from fathers to sons or not.

For boys who can inherit property from their fathers, I find no effect of sibling gender on education. For boys who are not in line to receive an inheritance from their father, I find a negative effect of having a brother of 9.4% of a year of schooling. For girls, I find a small negative effect of having a brother, regardless of inheritance customs. I show that the introduction of new laws guaranteeing that children can inherit from their parents leads to convergence in the effect sibling gender across ethnic groups. This shows that parents view the transfer of property and investment in education as substitutes in investing in the future of their offspring. In addition, I show that free primary education reduces sibling gender effects in educational attainment for boys who do not inherit from their fathers, providing notable implications for the role of policy in reducing gender inequalities.

This paper provides a number of important contributions. First, it contributes to our understanding of how sibling gender affects education outcomes in less developed countries. Specifically, this paper provides a comprehensive examination of how sibling gender affects education in sub-Saharan Africa and the mechanisms through which it operates.

Second, my work provides important insights on how policy interacts with culture and traditions. I show that inheritance laws guaranteeing inheritance to children and the introduction of free primary education can shape the inequalities in the allocation of educational resources that stem from different inheritance customs.

Third, I contribute to our understanding of how kinship affects individual-level outcomes. By identifying how the effect of sibling gender varies according to inheritance customs, I show that the ability to transfer inheritance to children can be used to reduce inequalities in the allocation of resources. Previous research has shown that kinship affects various individual and social characteristics of individuals, including preferences toward competition, risk preferences, co-operation among spouses and health outcomes. My work adds to this body of evidence by showing how kinship-induced inheritance customs act as a mechanism determining intra-household educational investments.

Paper III: The Long-Term Consequences of More Informative Grading

In education, it is commonplace for students to receive feedback in the form of grades and there are many reasons why more informative grading could matter for an individual's educational outcomes. If people are overconfident, more informative grading may lead to reduced self-belief (Möbius et al., 2014), lower effort (Fischer and Sliwka, 2018) and increased dropout rates (Stinebrickner and Stinebrickner, 2014). On the other hand, more informative grading should provide students with more accurate information on their comparative advantages, helping students to find better matches in education and the labour market.

Studying the effect of more informative grading is challenging. Comparing school systems with different types of grades is not informative, as there is a myriad of institutional, cultural and demographic factors affecting outcomes that vary and may be difficult to account for. In this paper, which is co-authored with Jonas Lundstedt, we identify the causal effect of more informative grading. We exploit a natural experiment, where a reform to the compulsory school grading system in Sweden in 2011 led to the introduction of more informative grades.

The reform affected the information content of grades by replacing the previous scale, which included three passing grades, with a more granular scale with five passing grades. As children in Sweden are assigned to school cohorts based on their date of birth, we are able to exploit this reform in a difference-in-discontinuity research design. Children born just after January 1st, 1997 were exposed to more informative grading, while those born just before were not. Those born just after the admissions cut-off date are also subject to school-starting age effects which arise across the January 1st assignment cut-off in every year (Black et al., 2011; Fredriksson and Öckert, 2014). By comparing the difference in outcomes between those born just before and just after the admissions cut-off of January 1st, 1997 to the difference in outcomes of those born just before and just after the January 1st admissions cut-offs from 1992 to 1996, we are able to separate the confounding effect of school starting age and identify the effect of exposure to more informative grading.

We find that more informative grading has negative consequences for educational outcomes, with treated students being less likely to graduate from high school, academic high school and STEM high school tracks, in addition to being less likely to enrol in STEM university tracks. Using survey data from an internationally representative sample of school children we find that treated children are less likely to see themselves as performing well or very well in school, suggesting that these negative effects are driven by reduced confidence in one's own academic ability.

The main contribution of this paper is to add to our understanding of how absolute performance feedback can be leveraged to improve educational achievement. Previous research has examined whether receiving any feedback or whether different timing of feedback can improve performance. In this paper, we provide the first evidence on the effects of receiving more informative grades in school on longer-term educational and labour market outcomes.

This paper also contributes to the broader empirical literature on how grades can be leveraged to improve educational performance. More recently, this literature largely examines the benefits of relative grading systems, finding mixed results. We contribute to this literature by examining effects on individuals who are treated between the ages of 13 and 15, ages at which individuals' responses to treatment may be quite different to those treated at the high school and university levels.

Paper IV: The Effect of University Grade Inflation on Graduate Outcomes

An increasing share of university students are receiving better grades. The proportion of A grades awarded in US universities rose from 33% to 45% between 1988 and 2008 (Rojstaczer and Healy, 2012), while in the UK, the proportion of first class honours degrees awarded rose from 18% to 28% between 2014 and 2018 (HESA, 2018, 2019). Similar trends have been noted across other European countries like Germany (Müller-Benedict and Gaens, 2020) and Italy (Biancardi, 2017). Increasing proportions of good grades and degrees could be the result of better quality or better prepared students being admitted to university, improved teaching, technological advancements leading to improved learning or greater effort by students.

An alternative explanation, which has received growing attention and concern among policy makers and the media, is that universities have been grading students more leniently over time, resulting in a trend of grade inflation. In other words, higher grades and good degrees are being awarded without a corresponding improvement in students' abilities or achievements. While the occurrence of grade inflation in universities has been widely documented, we know little about whether grade inflation actually matters for education and labour market outcomes.

Comparing students in universities with more or less stringent grading requirements could be confounded if weaker students sort into more lenient universities. It is also difficult to disentangle whether rising grades are due to better technology, better teaching methods, more diligent students, or grade inflation. In this paper, which is co-authored with Judith Delaney and Therese Nilsson, we avoid these issues and separate the effects of grade inflation from other factors affecting grades. In England, universities typically award a final grade based on a 4 or 5 point scale. Universities have full autonomy over how final grades are calculated and many universities have reformed their degree algorithm in recent years, making it easier for students to achieve a better final grade than students with the same profile of course grades who graduated prior to a reform. We employ a difference-in-differences identification strategy, comparing students in universities that implemented grade inflation to universities that did not.

Our results show that grade inflation does increase the proportion of students graduating with top grades. We find that grade inflation had a significant impact on short-term labour market outcomes and educational choice. Specifically, grade inflation reduces the likelihood of full-time employment but not the likelihood of being in any employment, while at the same time increasing the likelihood of pursuing further studies six months after graduation. Grade inflation increases the salary of graduates in fields of study with higher levels of salary variation, who are more likely to be employed in the private sector.

Previous literature has identified how some of the specific mechanisms mentioned above determine outcomes, including signalling, student motivation, educational aspirations and employer hiring practices, among others. Yet it is unclear whether and to what extent the incidence of grade inflation itself affects outcomes. The main contribution of this work is therefore to identify the overall effect of grade inflation on education and labour market outcomes.

Despite the substantial rise in university grades and the accompanying public debate around grade inflation and grading standards, the effects of grade inflation in higher education is an empirical question that is not very well examined. Previous research has shown that grade inflation reduces student effort, increases enrolment in more lenient courses and improves students' teacher evaluations. By examining the effect of grade inflation among the population of university graduates, we provide evidence that grade inflation has a causal impact on further study and salaries in the short term.

Finally, we contribute to the understanding of grade inflation more generally. Several studies examine grade inflation at high school level but the effects of high school grade inflation on labour market outcomes tend to be indirect and operate through the impact of grade inflation on sorting into higher education. In comparison, we shed light on the direct relationship between university grade inflation and labour market outcomes.

References

Becker, G. S. (1981). *A Treatise on the Family*. Harvard university press.

Biancardi, D. (2017). Empirical essays in education and labor economics.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, 93(2):455–467.

Fischer, M. and Sliwka, D. (2018). Confidence in knowledge or confidence in the ability to learn: An experiment on the causal effects of beliefs on motivation. *Games and Economic Behavior*, 111:122–142.

Fredriksson, P. and Öckert, B. (2014). Life-cycle effects of age at school start. *The Economic Journal*, 124(579):977–1004.

Hanushek, E. A. and Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4):267–321.

HESA (2018). Higher education student statistics: UK, 2016/17 - qualifications achieved. Technical report, Higher Education Statistics Agency.

HESA (2019). Higher education student statistics: UK, 2017/18 - qualifications achieved. Technical report, Higher Education Statistics Agency.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal*, 125(588):F397–F424.

Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing self-confidence. *NBER Working Paper*.

Müller-Benedict, V. and Gaens, T. (2020). A new explanation for grade inflation: the long-term development of german university grades. *European Journal of Higher Education*, 10(2):181–201.

Rojstaczer, S. and Healy, C. (2012). Where a is ordinary: The evolution of american college and university grading, 1940-2009. *Teachers College Record*.

Stinebrickner, R. and Stinebrickner, T. (2014). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3):601–644.

Paper I

New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments

Co-authored with Jan Bietenbeck[1]
Accepted for Publication in the *Journal of Applied Econometrics*

Abstract

We re-examine the importance of instruction time for student achievement on international assessments. We successfully replicate the positive effect of weekly instruction time in the seminal paper by Lavy (*Economic Journal*, 125, F397-F424) in a narrow sense. Extending the analysis to other international assessments, we find effects that are consistently smaller in magnitude. We provide evidence that this discrepancy might be partly due to a different way of measuring instruction time in the data used in the original paper. Our results suggest that differences in instruction time are less important than previously thought for explaining international gaps in student achievement.

Keywords: instruction time; student achievement; PISA; TIMSS
JEL Classifications: I21

[1]Lund University, CESifo, DIW Berlin, IZA

1 Introduction

Student achievement on international assessments differs widely across countries, and research shows that these achievement gaps are important drivers of cross-country differences in economic growth (Hanushek and Woessmann, 2012). This has spurred interest in the question of what explains international variation in student achievement, with one line of research focusing on the importance of instruction time. In the seminal study in this literature, Lavy (2015) uses student-level data from the Programme for International Student Assessment (PISA) in 2006 to show that weekly instruction time positively affects achievement. Given large variation in weekly instruction time across countries, this suggests that international achievement gaps are partly due to differences in the amount of hours students spend learning in the classroom.

In this paper, we re-examine the importance of instruction time for student achievement on international assessments. We first successfully replicate the results by Lavy (2015) in a narrow sense, using the same student fixed-effects specification and data from PISA 2006 for a sample of OECD countries. We then show that the effect of instruction time is also positive but smaller in data from five further waves of PISA: whereas Lavy (2015) estimates that a one-hour increase in weekly instruction time raises achievement by 0.058 standard deviations (SD), the estimates for the other waves range from 0.014 SD to 0.031 SD. Using data from six waves of the Trends in International Mathematics and Science Study (TIMSS), another international assessment of student competencies, we similarly find smaller effects ranging from 0.015 SD to 0.037 SD. While we are unable to fully explain this discrepancy in results, we provide evidence that the original estimate might be larger partly due to a different way of measuring instruction time in PISA 2006.

In additional analyses, we extend our samples to further countries and show that the effect of instruction time is larger in high-income countries than in low- and middle-income countries, in line with results by Lavy (2015). We also conduct a range of sensitivity checks to gauge whether our estimates are confounded by unobserved factors that the student fixed-effects specification cannot account for. We find no evidence of such bias, but the non-experimental nature of the data does not allow us to completely rule out the influence of confounding unobservables.

Our paper adds to the growing literature on the effect of instruction time on student achievement. Besides the study by Lavy (2015), this research includes important work by Rivkin and Schiman (2015), who use data from PISA 2009 and two different identification strategies. They estimate impacts of between 0.023 SD and 0.031 SD per weekly hour of instruction. Our results suggest that the smaller magnitude of these estimates could be due to the inclusion of low- and middle-income countries in their sample, or due to the different measurement of instruction time in PISA 2009 compared to PISA 2006. Several other related studies estimate the impact of instruction time on achievement using data for individual countries, including Switzerland, Denmark, and Israel (Cattaneo et al., 2017; Bingley et al., 2018; Lavy, 2020). We contribute to this research by providing comparable international evidence from many different datasets.

2 Empirical strategy

PISA is an international repeated cross-sectional study that assesses the competencies of 15-year-old students in math, reading, and science. To estimate the causal effect of instruction time in the resulting individual-level data, Lavy (2015) exploits the fact that each student is observed in three subjects in the following student fixed effects specification:

$$A_{iks} = \beta WeeklyHours_{ks} + \mu_i + \eta_k + \varepsilon_{iks} \tag{1}$$

Here, i denotes students, k denotes subjects (math, reading, science), and s denotes schools. A_{iks} is the achievement of student i in subject k. $WeeklyHours_{ks}$ are the weekly hours of instruction received in subject k, measured at the school level. μ_i is a student fixed effect, which controls for all student-level determinants of achievement that do not vary across subjects, such as general academic ability. η_k is a subject fixed effect, which controls for any level differences in achievement across subjects. ε_{iks} is the error term. Lavy (2015) estimates this specification by ordinary least squares and computes standard errors that are robust to clustering at the school level.

The regression in Equation 1 identifies the effect of instruction time from differences between subjects. A causal interpretation of the coefficient of interest β relies on the key assumption that there are no other subject-specific determinants of achievement that are correlated with instruction time. We assess the validity of this assumption in Section 4.3 below.

3 PISA and TIMSS: background and data

3.1 PISA

PISA was first conducted by the OECD in 2000 and has since been repeated every three years. The number of countries participating in the study differs somewhat between waves, but it usually covers more than 50 developed and developing countries. In each wave, PISA draws nationally representative samples of 15-year-old students and assesses them on their math, reading, and science skills using standardized tests. The tests measure students' ability to use their knowledge of the subject to solve real-life problems. Test scores are scaled to have mean 500 and SD 100 across OECD countries participating in PISA 2000. Scores from other countries and later waves are then put onto the same scale, which makes achievement comparable across countries and over time.

Students participating in PISA are asked to complete a questionnaire which, among other things, asks about the weekly amount of school-based instruction time received in each subject. In the 2006 wave of the study, this information was gathered by asking students how much time they typically spend per week attending school lessons in each subject, with possible answers being "no time," "less than 2 hours," "2 or more but less than 4 hours," "4 or more but less than 6 hours," and "6 or more hours." In the other waves, students were instead asked open-ended questions about the number of lessons per week they have in a given subject and how long a typical lesson lasts. Table 1 shows the exact questions used to measure instruction time in each wave.

Our narrow replication uses data from PISA 2006 as in Lavy (2015). For our extension, we use data from five further waves of PISA conducted between 2000 and 2018. We do not analyze data from PISA 2003 because instruction time was measured only in math in that wave, such that we cannot identify its impact using between-subject differences. Following the original paper, we restrict our samples to students who are observed with achievement scores and who answered the questions on instruction time in all subjects, resulting in a balanced panel with three observations per student. The main analysis further restricts attention to a group of 22 OECD developed countries.[1]

The key independent variable in our regressions measures the weekly hours of

[1] The 22 OECD developed countries included in the main sample in Lavy (2015) and our replication are: Australia, Austria, Belgium, Canada, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Iceland, Italy, Japan, Luxembourg, Netherlands, Norway, New Zealand, Portugal, Sweden, Switzerland, United Kingdom.

school-based instruction received in a given subject. In the data from PISA 2006, we follow Lavy (2015) and transform the categorical answers into continuous hours by recoding "no time" to missing, "less than 2 hours a week" to 1 hour, "2 or more but less than 4 hours" to 2.5 hours, "4 or more but less than 6 hours" to 4.5 hours, and "6 or more hours" to 6 hours.[2] In the data from the other PISA waves, we multiply the number of lessons by the number of minutes per lesson and divide by 60. For all waves, we then average instruction time at the school-by-subject level.

The outcome variable in our regressions is the subject-specific test score. As in the original paper, we transform raw scores into z-scores by subtracting 500 and dividing by 100. In this way, we can interpret the estimated effects in terms of standard deviations of the test score distribution among OECD countries participating in the first PISA assessment in 2000.[3]

3.2 TIMSS

TIMSS is an international assessment of the math and science knowledge of fourth- and eighth-grade students. The study has been conducted by the International Association for the Evaluation of Educational Achievement every four years since 1995 and usually covers more than 40 countries. In each country, nationally representative samples of students are assessed using standardized tests, which measure students' knowledge of the common part of the math and science curricula of participating countries. Test scores are scaled to have mean 500 and SD 100 across countries in TIMSS 1995, with scores from later waves put onto the same scale. TIMSS also asks all math and science teachers of participating students to complete a questionnaire, which collects information on how many minutes per week they teach their subject to the students' class, among other things. Table 1 shows the exact wording of the questions about instruction time fielded in each wave.

TIMSS shares some key features of PISA, such as the international scope and the focus on more than one subject, which notably allows us to estimate student

[2]In his paper, Lavy (2015) writes that he merges the "no time" and "less than 2 hours a week" categories, but the publicly available code on the journal website actually changes "no time" to missing. We choose to follow the code in order to replicate exactly the original estimates, but in practice this makes very little difference.

[3]Both PISA and TIMSS use Item Response Theory to score tests and report scores as a set of five or ten so-called plausible values. In our analysis, we follow Lavy (2015) and use the first plausible value for each student in each subject as our outcome. We checked that all of our results are insensitive to choice of plausible value.

Table 1: Questions used to measure school-based instruction time in PISA and TIMSS

Study	Respondent	Question	Type
PISA 2000	Student	In the last full week you were in school, how many <class periods> did you spend in <subject>?	open-ended
	Principal	How many instructional minutes are there in the average single <class period>?[a]	open-ended
PISA 2006	Student	How much time do you typically spend per week studying the following subjects? Regular lessons in <subject> at my school:	categorical[b]
PISA 2009 and 2012	Student	How many <class periods> per week do you typically have for the following subjects?	open-ended
	Student	How many minutes, on average, are there in a <class period> for the following subjects?	open-ended
PISA 2015 and 2018	Student	How many <class periods> per week are you typically required to attend for the following subjects?	open-ended
	Student	How many minutes, on average, are there in a <class period>?[a]	open-ended
TIMSS 1995 and 1999	Teacher	How many minutes per week do you teach <subject> to your <subject> class?	open-ended
TIMSS 2003 and 2007	Teacher	How many minutes per week do you teach <subject> to the TIMSS class?	open-ended
TIMSS 2011 and 2015	Teacher	In a typical week, how much time to you spend teaching <subject> to the students in this class?	open-ended

Notes: The table gives an overview of the questions used to measure school-based instruction time in each wave of PISA and TIMSS. The term "<class period>" is translated to the locally used term in each country. The term "<subject>" is replaced by "mathematics", "science," or "reading" (only in PISA) for the corresponding subject-specific question. In TIMSS, teachers' answers always refer to the class of students participating in the assessment. [a]This question is not asked separately for each subject. [b]Answer categories for this question: no time, less than 2 hours a week, 2 or more but less than 4 hours a week, 4 or more but less than 6 hours a week, 6 or more hours a week.

fixed effects models like in Equation 1. However, there are also important differences between the two assessments. Thus, TIMSS tests curriculum knowledge rather than problem-solving skills and does not cover reading. It also focuses on students in specific grades, who do not correspond exactly to the population of 15-year olds surveyed in PISA (eighth-grade students are about 13.5 years old on average). Moreover, only a subset of the 22 OECD countries examined in Lavy (2015) participated in each TIMSS wave. If the effect of instruction time varies along any of these dimensions, we can expect estimates to differ between the two assessments. By examining data from both PISA and TIMSS, we can gain a more general understanding of the importance of instruction time for student achievement.

Our analysis uses data from six waves of TIMSS conducted between 1995 and 2015. We focus on eighth-grade students and construct our data in a way that closely follows Lavy (2015). Specifically, we restrict the sample to students observed with achievement scores and instruction time in both subjects, and we keep only those countries that are included among the 22 OECD countries described above. To measure instruction time, we first assign each student the total hours received in each subject as reported by her teachers and then compute the school-by-subject average. We measure achievement using the subject-specific test scores, which we transform into z-scores by subtracting 500 and dividing by 100. The estimated effects are thus scaled in terms of standard deviations of the test score distribution among countries participating in TIMSS 1995.

4 Results

4.1 Main results

Table 2 presents our main results. Column 1 of Panel A shows the effect of instruction time on student achievement in the PISA 2006 data used by Lavy (2015). The results indicate that a one-hour increase in weekly instruction time raises achievement by 0.058 SD. This estimate is exactly identical to the one reported in the original paper and thus constitutes a successful replication in a narrow sense. Columns 2 to 6 of Panel A show results for five further waves of PISA. The impact of instruction time in these samples is also positive but smaller in magnitude, with estimates ranging from 0.014 SD to 0.031 SD. Panel B reports estimates from six waves of TIMSS, which similarly range from 0.015 SD to 0.037 SD. Thus, effects in TIMSS are comparable to those in PISA (except for PISA 2006) despite the differences in sample and test design. However, note that even among these smaller estimates, the effect of instruction time still varies by a factor of more than two.

Table 2: Effect of instruction time on student achievement in OECD developed countries

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: PISA data						
	Orig. data: PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
Weekly hours	0.058	0.019	0.027	0.031	0.020	0.014
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
# of schools	6,577	4,352	7,176	7,774	6,204	6,070
# of countries	22	21	22	22	22	21
Mean weekly hours	3.4	3.8	3.7	3.7	3.8	3.9
Panel B: TIMSS data						
	TIMSS 1995	TIMSS 1999	TIMSS 2003	TIMSS 2007	TIMSS 2011	TIMSS 2015
Weekly hours	0.037	0.037	0.024	0.015	0.019	0.017
	(0.006)	(0.009)	(0.007)	(0.004)	(0.007)	(0.006)
# of observations	83,200	43,036	46,840	41,134	48,322	81,092
# of students	41,600	21,518	23,420	20,567	24,161	40,546
# of schools	1,770	949	915	804	918	1,324
# of countries	16	6	6	5	6	8
Mean weekly hours	3.3	3.2	3.1	3.0	3.3	3.3

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement in a sample of 22 OECD developed countries. Panel A shows results based on PISA data. Column 1 reports estimates from the 2006 wave of PISA used in Lavy (2015) and columns 2-6 report estimates from five further waves of PISA. Column 2 covers only 21 countries because data on instruction time are not available for Norway in PISA 2000 and column 6 covers only 21 countries because data for reading achievement are not available for Spain in PISA 2018. Panel B shows estimates based on six waves of TIMSS data. The samples in these regressions include the subset of the 22 OECD countries which participated in the corresponding wave of TIMSS. All specifications in both panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

4.2 Investigating differences in estimates between samples

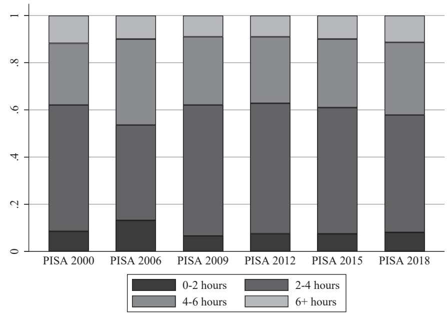
We now explore why the estimates in Table 2 differ so much between assessments. We focus mostly on the differences between PISA waves and consider two possible explanations: heterogeneous treatment effects and the measurement of instruction time. Heterogeneous treatment effects, for example by student characteristics, could account for the divergent results if samples differed between assessments. We discuss this possibility in detail in Appendix B. We show that there are only small differences in student characteristics between PISA samples, and we argue that any change in other, including unobservable, sample characteristics is likely gradual over time. We conclude that heterogeneous treatment effects are unlikely to be the main explanation for the differences in estimates between assessments, and especially for the much larger estimate in PISA 2006.

Another possible explanation for the divergent estimates is that the measurement of instruction time varies across assessments. This possibility is closely linked to the variation in the format of instruction-time questions shown in Table 1. As noted before, PISA 2006 is the only assessment in which students reported instruction time in broad categories of hours. There are also more minor differences in question format among the other assessments, which measure instruction time in minutes. Interestingly, there appears to be an association between question format and effect size: PISA 2006 with its categorical measurement produces by far the largest point estimate; PISA 2009 and PISA 2012 use identical questions and produce similar point estimates, and the same is true for PISA 2015 and PISA 2018 and three pairs of adjacent TIMSS waves.

One way in which question format could affect estimates is by changing the actual answers given by students, teachers, and principals. While we do not observe how the same respondents report instruction time under different question formats, we show that the categorical format used in PISA 2006 is associated with markedly different response patterns compared to the other PISA waves.[4] In Figure 1, we graph the share of reported instruction time falling into each of the PISA 2006 categories separately for each wave. Ignoring PISA 2006, about half the students report instruction time in the 2–4-hours category, with only little variation across waves. In contrast, only about 40% of answers fall into this category in PISA 2006. In theory, this difference could reflect an actual change in the distribution of instruction time in 2006. However, Figure A.1 shows that hours distributions before and after 2006 are similar, making this explanation unlikely. Instead, the results suggest that the different question format in PISA 2006 influenced students' answers. Since we do not observe exactly how answers were affected, however, we are unable to establish whether changed response patterns can account for the much larger estimate in that wave.

Another way in which the categorical question format in PISA 2006 could bias the estimate is by introducing an aggregation problem. In particular, the format requires researchers to impute an hours value for each category in order to arrive at a continuous measure, with Lavy (2015) using category mid-points. If the

[4] The different distribution of student answers in PISA 2006 was also noted by Rivkin and Schiman (2015).



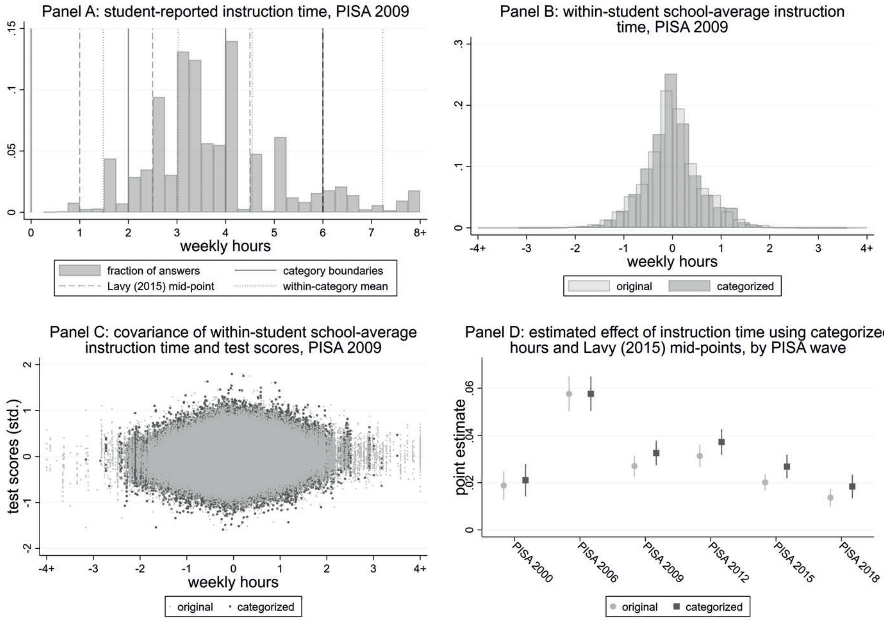
Notes: The figure shows the share of student-reported instruction time falling into each of the categories used in the PISA 2006 student questionnaire, separately for each PISA wave.

Figure 1: Distribution of student-reported instruction time across PISA 2006 categories

effect of instruction time is linear, aggregation of hours to the within-category mean does not affect the estimate, which implies that this imputation does not matter as long as the mid-point values are equal to the within-category averages. However, if mid-points and averages differ, the estimated effect of hours could be biased either upward or downward.

We illustrate this problem by artificially imposing the answer categories from PISA 2006 onto the hours distribution in PISA 2009. Panel A of Figure 2 shows the distribution of student-reported instruction time together with category boundaries, within-category means, and the Lavy (2015) mid-points. As can be seen, most mid-points differ substantially from the corresponding within-category mean. Panels B and C show that using these mid-points to aggregate answers leads to a reduction in the variance of within-student school-average instruction time, whereas the covariance between within-student test scores and school-average instruction time does not appear to change much. Put differently, the aggregation reduces the variance of the explanatory variable without markedly decreasing its covariance with the dependent variable, leading to an upward bias in the estimate. Panel D quantifies this bias and shows that the estimated effect increases by 20 percent in PISA 2009 and by between 12 and 34 percent in the other PISA waves if the PISA 2006 categories and Lavy (2015) mid-points are artificially imposed there.

The results in Figure 2 suggest that aggregation using mid-points can account for part of the difference in estimates between PISA 2006 and the other waves. An implication is that using correct within-category averages should reduce the PISA 2006 estimate. While the hours distribution in that wave is unobserved, we can impute within-category averages using the distributions from the five other



Notes: Panels A, B, and C of the figure show the effects of artificially discretizing instruction time in PISA 2009 by imposing the PISA 2006 answer categories and Lavy (2015) mid-points. Panel A shows the distribution of student-reported instruction time across subjects together with category boundaries, within-category mean hours, and mid-points. Panel B shows the distributions of subject-specific school-average instruction time for the original and categorized variables after residualizing on student and subject fixed effects. Panel C shows a scatter plot of test scores and subject-specific school-average instruction time, both of which have been residualized on student and subject fixed effects, for both variables. Panel D shows how imposing the categorical measurement and Lavy (2015) mid-points affects point estimates in each PISA wave. For details on how the differences between the main estimates and the estimates based on categorized hours in this panel materialize, see Appendix Table A.1.

Figure 2: Effects of imposing categorical measurement of instruction time

PISA waves. Doing this reduces the point estimate from 0.058 SD to 0.047 SD.[5] Taken together, our results suggest that the different measurement of instruction time in PISA 2006 at least partly explains the much larger estimated effect of instruction time in that wave.

[5]Note that this imputation cannot account for the changed response pattern and possible changes in the actual distribution of hours in PISA 2006. This could explain why the estimate using the imputed values is still higher than the estimates for the other PISA waves. For the imputation, we used all waves other than 2006 and computed within-category average hours separately by country and subject.

4.3 Evidence on the validity of the empirical strategy

A potential concern with our results is that they may be biased by subject-specific confounders. For example, if students who are especially gifted in math selected into schools offering relatively more math hours, our estimates would overstate the effect of instruction time. Lavy (2015) conducts a variety of sensitivity checks to test for such bias. For example, he restricts the sample to schools that do not track students by ability, with the intuition being that these schools should be less likely to admit students based on subject-specific academic ability. He finds that his estimates are broadly robust to this and various other sensitivity checks. We successfully replicated his results using the PISA 2006 data and ran comparable regressions for the five other waves of PISA.

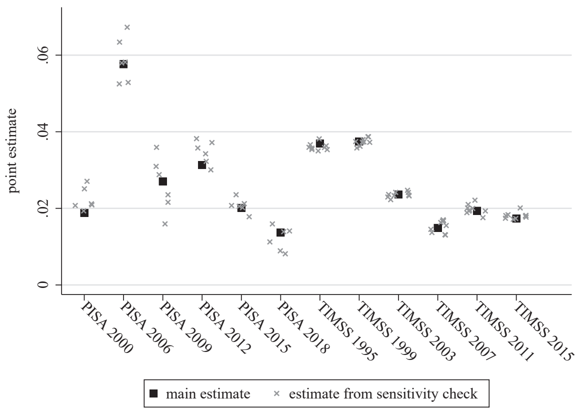
TIMSS collects detailed subject-specific information, which allows us to conduct additional sensitivity checks. Among other things, we observe whether a school offers subject-specific enrichment activities and a proxy for parents' perceived importance of each subject. We included each of these variables as a control in a separate regression. If subject-specific confounders were driving our results, we would expect the estimated effect of hours to change with the inclusion of these controls.

Figure 3 summarizes the results from our sensitivity analysis (full details are presented in Appendix C): it plots, separately for each wave of PISA and TIMSS, the main estimate from Table 2 and the estimates from the corresponding sensitivity checks. As can be seen, sensitivity estimates tend to cluster closely around the main estimates, which suggests that subject-specific confounders do not bias our results much. This finding is in line with previous results by Rivkin and Schiman (2015), who also provide evidence that bias due to any such confounders is likely to be small in magnitude.

4.4 Results for further countries

An interesting question is whether the effect of instruction time differs between developed and developing countries. In the original paper, Lavy (2015) shows that the impact is of similar magnitude to the one found for OECD developed countries in a sample of 14 Eastern European countries, but that it is only about half as large in a sample of 13 developing countries. We successfully replicated these results in a narrow sense and estimated equivalent specifications for the five other PISA waves.

We also estimated the effect separately for high-income and non-high-income



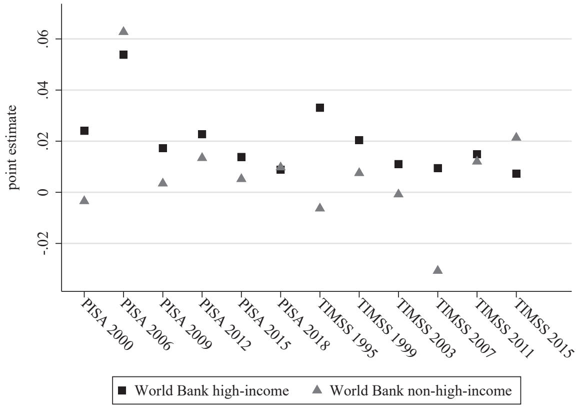
Notes: The figure shows point estimates of the effect of instruction time on student achievement. The black square reproduces the main estimate from Table 2 for the assessment indicated on the horizontal axis. The grey crosses depict point estimates from the corresponding sensitivity checks. See Appendix C for further details.

Figure 3: Estimates from sensitivity checks

economies as classified by the World Bank. This more comprehensive classification, together with the greater coverage across our 12 assessments, means that our results include a total of 59 high-income and 49 non-high-income economies. Figure 4 summarizes our findings (detailed estimation results are shown in Appendix Table A.2). It reveals that the effect of instruction time tends to be larger in high-income economies than in non-high-income economies, in line with the conclusion by Lavy (2015). While studying the exact reasons for this difference is beyond the scope of our paper, one potential explanation is that teachers in developing countries are frequently absent from the classroom, reducing actual instruction time (Chaudhury et al., 2006; Bold et al., 2017).

5 Conclusion

We re-examine the importance of instruction time for student achievement on international assessments. We successfully replicate the estimate of a positive effect of weekly instruction time in the seminal study by Lavy (2015) in a narrow sense. However, when we extend the analysis to data from 11 other international assessments, we find effects that are consistently smaller in magnitude than those reported in the original paper. We show that this discrepancy might be partly due to the different measurement of instruction time in the PISA 2006 data used



Notes: The figure shows point estimates of the effect of instruction time on student achievement separately for high-income and non-high-income economies as classified by the World Bank. For further details and additional estimates, see Appendix Table A.2.

Figure 4: Estimates for high-income and non-high-income economies

by Lavy (2015).

Our results suggest that the true effect of instruction time on student achievement is smaller than previously thought. However, some uncertainty about the exact magnitude of the effect remains. One reason is that our smaller estimates still vary by a factor of more than two. Another reason is that the identification strategy relies on the assumption that there are no subject-specific confounders. However, we provide new evidence which suggests that such confounders do not bias our estimates much.

References

Bingley, P., Heinesen, E., Krassel, K. F., and Kristensen, N. (2018). The Timing of Instruction Time: Accumulated Hours, Timing and Pupil Achievement. IZA Discussion Paper No. 11807.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., and Wane, W. (2017). Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa. *Journal of Economic Perspectives*, 31(4):185–204.

Cattaneo, M. A., Oggenfuss, C., and Wolter, S. C. (2017). The More, the

Better? The Impact of Instructional Time on Student Performance. *Education Economics*, 25(5):433–445.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, 20(1):91–116.

Hanushek, E. A. and Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4):267–321.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal*, 125(588):F397–F424.

Lavy, V. (2020). Expanding School Resources and Increasing Time on Task: Effects on Students' Academic and Noncognitive Outcomes. *Journal of the European Economic Association*, 18(1):232–265.

Rivkin, S. G. and Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal*, 125(588):F425–F448.

Appendix A: Additional results

Table A1: Effects of imposing categorical measurement of instruction time in PISA waves other than 2006

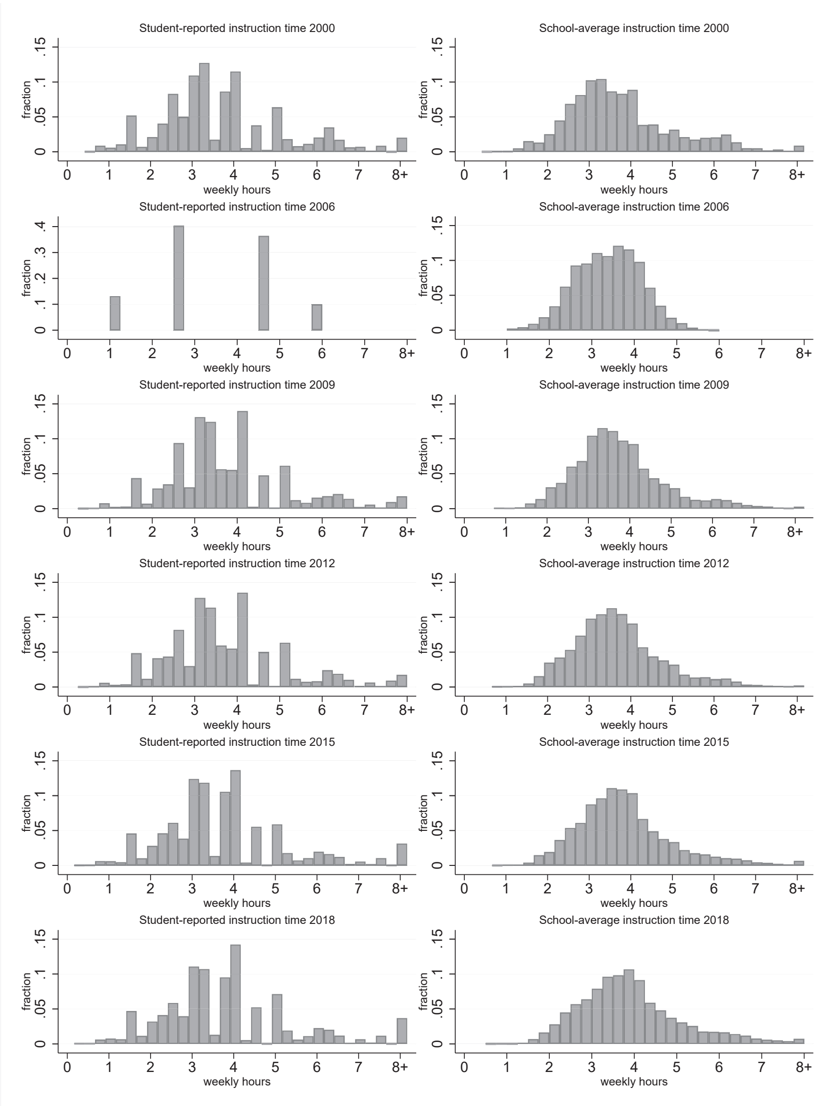
Wave	Variable Construction	Residual Variance	Residual Variance Factor	Residual Covariance	Covariance Factor	Point Estimate	Point Estimate Factor
2000	Original	0.5165	0.7806	0.0090	0.8750	0.0174	1.1210
	Discretized	0.4032		0.0079		0.0195	
2009	Original	0.3952	0.8219	0.0104	0.9892	0.0264	1.2036
	Discretized	0.3248		0.0103		0.0318	
2012	Original	0.4179	0.7632	0.0133	0.9073	0.0318	1.1889
	Discretized	0.3189		0.0121		0.0378	
2015	Original	0.5541	0.5638	0.0111	0.7517	0.0201	1.3333
	Discretized	0.3124		0.0084		0.0268	
2018	Original	0.4345	0.5818	0.0059	0.7835	0.0135	1.3467
	Discretized	0.2528		0.0046		0.0182	

Notes: The table shows how artificially discretizing instruction time in PISA waves other than 2006 affects the variance of the explanatory variable (column 3), the covariance between the dependent and explanatory variables (column 5) and, in turn, the point estimate in our main regression (column 7). Instruction time is discretized by imposing the answer categories used in PISA 2006 and the mid-points used in Lavy (2015). Instruction time and test scores are residualized on student and subject fixed effects. Factors reflect the magnitude of the residual variance (column 4), residual covariance (column 6), and point estimate (column 8) when using discretized instruction time, relative to undiscretized instruction time. For further details, see Figure 2 in the main text.

Table A2: Estimates for different groups of countries

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: PISA data						
	Orig. data: PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
A.1: Lavy (2015) 14 Eastern European countries						
Weekly hours	0.061	0.023	0.004	0.018	0.005	0.009
	(0.006)	(0.007)	(0.004)	(0.004)	(0.004)	(0.003)
# of students	59,005	6,416	61,147	39,062	62,932	64,984
# of countries	14	7	14	14	12	14
A.2: Lavy (2015) 13 developing countries						
Weekly hours	0.030	0.004	0.003	–0.005	0.005	0.033
	(0.008)	(0.006)	(0.003)	(0.003)	(0.004)	(0.003)
# of students	79,646	5,501	100,371	53,458	60,069	57,170
# of countries	13	6	13	11	8	10
A.3: World Bank high-income economies						
Weekly hours	0.054	0.024	0.017	0.023	0.014	0.009
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of students	227,445	28,767	273,032	169,342	256,392	241,117
# of countries	40	31	47	43	42	46
A.4: World Bank non-high-income economies						
Weekly hours	0.063	–0.003	0.003	0.013	0.005	0.010
	(0.007)	(0.005)	(0.002)	(0.003)	(0.003)	(0.002)
# of students	91,457	9,997	144,201	76,466	79,796	138,710
# of countries	16	11	26	20	12	26
Panel B: TIMSS data						
	TIMSS 1995	TIMSS 1999	TIMSS 2003	TIMSS 2007	TIMSS 2011	TIMSS 2015
B.1: World Bank high-income economies						
Weekly hours	0.033	0.020	0.011	0.009	0.015	0.007
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
# of students	79,714	81,722	97,871	95,659	83,833	127,555
# of countries	34	22	26	25	18	27
B.2: World Bank non-high-income economies						
Weekly hours	–0.006	0.008	–0.001	–0.031	0.012	0.021
	(0.011)	(0.004)	(0.004)	(0.003)	(0.009)	(0.005)
# of students	10,024	69,039	81,327	101,358	51,527	80,610
# of countries	4	15	21	25	9	13

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement separately for different groups of countries. Following Lavy (2015), Panel A.1 restricts the sample to 14 Eastern European countries and Panel A.2 restricts the sample to 13 developing countries. The number of countries is lower in some columns because not all countries participated in all rounds of PISA. Panels A.3 and A.4 (based on PISA data) and Panels B.1 and B.2 (based on TIMSS data) show results for high-income and non-high-income economies as classified by the World Bank as of June 2020. The number of countries included in these regressions varies across samples because not all countries participated in all assessments. All specifications in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.



Notes: The figure shows the distribution of student-reported instruction time (left panel) and school-average instruction time (right panel) separately for each PISA wave.

Figure A1: Distribution of instruction time by PISA wave

Appendix B: Differences in estimates due to heterogeneous effects?

In Section 4.2, we mention the possibility that estimates differ across waves due to heterogeneous treatment effects. In this Appendix, we discuss this possibility in more detail. Due to the inherent differences in design between PISA and TIMSS, we concentrate on differences in estimates between PISA waves, with a special focus on the PISA 2006 estimate.

One important dimension of heterogeneity in the effect of instruction time is student background. For example, Lavy (2015) shows that the impact of hours is larger for students with an immigrant background and for students with less educated parents. Similarly, Bingley et al. (2018) find that the effect varies by students' gender and socioeconomic status. If student background differed between samples, this heterogeneity could explain the differences in estimated effects. To explore this possibility, panel A of Appendix Table B1 shows means of students' socio-demographic characteristics separately for each PISA wave. All samples are balanced on gender, but immigration status and parental education trend upwards over time. However, these smooth trends cannot account for the much larger estimate in PISA 2006 compared to all other waves.

The effect of instruction time likely also differs by other, unobserved dimensions of student background. Moreover, it varies with school and class characteristics: for example, Rivkin and Schiman (2015) show that the effect differs by classroom quality. While we cannot determine whether the PISA samples are comparable on all possible dimensions of effect heterogeneity, any changes in such characteristics likely follow similarly smooth time trends as the characteristics observed in panel A of Appendix Table B1 and as such cannot account for the much larger estimate in PISA 2006 compared to the other waves.

A related alternative explanation for the differences in estimates is that the distribution of achievement changes across waves: even in the absence of heterogeneous treatment effects, if the standard deviation of achievement was much larger in PISA 2006, this could explain the higher estimate for this wave. However, panel B of Appendix Table B1 shows that means and standard deviations of test scores are broadly similar across waves, making this explanation unlikely.

Finally, non-linearities in the effect of instruction time could be at play if the distribution of hours changed between waves. In Appendix Figure A1, we show that hours distributions in PISA waves other than 2006 are comparable, and we argue in the main text that the true distribution of instruction time in PISA

2006 likely looks similar. However, due to the very different measurement of instruction time, the observed distribution of hours in that year differs from those in the other years. This means that we unfortunately cannot establish to what extent non-linearities in the effect of instruction time can account for the larger estimate in PISA 2006.

Appendix Table B1: Summary statistics of students' socio-demographic characteristics and achievement by PISA wave

	Orig. data					
	PISA 2006 (1)	PISA 2000 (2)	PISA 2009 (3)	PISA 2012 (4)	PISA 2015 (5)	PISA 2018 (6)
Panel A: means of socio-demographic characteristics						
Female	0.51	0.51	0.50	0.50	0.50	0.51
First-generation immigrant	0.05	0.05	0.06	0.07	0.07	0.08
Second-generation immigrant	0.05	0.05	0.06	0.07	0.08	0.10
Father has college education	0.24	–[a]	0.26	0.28	0.33	0.36
Mother has college education	0.22	–[a]	0.25	0.28	0.34	0.39
Panel B: mean and standard deviation of achievement						
Mean	513.42	521.62	509.76	513.17	509.38	510.34
Standard deviation	93.28	96.12	92.69	90.92	92.56	93.20

Notes: The table shows means of students' socio-demographic characteristics (Panel A) and the mean and standard deviation of student achievement (Panel B) separately by PISA wave as indicated in the column headers. Statistics for each wave are computed for the students included in the estimation sample of 22 OECD countries. [a]Data on parental education in PISA 2000 are not directly comparable to data in the other waves because of a change in the PISA student questionnaire after this wave.

Appendix C: Details on sensitivity checks

Section 4.3 summarizes results from sensitivity checks that gauge the extent of bias in our estimates due to subject-specific confounders. In this Appendix, we present additional details on these checks.

Checks in the PISA data

Our analysis for the PISA data closely follows Lavy (2015) and comprises five sensitivity checks, the results of which are shown in Appendix Table C1. First, we restrict the sample to schools that do not consider students' academic record in the admission process.[6] Intuitively, such schools should be less likely to select students based on subject-specific academic ability, reducing the potential for bias. Panel B presents the corresponding estimates (Panel A reproduces the main estimates from Panel A of Table 2 to facilitate comparison). Second, based on similar reasoning, we restrict the sample to schools that do not consider students' needs or desire for a particular program as a criterion for admission. The results for this check are presented in Panel C.

Third, we restrict the sample to schools that do not practice tracking (in any subject) between or within classes. The intuition is that schools that practice tracking will be more likely to admit students based on subject-specific academic ability. These schools could also place higher-ability students in classes with more instruction time. Panel D shows the results for the sample excluding these schools. Fourth, Panel E presents estimates for the subsample of public schools, for which subject-specific sorting is less of a concern according to Lavy (2015). Finally, we use information on teacher shortages in each subject. For example, schools that are mathematics-oriented might attract more effective math teachers, which could confound the estimates. Panel F presents estimates from regressions which control for an indicator for a lack of qualified teachers in a subject.

Overall, Appendix Table C1 shows that the estimated effect of instruction time is quite similar across the different specifications within a given PISA wave, which suggests that subject-specific confounders do not bias our results. However, one caveat of these checks is that they mostly rely on information that is not subject-specific, and that they therefore might not fully capture the influence of potential subject-specific confounders. As we describe below, the TIMSS data allows us to partially address this concern.

[6] Information on factors considered in the admission process was collected in all PISA waves, but the format of the question posed to principals changed somewhat over time. Question formats also changed for some of the other variables used in our analysis, and in a few cases the question was not asked at all. Whenever information is available, we define our variables such that they most closely resemble the original variables used by Lavy (2015).

Checks in the TIMSS data

In the TIMSS data, we use detailed background information from surveys to identify potential subject-specific confounders related to schools, teachers, and students. Appendix Table C2 presents the results of our sensitivity checks based on these variables. Note that not all variables are available in all waves. Moreover, the format of the underlying survey questions sometimes changes between waves; in these cases, we define variables as consistently as possible across waves.

Starting with school-related confounders, Panel B shows estimates from regressions in which the sample is restricted to schools that do not use students' academic record in the admission process (Panel A reproduces the main estimates from Panel B of Table 2 for convenience). This sensitivity check is equivalent to one of the checks conducted by Lavy (2015). Panel C shows estimates from specifications that add a control for subject-specific tracking by ability. Intuitively, such tracking could influence school choice and could also be related to instruction time, which in turn could lead to bias in our results. Panel D adds a control for whether the school offers subject-specific enrichment activities and Panel E adds a control for subject-specific remedial teaching. Such special teaching activities are likely to attract students with particularly high or low subject-specific ability, and they might also be related to instruction time. The results in Panels B to E show that our estimates from these checks are virtually identical to our main estimates.

Moving on to teacher-related confounders, Panel F adds a control for whether there is a shortage of teachers in a subject at the school, and Panel G adds a control for whether the school has had difficulties filling open teaching positions in a subject. These controls capture a lack of qualified teachers, which could be related to instruction time and also affect student achievement. Building on this same intuition, Panel H shows results from regressions that control for two observable dimensions of teacher quality: education, measured as an indicator for whether the teacher holds an advanced degree, and experience, measured as an indicator for whether the teacher has been teaching for at least five years. Our estimates in panels F to H are robust to these checks.

Finally, we estimate two specifications that add controls for subject-specific confounders related to students and parents. Panel I uses the fact that in two waves of TIMSS, students were asked to what extent their mother thinks that it is important for them to do well in each subject. This variable proxies for parents' valuation or preferences over subjects and intuitively relates to subject-specific sorting to schools. The specifications in Panel I add this variable as a control to our regressions. In Panel F, we control instead for an indicator for whether a student receives out-of-school extra lessons in a subject. This variable similarly proxies for parents' or students' subject-specific abilities and preferences. The results show that our estimates are robust to both of these sensitivity checks.

Taken together, the estimates in Appendix Table C2 show no indication of bias due to subject-specific confounders related to schools, teachers, or students. However, as we emphasize in the main text, we cannot control for all potential confounders and the key identification assumption in our empirical model is ultimately untestable.

Appendix Table C1: Sensitivity checks in PISA

	Orig. data					
	PISA 2006 (1)	PISA 2000 (2)	PISA 2009 (3)	PISA 2012 (4)	PISA 2015 (5)	PISA 2018 (6)
Panel A: main estimates (for comparison)						
Weekly hours	0.058	0.019	0.027	0.031	0.020	0.014
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
Panel B: academic record not considered for school admission						
Weekly hours	0.060	0.017	0.025	0.034	0.022	0.013
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
# of observations	266,769	29,799	265,005	146,370	162,897	171,039
Panel C: students' needs or desire not considered for school admission						
Weekly hours	0.066	0.027	0.035	0.037	0.023	0.015
	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
# of observations	171,687	22,122	182,931	117,144	124,785	138,258
Panel D: no tracking by ability between or within classes						
Weekly hours	0.052		0.018		0.018	0.009
	(0.007)		(0.004)		(0.003)	(0.003)
# of observations	160,188		173,958		170,250	123,432
Panel E: public schools						
Weekly hours	0.061	0.020	0.031	0.035	0.019	0.011
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
# of observations	330,492	37,899	387,117	253,281	271,068	218,034
Panel F: control for lack of qualified teachers in subject						
Weekly hours	0.058		0.027	0.031		
	(0.004)		(0.002)	(0.002)		
# of observations	460,734		493,800	327,891		

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement from various sensitivity checks. See text for details on these checks. No estimates are available for some specifications in some waves because the necessary information is not available in those waves. All regressions in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Appendix Table C2: Sensitivity checks in TIMSS

	TIMSS 1995 (1)	TIMSS 1999 (2)	TIMSS 2003 (3)	TIMSS 2007 (4)	TIMSS 2011 (5)	TIMSS 2015 (6)
Panel A: main estimates (for comparison)						
Weekly hours	0.037	0.037	0.024	0.015	0.019	0.017
	(0.006)	(0.009)	(0.007)	(0.004)	(0.007)	(0.006)
# of observations	83,200	43,036	46,840	41,134	48,322	81,092
Panel B: academic record not considered for school admission						
Weekly hours	0.036	0.037				
	(0.007)	(0.009)				
# of observations	76,704	37,808				
Panel C: control for subject-specific tracking by ability						
Weekly hours	0.036	0.038		0.015		0.018
	(0.006)	(0.009)		(0.004)		(0.006)
# of observations	83,200	43,036		41,134		81,092
Panel D: control for subject-specific enrichment activities						
Weekly hours	0.036	0.037	0.023	0.015		
	(0.007)	(0.009)	(0.007)	(0.004)		
# of observations	83,200	43,036	46,840	41,134		
Panel E: control for subject-specific remedial teaching						
Weekly hours	0.037	0.038	0.023	0.015		
	(0.006)	(0.009)	(0.007)	(0.004)		
# of observations	83,200	43,036	46,840	41,134		
Panel F: control for shortage of teachers in subject						
Weekly hours					0.019	0.017
					(0.007)	(0.006)
# of observations					48,322	81,092
Panel G: control for difficulty of hiring teachers in subject						
Weekly hours			0.024	0.015	0.019	0.018
			(0.007)	(0.004)	(0.007)	(0.006)
# of observations			46,840	41,134	48,322	81,092
Panel H: control for experience and education of subject teacher						
Weekly hours	0.037	0.036	0.024	0.015	0.020	0.017
	(0.006)	(0.009)	(0.007)	(0.004)	(0.007)	(0.006)
# of observations	83,200	43,036	46,840	41,134	48,322	81,092
Panel I: control for mother's stated importance of doing well in subject						
Weekly hours	0.036	0.038				
	(0.006)	(0.009)				
# of observations	83,200	43,036				
Panel J: control for extra lessons in subject						
Weekly hours	0.037	0.036	0.024			
	(0.006)	(0.009)	(0.007)			
# of observations	83,200	43,036	45,433			

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement from various sensitivity checks. See text for details on these checks. No estimates are available for some specifications because the necessary information is not available in those waves. When information is available in a wave but the value on a control is missing for an observation, we impute this information at the sample mean and include a dummy indicating missing values in the regression. All regressions in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Paper II

Sibling Gender, Inheritance Customs and Educational Attainment: Evidence from Matrilineal and Patrilineal Societies

Abstract

I identify the causal effect of sibling gender on education and how this effect varies according to traditional inheritance customs. Using data from 27 sub-Saharan African countries, I find that boys who inherit their father's property experience no effect of sibling gender, while boys who do not inherit experience a significant negative effect of having a brother. Having a brother has a small negative effect on the education of girls, regardless of inheritance customs. The effect of sibling gender converges after the introduction of laws guaranteeing that children inherit from their parents, suggesting that parents substitute between transferring property to their children and investing in their education. Exploiting quasi-random variation in national reforms, I show that free primary education reduces the negative effect of having a brother.

Keywords: sibling gender; patriliny; matriliny; educational attainment
JEL Classifications: JEL

1 Introduction

Inequalities in the allocation of resources within households can occur because of differences in cognitive and health endowments but also due to other factors such as gender, expected inheritances, labour market prospects and cultural practices. According to classical models of human capital investment, parents will invest more in the education of children who they believe have a higher marginal return to schooling (Becker, 1981). If parents invest in their children as predicted, this will serve to exacerbate inequalities. Children cannot choose their endowments or early-life circumstances, and the family and society into which a child is born has ramifications for the allocation of resources among siblings. While a large literature examines how parental investments respond to the cognitive and health endowments of their children, we know little about how these investments respond to other factors that determine economic opportunity.[1]

In this paper, I study the effect of sibling gender on education and how that effect varies according to the inheritance customs of different ethnic groups. I first consider the mechanism through which sibling gender and inheritances interact to determine educational attainment. With constrained resources, if there are greater returns to educating boys or if parents' preferences are biased toward sons, then for a given child, having a brother rather than a sister leads to a greater diversion of resources away from that child. This loss of resources reduces educational attainment, regardless of that child's gender. If parents view education and inheritance as substitutes, then for boys who can inherit property, having a brother leads to a loss of inheritance, for which parents compensate with greater educational investment, mitigating the negative effect of having a brother.

I then test these predictions empirically. The gender composition of one's siblings is not random and stems from parental preferences over the number of children of each gender they wish to have. I circumvent this problem by exploiting the fact that for first-born children who have a second-born sibling, the gender of that sibling is as good as random. This allows me to identify the causal effect of having a second-born brother, relative to a second-born sister, on the education of the first-born child. Doing so avoids the issue of the endogeneity of family size with respect to child gender composition.[2] Due to the presence of many matrilineal and patrilineal ethnic groups, Sub-Saharan Africa provides the

[1] See Section 4 of Almond et al. (2018) for a review of the literature on how parental investments respond to the endowments of children.

[2] This is consistent with the literature using child gender as an instrument for family size including Angrist et al. (2010), Black et al. (2005) and Hank and Kohler (2000).

perfect setting to test how the effects of sibling gender vary according to whether sons will inherit from their father. In both patrilineal and matrilineal societies, property is predominantly held by males but these kinship structures give rise to different patterns of inheritance. In many patrilineal societies, inheritances are passed directly from a man to his sons. In other patrilineal societies and in matrilineal societies, inheritances are typically passed to other male heirs, which could include brothers, cousins or nephews, among others. This setting allows me to examine variation in the effect of sibling gender according to whether inheritance customs allow inheritances to be passed from fathers to sons or not.

I use nationally representative data from the Demographic and Health Survey in 27 countries, which includes information on the ethnicity and full birth history of mothers, along with the education outcomes of their children. To identify inheritance customs, these data are matched to the Ethnographic Atlas, an anthropological dataset containing information on the cultural characteristics of ethnic groups across the world.

On average, I find small yet statistically significant negative effects on education of having a brother relative to a sister. This result masks significant heterogeneity. For boys who can inherit property from their fathers, I find no effect of sibling gender on education. For boys who do not inherit from their father, I find a negative effect of having a brother of 9.4% of a year of schooling. For girls, I find a small negative effect of having a brother, regardless of inheritance customs. Overall, these findings are consistent with the hypothesis that parents view the transfer of property and investment in education as substitute goods in investing in the future of their offspring.

I provide evidence in support of this hypothesis using quasi-experimental methods. I identify five countries where, during my period of observation, inheritance laws were introduced guaranteeing the inheritance of one's property by children, regardless of gender. I identify how these reforms interact with sibling gender effects using a fixed effects strategy to identify within-country, over-time variation in inheritance laws. After the introduction of these laws, as the ability to transfer inheritance to children is the same for all parents, the effect of having a brother converges across ethnic groups according to traditional inheritance customs.

In addition, I show that government education policy can also play a large role in reducing sibling gender effects. During my period of observation, 19 countries in my sample introduced free primary education. Using a within-country fixed effects identification strategy, I show that removing primary school tuition fees reduces sibling gender effects in educational attainment for boys who do not inherit from their fathers. This result provides notable implications for the role of policy in reducing gender inequalities.

To identify other mechanisms driving the baseline results, I examine how sibling gender affects family size. In ethnic groups not practising inheritance from fathers to sons, boys who have a brother go on to have relatively more siblings, potentially contributing to the sibling gender effects identified. For girls, regardless of inheritance customs, having a brother leads to fewer total siblings, potentially mitigating the negative effect of having a brother. I also show that differences across ethnic groups in child labour, gender biases, returns to education and access to education are unlikely to contribute to the results.

I provide evidence that the estimated effect of inheritance customs can be given a causal interpretation by showing that the effects are robust to a variety of alternative identification strategies. In particular, I find similar results when using a regression discontinuity design, exploiting spatial discontinuities in inheritance rules occurring at borders between the historic homelands of different ethnic groups. The results are also robust to alternative definitions of the inheritance customs variable and to conditioning on a battery of observable predictors of educational attainment which may confound the effects of inheritance customs. I show that the results are unlikely to be driven by mis-reporting of inheritance rules by restricting my sample to only ethnic groups whose inheritance rules can be validated by other sources. I also show that my results are not driven by selection, including selection into having a sibling, childhood mortality, selection into being observed by the DHS survey and sibling sex-selection.

This paper provides a number of important contributions. First, it contributes to our understanding of how sibling gender affects education outcomes in less developed countries.[3] Specifically, this paper provides causal evidence on the effect of sibling gender on education in Sub-Saharan Africa, a region where boys continue to attain more education than girls and where there exist diverging trends in the gender gap in education across countries (Evans et al., 2021). Morduch (2000) finds a positive association between having more sisters and years of schooling in Tanzania but finds no association in South Africa. This analysis does not account for the endogeneity of family size, however, so this finding can be seen as descriptive rather than causal. Vogl (2013) studies the effects of sibling gender on the marriage outcomes of females in South Asia. In a comparative analysis using a sample of women in Sub-Saharan Africa, he

[3]The effect of sibling gender on various outcomes has examined in developed countries, including earnings (Cools and Patacchini, 2019; Gielen et al., 2016; Peter et al., 2018; Rao and Chatterjee, 2018), education (Chen et al., 2019), gender conformity (Brenøe, 2018), family formation (Peter et al., 2018) and personality (Golsteyn and Magnée, 2020).

finds that having a sister leads to fewer years of schooling, lower enrolment and a higher likelihood of being illiterate. My work provides a comprehensive examination of how sibling gender affects education and the mechanisms through which it operates.

Second, my work provides important insights on how policy interacts with culture and traditions. While evidence on how policy affects cultural practices directly is relatively rare, Bau (2021) shows that the introduction of improved pension policies reduced the practice of patrilocality and matrilocality. As patrilocality and matrilocality incentivise investments in the education of sons and daughters, respectively, Bau (2021) shows that these pension plans also reduced inequalities in investments across siblings. Ashraf et al. (2020) show that school construction in Indonesia led to increased educational attainment for girls from ethnic groups practising bride price, which incentivises the education of girls. For other girls, school construction had no effect on education. I show that inheritance laws guaranteeing inheritance to children and the introduction of free primary education can shape the inequalities in the allocation of educational resources that stem from different inheritance customs.

Third, I contribute to our understanding of how kinship affects individual-level outcomes. By identifying how the effect of sibling gender varies according to inheritance customs, I show that the ability to transfer inheritance to children can be used to reduce inequalities in the allocation of resources. An increasingly large body of literature examines various mechanisms through which kinship affects outcomes. La Ferrara and Milazzo (2017), exploiting a reform that affected inheritance rules in Ghana, examine how patrilineal inheritances lead to reduced educational attainment among boys. Gneezy et al. (2009) examine how individuals from different kinship systems have different preferences toward competition, while Pondorfer et al. (2017) identify differences between patrilineal and matrilineal societies with regard to stereotypes of risk preferences. Lowes (2020b) shows that matrilineal kinship reduces spousal co-operation and Loper (2019) shows that matrilineal women are more likely to suffer from HIV. In addition to Bau (2021), Levine and Kevane (2003) examine how patrilocality affects investment in daughters' education. My work adds to this body of evidence by showing how kinship-induced inheritance customs act as a mechanism determining intra-household educational investments.

The remainder of the paper proceeds as follows. Section 2 discusses matrilineal and patrilineal inheritance customs and considers the theoretical context. Section 3 outlines the data and empirical strategy. Section 4 presents the results, investigates the potential mechanisms underlying them and examines their robustness. Section 5 provides evidence in support of the main hypothesis and section 6 investigates whether policy can has a role to play in reducing sibling gender effects. Section 7 concludes.

2 Kinship, Inheritance Rules and Resource Allocation

Inheritance customs stem from kinship structures, which define familial descent. The majority of ethnic groups in Sub-Saharan Africa follow a unilineal kinship rule, whereby kin membership is passed from one generation to the next via one gender. In patrilineal societies, kinship passes through males, while in matrilineal societies, kinship passes through females. Figure 1 outlines the structure of patrilineal kinship systems, where men and women are represented by male and female symbols, respectively, with colours denoting kin membership. Children are incorporated into the kin of their father. After marrying, sons maintain the kin of their father, which is passed on to their own children, while daughters effectively give up their father's kin and are incorporated into the kin of their husband. Matriliny and patrilineal kinship, however, are not symmetric. Figure 2 outlines the matrilineal kinship structure. Children are incorporated into the kin of their mother. After marrying, daughters maintain the kin of their mother, which is passed on to their own children. Contrary to patrilineal kinship, sons maintain the kin of their own mother even after marrying, with their children being incorporated into the kin of their wife.

In both patrilineal and matrilineal societies, males tend to own the majority of real property, i.e. land and buildings, with inheritance rules largely following kinship.[4] There also exists variation in the pattern of inheritance within kinship structures, with inheritances being passed to different heirs across different ethnic groups. While a patrilineal man's property is most commonly inherited by his sons, property may also be passed to others. For example, among the Mende of Western Africa, a man's property is passed to his brothers by order of age (Aguwa, 2010), while a Yoruba man's property is assigned to members of his patrilineage on the basis of need (Barnes, 2009). In matrilineal societies, however, a man's property is often inherited by his sisters' sons or by other matrilineal heirs.[5] For example, among the Southern African Tonga, a man's

[4]This is not true for all matrilineal ethnic groups. For example, in some matrilineal ethnic groups in Malawi, inheritance is passed from mothers to daughters, with the husbands of those daughters working on the land (Berge et al., 2014). Not accounting for this could potentially confound the results, an issue which I address in section 4.4.

[5]One commonly used example of inheritance passing from a man to his sisters' sons is among the Akan of western Africa, whereby lineage-owned land is inherited as such (Gilbert

heir is appointed from within his matrilineage after his death (Colson, 2014) and among the Central African Suku, property is passed to the elders of a matrilineage (Kopytoff, 2016).

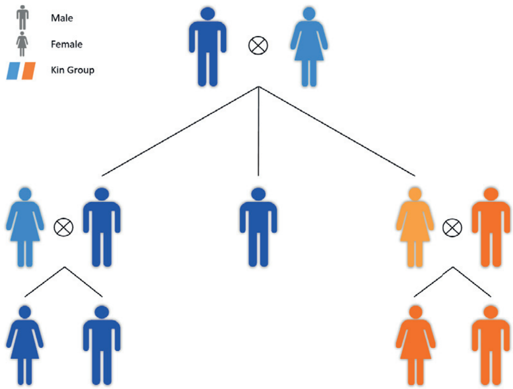


Figure 1: Patrilineal Kinship Structure

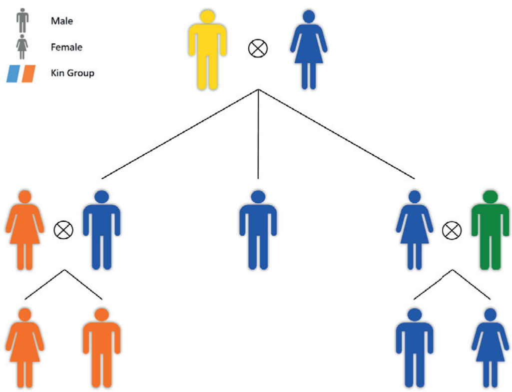


Figure 2: Matrilineal Kinship Structure

In Appendix A, I provide a theoretical framework outlining how sibling gender and the ability to transfer inheritance to sons affect the distribution of educational resources across children. Parents derive utility from preparing their children well to earn their own livelihood, either altruistically or because they need their children to care for them in their old age. One means by which parents prepare their children is through investing in their education. In line with

et al., 2000).

typical models of the household, if parents invest more in children with higher rates of return (Becker, 1981), investments in boys will be greater if there exist higher returns to education for boys. Having a brother will therefore lead to a greater diversion of limited educational resources toward that brother than would otherwise be diverted toward a sister. This relative loss of resources manifests in reduced educational attainment. This produces the framework's first testable prediction.

PREDICTION 1: *For all children, having a brother will lead to lower educational attainment, relative to having a sister.*

For parents from ethnic groups whose customs allow for inheritance to pass from fathers to sons, the transfer of inheritance provides an additional means by which parents can prepare their children. If parents view inheritance and education as substitutes, then boys who can inherit property will receive less educational resources. If inheritance is divisible, this reduces the share of resources diverted towards brothers, in turn reducing the negative effect of having a brother. Moreover, for sons who can inherit property, in addition to a reduced relative loss of educational resources, having a brother also leads to a loss of one's own expected inheritance. Parents compensate for their son's reduced inheritance with increased educational investment, reducing the negative effect of having a brother on education even further. This produces the theoretical framework's second testable prediction.

PREDICTION 2: *The negative effect of having a brother on education will be lower for children in ethnic groups where sons inherit property, particularly for boys.*

Prediction 2 relies on the assumption that parental preferences toward investing in the education of sons do not dominate any reduction in educational investments that are associated with receiving an inheritance. More crucially, however, these predictions rely on the assumption that parents in ethnic groups where sons do not inherit property from their fathers do not consider potential inheritances from other relatives when deciding on the allocation of educational resources. This is despite the fact that many boys who do not inherit directly from their father may inherit at some point from an uncle, from another family member or be distributed land at some point by the elders or leaders of their kin group.

There are various reasons why this assumption holds in the absence of inheritance from fathers to sons, some of which have received attention in the economic and anthropological literature. Even if sons will inherit from an uncle or another family member, parents face much uncertainty over if or when property will be inherited and how much will be inherited, so they may want to insure against the risk of an inadequate inheritance (La Ferrara and Milazzo, 2017).[6] Similarly, Matlon (1994) notes in a study of land use in Burkina Faso, that inheritances from fathers to sons are considered more secure than other lineage inheritances. An uncle may also have more nephews than sons or a kin leader may have more heirs than sons, leading to greater division of inheritance and, in the case of land inheritance, separation of lands inherited from different relatives, which can make inherited lands less productive. In addition, while all children have a biological father, a child may not have a maternal uncle or other relative from whom to inherit. Moreover, the substitution between education and inheritance has been noted in various contexts.[7]

3 Data and Empirical Strategy

3.1 Contemporaneous Data

Data on education outcomes and sibling sex composition are taken from the Demographic and Health Survey (DHS). The DHS are a series of household surveys implemented in developing countries across the world, providing nationally and/or regionally representative data on health and demography, with a particular focus on female respondents. In my main analysis, I use data from the household and woman's questionnaires. The household questionnaire can be completed by any knowledgeable person age 15 or older living in the household and provides information on the age, sex and education of all usual members and visitors of a household, in addition to background information on the household.

[6] Moreover, ethnic groups for whom I observe inheritance passing from uncles to nephews make up only 2% of my sample and I show in section 4.4 that the results are robust to excluding these from the analysis.

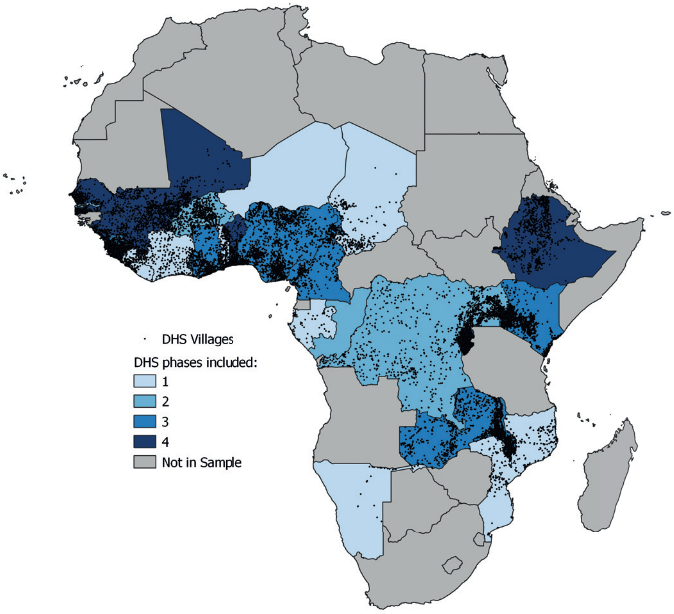
[7] In Ghana, Duncan (2010) describes how, in cocoa producing regions where matrilineal women traditionally have the right to claim a portion of their husband's land, they are "willing to accept the education of their children by the men with whom they are involved as acceptable substitutes for land." La Ferrara and Milazzo (2017) find that guaranteeing land inheritance to children in Ghana reduced the education of matrilineal boys. Similarly, Congdon Fors et al. (2019) find in Ethiopia that improving the security of land tenure reduced the education of sons who were in line to inherit, while in Kenya, Migot-Adholla et al. (1994) observe that individuals with less secure land tenure tend to have higher higher levels of education. Outside of Africa, Quisumbing and Otsuka (2001) find in Sumatra, Indonesia, where lineage-owned property was traditionally inherited by women, that a gradual evolution toward sons inheriting land is associated with rising female education.

The woman's questionnaire is given to all women between the ages of 15 and 49 residing in the household. The questionnaire includes a number of sections, with ethnicity and birth histories being of most importance in my analysis. I assign ethnicity according to a child's mother as this is the level at which I observe birth records and thus observe birth order and sibling gender.[8] My sample is comprised of first born children who are residing in their mother's household at the time of the survey. While I observe birth records for children who live elsewhere, I do not observe education outcomes for those children. In section 4.4, I show that sibling gender and inheritance customs do not predict the likelihood of being observed in my sample.

I include data from all national DHS surveys using two criteria: (1) surveys in which children's information in the household survey can be linked to the birth history of their mother from the women's survey; and (2) those in which the woman's survey includes a question on the respondent's ethnicity, where that ethnicity can be matched to an ethnic group in the ethnographic atlas which is observed to practise a matrilineal or patrilineal inheritance custom. My main sample includes data from 71 surveys implemented across DHS phases 4-8 in 27 Sub-Saharan African countries. A full list of surveys included is provided in appendix E2. The data contain geo-spatial information, which is provided by the DHS and includes the GPS co-ordinates of sampled villages and standardised geo-spatial characteristics of the villages, including data on temperature, rainfall, population density and travel times to nearby cities or international borders, among others. Figure 3 shows the countries and locations of clusters included in my main sample.

As outcome variables, I use three education outcomes, which are a child's highest grade completed, whether a child is attending school at the time of the survey and whether a child has ever attended school. Highest grade completed is calculated based on answers to questions on the highest level of school attended (e.g. primary, lower secondary, etc.) and the highest grade completed at that level. The variable is calculated based on the number of years of schooling required to complete that grade in each country. Whether a child is currently attending school is created based the answer provided to a question asking whether that child had attended school during the current school year. Whether a child ever attended school is based on the answer to a question on that child's education status, where answers may be chosen from a list of possibilities, one of which includes never having attended.

[8] I consider inter-ethnic households in section 4.4.



Notes: This map shows for the main sample the number of DHS phases included from each country and the location of clusters sampled. Phases correspond to periods between which the standard model DHS questionnaire is reviewed and modified. Participating countries are asked to adopt the model questionnaire in full but can add or delete questions where appropriate. GPS co-ordinates were not collected for all country-phase combinations and thus not all cluster locations are represented on the map but remain in the main sample.

Figure 3: Countries and location of respondents in main sample

3.2 Ethnographic Data

To obtain data on inheritance customs, I use the Ethnographic Atlas. The ethnographic atlas is a worldwide anthropological dataset containing detailed information on cultural, institutional and economic characteristics of 1,291 ethnic groups prior to industrialisation and colonial contact (Murdock, 1967).[9] As a proxy for present-day inheritance customs, I use as an explanatory variable the inheritance rule practised for real property by each ethnic group. The ethnographic atlas categorises inheritance rules into seven categories: (i) *patrilineal by sons*, (ii) *matrilineal by sister's sons*, (iii) *patrilineal by (other) heirs*, (iv)

[9]The ethnographic atlas is made publicly available online by the Database of Places, Language, Culture, and Environment (D-PLACE) and was compiled by Kirby et al. (2016).

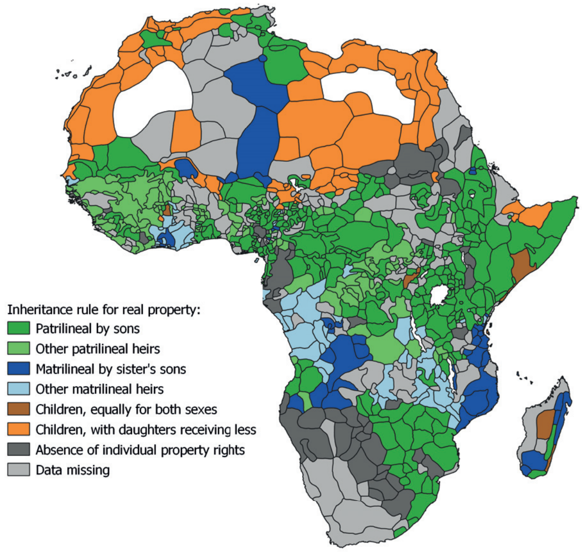
matrilineal by (other) heirs, (v) *children*, (vi) *children, less for daughters* and (vii) groups without individual property rights. Although, categories (iii) and (iv) condense many heterogeneous inheritance rules into broad categories, what is important for the analysis is that all of the ethnic groups in these categories follow customs whereby inheritance is passed to heirs other than sons.

Figure 4 shows the spatial distribution of inheritance rules observed in the ethnographic atlas.[10] Patrilineal inheritance rules are found most commonly in the data but there remains much variation in the rules observed. Matrilineal inheritance rules are common across what is known as the matrilineal belt, which spans the area intersecting modern-day Angola and the Republic of Congo on the west coast of the continent, to Mozambique and southern Tanzania on the east. A number of groups practising matrilineal inheritance rules are also observed in north and west Africa.

To examine how sibling gender affects education outcomes across ethnic groups with different inheritance customs, I match data from the ethnographic atlas with the individual-level DHS data. The classification of the DHS respondents' ethnic groups do not always coincide with those of the ethnographic atlas. For example, ethnic groups may have different names in different regions within and across countries or ethnic groups may be divided into different sub-groups in one dataset, relative to the other. In some cases, matching was straightforward as the names of ethnic groups observed in the same region in both datasets matched exactly. In other cases, a multitude of sources were used to identify matchings, such as Ethnologue, People Groups and the Joshua Project. I observe ethnicity for 169,525 individuals, which corresponds to 93.05% of my sample. Of those for whom I observe ethnicity, I am able to match 158,043, or 93.22%, to the Ethnographic Atlas. Of those matched, I observe 110,675 from ethnic groups observed to practise matrilineal or patrilineal inheritance rules.[11]

[10]This map is created using the Murdock (1959) map of the homelands of ethnic groups across Africa. The ethnic group classifications used in this map and those included in the Ethnographic Atlas differ. I therefore match ethnic groups to polygons in the map using the matchings of Teso (2019).

[11]These figures include information for Burundi, Rwanda and Lesotho which are not observed in the DHS but are imputed to the rundi, ruanda and sotho ethnic groups respectively, as these are the dominant ethnicities in these countries. In addition, ethnicity information for the Democratic Republic of Congo is categorised by the DHS into broad categories of culturally similar groups. For more information on how these are matched to the ethnographic atlas, see appendix E1. As discussed in section 4.4, the results are robust to excluding these countries from the analysis.



Notes: This map presents the inheritance rule for real property recorded in the ethnographic atlas for all ethnic groups observed by Murdock (1967) in Africa. This map is created using the Murdock (1959) map of the homelands of ethnic groups across Africa. The ethnic group classifications used in this map and those included in the ethnographic atlas differ. I therefore match ethnic groups to polygons in the map using the matchings of Teso (2019).

Figure 4: Inheritance rules in pre-colonial African ethnic groups

3.3 Empirical Strategy

Under the assumption that changes to parental investments in children's education will be reflected in observable education outcomes, I am able to test the implications of the theoretical framework. In estimating the effect of sibling gender on education outcomes, it would not be appropriate to simply compare children from families with different gender compositions. This is because sibling gender may be endogenous, with parents potentially deciding on the total number of children to have based on the gender of children who are already born. If parents decide to have further children based on the gender of existing children, it is not possible to identify the causal effect of the gender of previously born children or all children on the outcomes of any individual child.

To estimate the causal effect of sibling gender, I restrict my sample to first-born children who have at least one sibling. When parents decide to have a second child, the gender of that second child is unknown. By conditioning on children whose parents went on to have a second child, I can exploit the exogenous variation in the gender of that next sibling. By comparing first born children who have a brother to first born children who have a sister as their next sibling, I identify the causal effect of sibling gender on outcomes.[12]

I restrict my sample to individuals from ethnic groups observed to practise either patrilineal or matrilineal inheritance rules. I create an indicator variable, NDI, which represents *No direct inheritance* and takes a value of 1 if an individual's mother reports belonging to an ethnic group which is observed to practise an inheritance rule whereby a father's property is not directly inherited by his son or sons. I thus combine together groups practising patrilineal inheritance to heirs other than sons and groups practising matrilineal inheritance rules. The comparison is therefore not between patrilineal and matrilineal groups but rather between groups where sons inherit directly from their fathers and groups where sons do not.[13] The empirical specification for the main analysis is as follows:

$$\begin{aligned} y_{iect} &= \beta_1 * Brother_i + \beta_2 * Brother_i * NDI_e \\ &+ \sum_{j=0}^{1} \mathbb{1}[NDI = j] * (\alpha + X_i'\rho + \gamma_c + \delta_t) + u_{iect} \end{aligned} \tag{2}$$

where y_{iect} represents the outcome of interest for individual i from ethnic group e, observed in country c and interviewed in year t. $Brother_i$ is an indicator variable taking a value of 1 if individual i's next sibling is a brother and NDI is assigned at the ethnic group level. As inheritance rules are not randomly assigned, I control for a series of individual characteristics, which are contained in the vector X, comprising DHS phase, age at the time of the survey, gender, birth year, the interval to next sibling birth (in months), maternal age, along with indicators for living in a rural area and whether one's mother was married at birth. γ_c and δ_t represent country and interview year fixed effects, respectively. u_{iect} is the error term, which is two-way clustered at the DHS sampling cluster (village) and ethnic group levels.

[12]This identification strategy is widely used in research examining the effects of sibling gender on various outcomes, including, but not limited to, Brenøe (2018), Cools and Patacchini (2019), Golsteyn and Magnée (2020) and Peter et al. (2018)

[13]By comparing a group comprising only children from patrilineal ethnic groups to one comprising children from both patrilineal and matrilineal groups, it is possible that my results may be confounded by factors that differ across kinship structure. In section 4.4, I show that this is not the case by re-estimating the model including only patrilineal ethnic groups.

β_1 and β_2 are the key parameters of interest. β_1 reflects the effect of having a brother rather than a sister for children in direct inheritance groups. If parents place greater value on educating boys and view educational investments and inheritances as substitutes, this coefficient should be negative for girls and weakly negative for boys. This is because for boys in direct inheritance groups, while having a brother leads to a loss of educational investment, it also leads to a loss of inheritance, for which parents compensate by increasing educational investments. β_2 reflects the effect of having a brother rather than a sister for children in no direct inheritance groups, relative to direct inheritance groups. For boys, we should expect this coefficient to be negative as having a brother leads to a reduction in educational investment. For girls, we should expect this coefficient to be weakly negative. This is because the diversion of resources toward a brother should be larger for brothers who do not inherit.

While the main concern of this paper is the causal effect of sibling gender, it is of interest to identify whether the effect of inheritance customs can also be considered causal as ethnic group membership is not randomly assigned. One concern is that other drivers of education may be correlated with inheritance customs, spatially and/or across ethnic groups. In section 4.4, using a series of alternative estimation strategies and by investigating a battery of potentially confounding phenomena, I provide evidence that the estimated effect of inheritance customs has a causal interpretation.

I limit my sample to first born children who have at least one sibling. To identify the effect of sibling gender independently of family size effects, I exclude from my sample anyone who was born or whose next sibling was born as part of a twin birth or any higher order multiple birth. Furthermore, I limit my sample to children whose mothers report giving birth at normal child-bearing ages (15-49, which is the cut-off for taking part in the woman's survey) and those who are of school-going age (6-18). This leaves an overall sample of 110,675 children.

Descriptive statistics are presented in appendix table B1. The average age of children observed in my sample is around 11 years old. 67% of the sample are living in rural areas. Siblings are an average of 37 months younger than the children I observe and children in my sample are from families with an average of four total children at the time of the survey. Children in my sample completed an average of 2.97 years of schooling with 79% having attended in school at some point and 73% attending at the time of the survey. 59% of children belong to ethnic groups observed to practise direct inheritance to sons, with the remaining 41% observed practising other inheritance rules. Although second-born gender and education outcomes are very similar, differences do exist in background characteristics across inheritance groups. Relative to children in the direct inheritance group, those in the no direct inheritance group are less likely to live in rural areas, have longer intervals to the birth of their next sibling and have mothers who are less likely to have been married at birth. As will be shown in section 4.4, the results presented in this paper are not sensitive to the inclusion or exclusion of these controls.

The key identifying assumption for this analysis is that sibling sex is exogenous. The phenomenon of missing women has been observed across many developing countries and various continents. Many societies possess fewer women than men, both at birth and surviving into adulthood, which is due to pre-natal sex selection and post-natal mortality (Bongaarts and Guilmoto, 2017). If pre-natal sex selection, i.e. sex-selective abortion, is an option, then sibling gender may not be exogenous. Furthermore, the Trivers-Willard hypothesis proposes that child gender is affected by the maternal condition of the mother (Trivers and Willard, 1973). Sex-selective abortion, in particular, would require the use of ultrasound imagery to identify the sex of the unborn and it has been noted by Wanyonyi et al. (2017) that most women in Sub-Saharan Africa do not have access to ultrasound during pregnancy. So while it has been noted that mothers in Sub-Saharan Africa often possess strong preferences for the gender of their children (Fuse, 2010), the ratio of female to male births across the region is centred around the natural rate of 105:100 (Morse and Luke, 2021).

Were sibling sex to be endogenous, one might expect sibling sex to be predicted by pre-determined characteristics. Table 1 shows that this is not the case for the whole sample or for boys and girls separately. Of 34 tests of individual pre-determined characteristics, only three are significantly different at the 10% level, two of which are significant at the 5% level. In particular, two characteristics are of increased importance. If parents could engage in sex-selective abortions in order to favour sons, one would expect a higher likelihood of mothers reporting terminated pregnancies before the birth of a brother than before the birth of a sister. Table 1 shows that this is not the case. As terminated pregnancies are self-reported by the mother, terminated pregnancies before the birth of a brother might not be observable. In that case, one would expect a longer average interval to the birth of a brother than to a sister, which again is not the case. A series of joint F-tests also shows that the set of pre-determined characteristics and their interactions with own sex and inheritance rule do not predict sibling sex. In section 4.4, I discuss in greater detail various other issues of selection which may affect the results, including selection into having a sibling and excess childhood mortality.

Table 1: Balancing Tests

Panel A: Covariate Balance	Full sample				First-Born Boys				First–Born Girls			
	Sister	Brother	Difference	N	Sister	Brother	Difference	N	Sister	Brother	Difference	N
Female	0.490	0.492	0.002	110,675	0.000	0.000	0.000	56,060	1.000	1.000	0.000	54,615
Age	11.035	11.074	0.038	110,675	11.092	11.135	0.042	56,060	10.976	11.011	0.035	54,615
Rural	0.675	0.675	-0.000	110,675	0.680	0.679	-0.001	56,060	0.670	0.671	0.001	54,615
Terminated pregnancy before birth	0.034	0.035	0.001	106,557	0.032	0.034	0.002	54,025	0.036	0.035	-0.000	52,532
Terminated pregnancy before sibling birth	0.030	0.029	-0.001	106,698	0.029	0.028	-0.001	54,101	0.031	0.030	-0.001	52,597
Interval to sibling birth (months)	37.060	37.119	0.059	110,675	37.052	37.132	0.080	56,060	37.068	37.106	0.038	54,615
Mother married at birth	0.893	0.893	0.000	110,675	0.895	0.896	0.000	56,060	0.890	0.891	0.000	54,615
Mother's age at birth	20.119	20.098	-0.021	110,675	20.167	20.101	-0.066*	56,060	20.069	20.094	0.026	54,615
Mother is head of household	0.146	0.152	0.006**	110,675	0.144	0.149	0.005	56,060	0.147	0.155	0.007**	54,615
Mother has primary education	0.348	0.345	-0.002	110,675	0.343	0.343	0.000	56,060	0.352	0.348	-0.005	54,615
Mother has secondary education	0.234	0.234	0.000	110,675	0.232	0.232	-0.000	56,060	0.237	0.237	0.000	54,615
No direct inheritance	0.412	0.415	0.003	110,675	0.418	0.419	0.001	56,060	0.406	0.410	0.004	54,615
Panel B: Balancing Test												
		Joint-F	Prob>F	N		Joint-F	Prob>F	N		Joint-F	Prob>F	N
All covariates		0.8255	0.6242	106,557		0.6393	0.7964	54,025		0.598	0.8322	52,532
All, interacted with female		0.6236	0.9166	106,557								
All, interacted with no direct inheritace		0.8221	0.7065	106,557		0.7638	0.7674	54,025		1.079	0.3625	52,532
All, interacted with female and no direct inheritance		0.9168	0.6272	106,557								

Notes: In panel A, Columns (1)-(3) show the mean of a series of pre-determined characteristics for individuals with a sister and brother, respectively, followed by the difference in means and the significance level of that difference. * p<0.10 ** p<0.05 *** p<0.01. Columns (4)-(6) and (7)-(9) shows the same for boys and girls, respectively. Panel B tests whether the pre-determined characteristics included can jointly predict sibling gender.

4 Results

4.1 Main Results

Table 2 reports the main results for highest grade completed, attending school at the time of the survey (hereafter referred to as current attendance) and ever having attended school, separately for the full sample, for boys only and for girls only. Columns 1, 3 and 5 show the average effect of having a brother relative to a sister, without consideration of inheritance customs. Overall, having a brother rather than a sister leads to a reduction in years of education of 0.03, which corresponds to just over 1% of the mean and 1% of a standard deviation for my sample. Having a brother also leads to a 0.6 and 0.5 percentage point decrease in the current attendance and ever having attended school. For all three outcomes, the estimated effects are slightly larger in magnitude for girls than for boys. Building on the predictions presented in section 2, this implies that for boys and girls, having a brother rather than a sister leads to a diversion of educational investment toward that brother, which leads to a reduction in the likelihood of attending school of about 0.5 percentage points and reduces educational attainment by 3% of a year of schooling. For girls, the effects on highest grade completed are similar in magnitude but with opposite sign to those of Vogl (2013), who in a comparative analysis of women at all birth orders, found that having a sister as a next sibling led to a reduction in years of schooling of 0.028 in 30 Sub-Saharan African countries.

Columns 2, 4 and 6 then show the interaction between sibling gender and inheritance. The first row in each panel presents the effect of having a brother rather than a sister for those in ethnic groups practising direct inheritance to sons and the second row presents the effect of having a brother rather than a sister for those in no direct inheritance groups, relative to those in direct inheritance groups. If there exist significant differences across ethnic groups according to inheritance customs, this will be reflected in the second row of each panel. As can be seen in column 2 of panel B, there exists significant heterogeneity in the effect of sibling gender across inheritance groups.

For boys who can inherit their fathers property, having a brother rather than a sister has no significant effect on years of education. For boys in no direct inheritance groups, however, there exists a larger negative effect with a coefficient on *Brother*No direct inheritance* of -0.114, which is significant at the 1% level. As this is an interaction term, the net effect of having a brother rather than a sister for boys who cannot inherit from their father is estimated at 0.094 years of schooling. Similar patterns are found with respect to current attendance and

Table 2: Effect of Sibling Gender and Inheritance Customs on Educational Outcomes

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
Brother	–0.030**	–0.004	–0.006**	–0.004	–0.005*	–0.004
	(0.014)	(0.018)	(0.003)	(0.003)	(0.003)	(0.004)
Brother*No direct inheritance		–0.063**		–0.004		–0.003
		(0.024)		(0.005)		(0.005)
N	110,675	110,675	110,675	110,675	110,675	110,675
Panel B: Boys only						
Brother	–0.028	0.020	–0.005	–0.000	–0.003	0.003
	(0.022)	(0.024)	(0.004)	(0.006)	(0.004)	(0.006)
Brother*No direct inheritance		–0.114***		–0.012		–0.015*
		(0.037)		(0.008)		(0.008)
N	56,060	56,060	56,060	56,060	56,060	56,060
Panel C: Girls only						
Brother	–0.032*	–0.029	–0.006	–0.009	–0.007*	–0.010**
	(0.017)	(0.023)	(0.004)	(0.005)	(0.003)	(0.005)
Brother*No direct inheritance		–0.008		0.006		0.009
		(0.033)		(0.008)		(0.007)
N	54,615	54,615	54,615	54,615	54,615	54,615
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Controls include dummies for age, gender, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

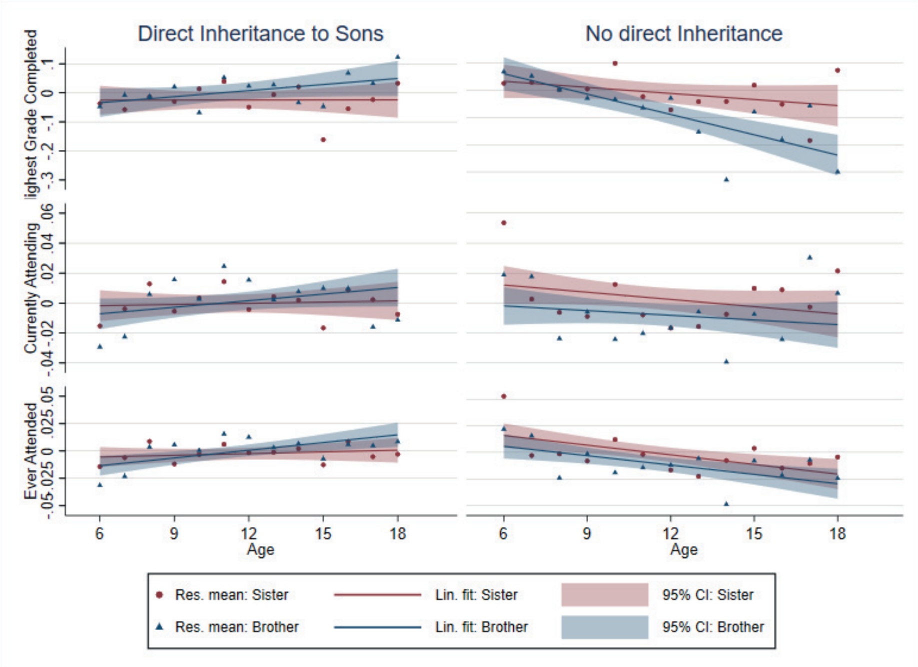
ever having attended school although the difference across inheritance groups with regard to current attendance is not statistically significant at conventional levels. This result means that for boys, having a brother leads to a loss of educational investment from parents, reducing the likelihood of ever attending school and attending school at the time of the survey by 1.2 percentage points, which leads to a reduction in years of schooling of 9.4% of a year. For boys who can inherit, however, the loss of educational resources is lower as that brother can also inherit. In addition, as parents compensate for this loss of inheritance with increased educational investment, the negative effect of having a brother is removed and thus boys who can inherit do not experience any net effect of sibling gender on education.

For girls in direct inheritance groups, having a brother leads to a smaller negative effect on highest grade completed of 0.029 and a 0.9 and 1 percentage point decrease in the likelihood of currently attending or ever having attended school, respectively. While the effects for girls in no direct inheritance groups are relatively smaller, as expected, the effects do not vary significantly across inheritance groups. It may be that, even though we should expect smaller negative effects for girls in direct inheritance group, that parental preferences toward investing more in sons still dominate, even when sons can inherit. So for girls, having a brother leads to a loss of educational resources, resulting in a lower likelihood of attending school and a reduction in years of schooling of around 3% of a year.

Appendix table B2 shows that the gender difference in the effect of sibling gender within the no direct inheritance groups is statistically significant at the 5% level with respect to highest grade completed and ever having attended school. The reason for why girls, who typically do not inherit property in either type of inheritance group, experience smaller effects of having a brother relative to boys who do not inherit is not clear. One potential reason is the competing effect of sibling gender on the future fertility decisions of parents, which I explore in more detail in section 4.2.

It is not uncommon for children to start school one year (or more) later than expected which could have implications for my results. If the effects I identify are driven by children with a brother starting school slightly later, the effects may dissipate as children get older. To investigate how the effects noted in table 2 accrue dynamically, I plot means of the residuals at each age by sibling gender separately for each inheritance rule and outcome for boys. For boys in no direct inheritance groups, the top-right panel of figure 5 shows that differences in years of education are close to zero at early ages and grow larger as those with a brother fall behind relative to those with a sister. When looking at the attendance outcomes, differences are rather constant across ages. This finding is perhaps logical, given that decisions around attendance will occur when children are at younger ages, particularly with regard to ever being enrolled, while differences in attainment will accrue over time due to grade retention and school drop-out.

Land and buildings are durable properties that typically maintain their value over generations. As populations become more urbanised, there will likely be a transition from agriculture toward other sectors, which may reduce the importance of land as a form of inheritance. These results may therefore be of less relevance to families who do not own much real property or do not work in agriculture. It is of interest to understand whether the results can be generalised to other inheritances. Table 3 shows that using the observed inheritance rule for movable property provides similar results to those using the inheritance of real property. This result implies that the findings of this analysis are relevant not just to land and buildings but also to the inheritance of other forms of wealth.



Notes: These figures plot the residualised means of the outcomes detailed on the y-axis by age and linear fits of the residualised outcomes as a function of age. Outcomes are residualised on dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth.

Figure 5: Effect of Sibling Gender on Educational Outcomes by Age - Boys only

Table 3: Effect of Sibling Gender and Inheritance Customs for Movable Property on Educational Outcomes

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls	(5) Boys	(6) Girls
Brother	0.006	–0.029	–0.007*	–0.008	–0.003	–0.008*
	(0.024)	(0.020)	(0.004)	(0.005)	(0.004)	(0.005)
Brother*No direct inheritance	–0.091**	0.000	–0.005	0.006	–0.011	0.006
	(0.037)	(0.031)	(0.007)	(0.007)	(0.007)	(0.006)
N	63,281	61,998	63,281	61,998	63,281	61,998
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. In this table, *No direct inheritance* is defined according to the observed inheritance rule for movable property. Point estimates in each column, within each panel, are taken from the same OLS regression. Controls include dummies for age, gender, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance.*

4.2 Mechanisms

In section 4.1, I identified a negative effect of having a brother rather than a sister for boys who cannot inherit from their father and for girls, regardless of inheritance customs. In this section, I investigate additional outcomes that may be affected by sibling gender and/or inheritance customs, with consequences for education. One mechanism to which I give particular consideration is family size. If parents have strong preferences over the gender composition of their children, then the gender of the first two children in a family can have a significant impact on the total number of children they have and the timing of future childbearing. While the literature on family size finds mixed results, there are a number of empirical studies pointing to a quality-quantity trade-off in the number of children parents choose to have (Booth and Kee, 2009; Chen et al., 2019; Mogstad and Wiswall, 2016; Åslund and Grönqvist, 2010), which means that overall sibling composition could have an impact on education outcomes.

I re-estimate equation 2, using as outcomes the total number of siblings a child has and the interval between the birth of their second- and third-born siblings, which is naturally conditional on having at least three children. The results of this analysis are presented in table 4. For first-born boys in direct inheritance groups, having a brother leads to a reduction in total siblings of 0.025, while for boys who cannot inherit property, there is a significant positive effect, with a coefficient of 0.101 corresponding to a 0.076 increase in total siblings. These results show that, relative to parents in direct inheritance groups, parents in no direct inheritance groups are more likely to have further children after having two sons rather than a son and a daughter, implying that those parents have a preference for a mix of sons and daughters. For boys in no direct inheritance groups, having a brother also leads to a shorter interval to the birth of a next sibling of 0.535 months. More siblings and a shorter interval to the next sibling leads to greater competition for resources, which is likely to contribute to the negative effect of having a brother for boys who cannot inherit. For girls on the other hand, having a brother leads to fewer siblings and a longer interval to the birth of a next sibling, a finding which is not found to vary significantly across inheritance groups. Descriptively, appendix table B3 shows that more siblings and a shorter interval to the birth of a third-born child are associated with worse education outcomes. Effects on family size could potentially mitigate the negative effect of having a brother, which could explain why the sibling gender effects identified are smaller in magnitude for girls.

In appendix C1, I examine a number of alternative mechanisms, specifically participation in child labour, differential gender biases according to inheritance

Table 4: Effect of Sibling Gender and Inheritance Customs on Family Outcomes

Dep. Var.:	Total Siblings		Interval to Next Sibling	
	(1)	(2)	(3)	(4)
	Boys	Girls	Boys	Girls
Brother	–0.025*	–0.050***	0.187	0.494**
	(0.014)	(0.012)	(0.207)	(0.200)
Brother*No direct inheritance	0.101***	0.032	–0.722**	–0.376
	(0.018)	(0.022)	(0.319)	(0.378)
N	56,060	54,615	46,792	45,678
Controls	X	X	X	X
Fixed Effects	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

customs, differential returns to education according to gender and inheritance customs and differential access and attitudes to education according to inheritance customs. I do not find evidence that any of these act as mechanisms underlying my results.

4.3 Heterogeneous Effects

If the effect of sibling gender is in part attributable to or affected by a family's environment, it could be expected that there exists heterogeneity in the effect of having a brother. If ownership of real property is more common and/or if customs and traditions are more persistent in rural areas, then we might expect to see larger effects in rural areas relative to urban areas. Panel A of table 5 shows that this is the case, although similar effects are still found in urban areas.

Heterogeneous effects may also occur according to household wealth. It can be argued that since less wealthy households are more budget-constrained, the effects of sibling gender may be larger due to greater relative competition for resources. On the other hand, if there is less wealth to be inherited, even boys who can inherit may experience negative effects of having a brother. In addition, if education levels in less wealthy households are generally lower, the marginal effect of having a brother may be smaller. It is therefore unclear ex ante how the results may vary according to household wealth.

Table 5: Effect of Sibling Gender and Inheritance Customs, Heterogeneity

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended		Highest Grade Completed		Currently Attending		Ever Attended	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls	(5) Boys	(6) Girls	(7) Boys	(8) Girls	(9) Boys	(10) Girls	(11) Boys	(12) Girls
	Panel A: Rural/urban status											
	Rural						Urban					
Brother	0.033	–0.043	0.002	–0.010	0.004	–0.014**	0.022	0.001	–0.005	–0.009	–0.002	–0.006
	(0.028)	(0.027)	(0.006)	(0.007)	(0.006)	(0.006)	(0.048)	(0.040)	(0.008)	(0.006)	(0.007)	(0.005)
Brother*No direct inheritance	–0.134**	0.011	–0.021**	0.008	–0.020*	0.008	–0.104	–0.033	0.002	0.007	–0.003	0.015*
	(0.053)	(0.037)	(0.009)	(0.010)	(0.010)	(0.009)	(0.071)	(0.065)	(0.016)	(0.010)	(0.013)	(0.009)
N	38,079	36,603	38,079	36,603	38,079	36,603	17,981	18,012	17,981	18,012	17,981	18,012
	Panel B: Household Wealth											
	Below median						Above median					
Brother	–0.018	–0.074**	0.000	–0.022***	0.002	–0.018**	0.052	–0.007	–0.000	–0.004	–0.003	–0.005
	(0.032)	(0.034)	(0.007)	(0.008)	(0.006)	(0.008)	(0.041)	(0.029)	(0.007)	(0.005)	(0.008)	(0.006)
Brother*No direct inheritance	–0.031	0.019	–0.011	0.011	–0.008	0.014	–0.162***	–0.009	–0.009	0.007	–0.013	–0.002
	(0.065)	(0.047)	(0.012)	(0.011)	(0.011)	(0.014)	(0.052)	(0.051)	(0.009)	(0.009)	(0.011)	(0.010)
N	26,216	25,180	26,216	25,180	26,216	25,180	25,861	25,534	25,861	25,534	25,861	25,534
	Panel C: Polygyny											
	Mother in Monogamous Union						Mother in Polygynous Union					
Brother	0.035	–0.042	0.003	–0.012*	0.004	–0.009	–0.002	–0.078	0.024***	–0.026***	0.008	–0.030***
	(0.029)	(0.029)	(0.006)	(0.006)	(0.006)	(0.007)	(0.044)	(0.054)	(0.009)	(0.007)	(0.008)	(0.009)
Brother*No direct inheritance	–0.128***	0.019	–0.013	0.011	–0.017**	0.007	–0.094	0.074	–0.041***	0.019	–0.018	0.027
	(0.049)	(0.040)	(0.009)	(0.009)	(0.008)	(0.011)	(0.081)	(0.080)	(0.013)	(0.013)	(0.013)	(0.017)
N	37,451	36,400	37,451	36,400	37,451	36,400	11,348	10,885	11,348	10,885	11,348	10,885
	Panel D: Bride Price											
	Ethnic Groups Not Practising Bride Price						Ethnic Groups Practising Bride Price					
Brother	0.077	0.009	0.003	–0.006	–0.010	–0.003	0.007	–0.032	0.003	–0.012**	0.002	–0.010*
	(0.052)	(0.048)	(0.008)	(0.015)	(0.013)	(0.013)	(0.026)	(0.025)	(0.006)	(0.005)	(0.006)	(0.006)
Brother*No direct inheritance	–0.180**	–0.014	–0.020	–0.003	–0.006	0.003	–0.097**	–0.018	–0.013	0.014**	–0.014	0.007
	(0.072)	(0.061)	(0.015)	(0.018)	(0.017)	(0.019)	(0.042)	(0.038)	(0.009)	(0.007)	(0.009)	(0.008)
N	12,986	12,737	12,986	12,737	12,986	12,737	43,074	41,878	43,074	41,878	43,074	41,878
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Countrys	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Household wealth is observed in the DHS according to an index which is based on assets owned by a household at the time of the survey. Malaria prevalence is calculated by the DHS as the percentage of children age 6-59 months tested using either microscopy or a rapid diagnostic test who are positive for malaria. HIV prevalence is calculated as the proportion of positive tests returned among adults tested in a village which produced a valid test result. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Panel B of table 5 shows how the effects vary by wealth, splitting the sample at the median level of household wealth observed in my sample.[14] In lower wealth households, having a brother does not appear to cause a reduction in years of education, although similar effects are found on both attendance outcomes for these boys as to the main analysis, albeit they are less precisely estimated. This heterogeneity points toward sibling gender effects manifesting at different margins in lower and higher wealth households. For girls, the effect of having a brother is larger in lower wealth households for all three outcomes, pointing toward greater relative investment in boys in lower wealth homes.

Sibling gender effects may also vary according to other ethnic customs and traditions. Panel C table 5 shows how the estimated effects of sibling gender and inheritance customs vary according to whether or not a child's mother is in a polygynous union. For this analysis, I partition my sample according to whether a mother reports being in a monogamous or polygynous union and re-estimate equation 2 in each sub-sample. For boys, the results are larger in magnitude and more precisely estimated among the children of monogamous mothers. As I identify status as a first-born and sibling gender at the level of the mother, some children who are identified as first-born children may not be the first born child of their mother's partner, which could attenuate the findings for children of polygynous mothers. Panel D of table 5 then shows how the results vary according to whether an ethnic group is observed in the ethnographic atlas to practise a bride price. If a group has a bride price, the higher bride prices associated with higher levels of education incentivise parents to invest in the education of girls, which could reduce sibling gender effects. For girls, however, sibling gender effects are slightly larger in bride price groups. The differences across groups according to bride price are not statistically significant, however.

4.4 Robustness

Alternative Identification Strategies

In this section, I use a variety of alternative identification strategies to show that the variation I identify in the effect of sibling gender according to inheritance customs can be given a causal interpretation. First, I use a spatial regression discontinuity (RD) design. My data includes the co-ordinates of villages sampled by the DHS and the boundaries of the ancestral homeland of each ethnic group. I can therefore exploit the spatial discontinuity in inheritance rules occurring

[14] Household wealth is observed in the DHS according to an index which is based on assets owned by a household at the time of the survey.

at the border between different ethnic groups. Using a fuzzy RD design, I instrument the *No direct inheritance* variable with an indicator for being located in the ancestral homeland of a no direct inheritance ethnic group, conditional on distance to the border of a no direct inheritance group. The main assumption is that while the inheritance rule practised changes discontinuously across ethnic group borders, other factors affecting educational attainment do not vary in space across those borders. I estimate the following system of equations:[15]

$$y_{iev} = \alpha + \beta_1 Brother_i + \beta_2 * NDI_{ev} + \beta_3 * Brother_i * NDI_{ev} + \gamma_1 * Dist_v + \gamma_2 * Dist_v^2 + u_{iev} \tag{3}$$

$$NDI_{iev} = \pi_1 + \lambda_1 * Brother_i + \rho_{11} * \mathbb{1}[Dist_v \geq 0] + \rho_{12} * Brother * \mathbb{1}[Dist_v \geq 0] + \delta_{11} * Dist_v + \delta_{12} * Dist_v^2 + \varepsilon_{iev} \tag{4}$$

$$Brother_i * NDI_{ev} = \pi_2 + \lambda_2 * Brother_i + \rho_{21} * \mathbb{1}[Dist_v \geq 0] + \rho_{22} * Brother * \mathbb{1}[Dist_v \geq 0] + \delta_{21} * Dist_v + \delta_{22} * Dist_v^2 + v_{iev} \tag{5}$$

where $Dist_v$ represents the distance, in kilometres, from village v to the nearest ancestral ethnic group border where direct inheritance to sons was not practised, re-centred such that positive values correspond to no direct inheritance areas. Distance is included using a quadratic functional form and I include only villages located within 300km of a relevant border. The exact location of urban and rural DHS sampling clusters are offset by up to 2 and 5 km from their true location, respectively. To avoid incorrectly assigning inheritance rules, I exclude individuals from villages observed within these distances of a border, resulting in a donut-RD design.

While other studies have used this approach (Loper, 2019; Moscona et al., 2020), the Murdock (1959) map is not entirely accurate in its representation of ethnic group borders and does not take into account overlapping boundaries (Michalopoulos et al., 2019). The results of this analysis may therefore not be robust to bandwidth adjustments but can be seen as indicative of a causal relationship between inheritance customs and sibling gender in determining education outcomes. These concerns are illustrated in the top panel of appendix figure B1, which shows the likelihood of a correct match between that the reported ethnicity of an individual's mother and the ethnic group whose ancestral homeland

[15] Equations 4 and 5 are first stage equations. Although the right hand side variables in both equations are identical, Angrist and Pischke (2008) and Wooldridge (2010) show that this approach is consistent when interacting an instrument with an exogenous explanatory variable.

an individual's village is located in. As can be seen, the likelihood of a correct match is much lower close to the border, although the bottom panel of appendix figure B1 does show that there remains a discrete jump in the likelihood of being in a no direct inheritance ethnic group when crossing the border. Appendix table B4 presents the results of the reduced form and fuzzy spatial RD designs. As can be seen, the results are very similar to those previously estimated. The corresponding results using a fuzzy design are larger in magnitude than those of the reduced form design.

Second, I assign inheritance customs according to the stated ethnicity of the mothers of the children in my sample. In inter-ethnic marriages it is likely that inheritance will follow the husband's customs. I test that this does not affect my results by assigning inheritance customs according to the ethnic group of the husband of the mothers in my sample, for those who can be matched. In addition, if a family migrates, they may take up the customs of where they live. To test that this does not affect my results, I assign inheritance customs according to the ethnic group homeland in which a village is located. The estimated results, using both of these alternative variables, are presented in appendix table B5. The results for boys with regard to highest grade completed are robust to both definitions. For current attendance, the results are robust to the latter definition for boys and for ever having attended, the results for girls are robust to assigning inheritance customs according to the husband.

Finally, I control for any factors correlated with inheritance customs which might confound the results. Specifically, I control for the presence of potential child-carers and foster children in the household, religion, distance to colonial religious missions, other customary characteristics from the ethnographic atlas, ethnic fractionalisation and polarisation, ethnic-group average wealth, geo-spatial controls, colonial power, exposure to the trans-Atlantic slave trade and historic crop yield data. Full details of the specific control variables included and potential biases they might cause are described in appendix D. As treatment effects could be correlated across DHS villages either spatially or due to other factors which may confound the results, I also include village fixed effects. The results of this analysis are presented in appendix tables B6, B7 and B8. With regard to highest grade completed, the results for boys are robust to all controls. In fact, the net effect of having a brother for boys in no direct inheritance increases from -0.096 years of schooling with no controls (or -0.094 with the standard controls) to -0.105 years of schooling with the full set of controls, or -0.097 when including all controls and village fixed effects. For girls, while the results do lose some significance when controls lead to lower sample sizes, the point estimates remain relatively consistent. For both current attendance and ever having attended school, the effects for boys are not robust to all controls but for girls, the effects estimated are robust to most controls but not the introduction of village fixed effects.

Selection Issues

My analysis relies on the assumption that, conditional on having a sibling, the gender of that sibling is exogenous. If parents favour boys, then the gender of the first-born may affect the likelihood of having a second child and the results could suffer from sample selection bias. Taking a sample of all first born children, I test whether gender and inheritance customs predict the likelihood of having a sibling. Column (1) of appendix table B9 shows that this is not the case.

If parents have a preference for boys and focus on the development of boys relative to girls in more than just education, then it may be that having a brother increases the likelihood of mortality. Assuming that children no longer alive would be the weakest in terms of physical health if they were to survive, they would likely be the weakest academically. This would positively bias the estimated effect of having a brother. As I observe each woman's birth history, I am able to test this and, as shown in columns (2)-(3) of appendix table B9, there is no effect of sibling gender on mortality.

If a child is living away from home, they are unlikely to be observed. If parents choose to send their brightest child or children to boarding school, then having a brother might increase the likelihood of being observed, negatively biasing the estimated effect of having a brother. Assuming that all children not observed in the household survey but who are reported to be alive are living away from home, I can test whether sibling gender affects the likelihood of being observed. As shown in columns (4)-(5) of appendix table B9, this is not the case.

The selection of ethnic groups may also introduce bias by comparing a subset of patrilineal groups to a combination of matrilineal groups and patrilineal groups. Kinship is correlated with residence patterns, which introduces incentives to invest differentially in sons or daughters (Bau, 2021). Lowes (2020b) shows that while matrilineal spouses co-operate less, matrilineal women have more autonomy and have healthier children. Patrilineal ethnic groups are more likely to practice a bride price (Lowes, 2020a), which incentivises daughters' education (Ashraf et al., 2020). In matrilineal groups, the wider kin network typically plays a larger role in investing in children, which could reduce the autonomy of parents in making decisions around investments. To test that these factors do not confound the results, I replicate the main analysis, including individuals from patrilineal groups only. Panel A of appendix table B10 shows that the results are robust to doing so.

Not all DHS surveys ask about ethnicity and in some cases I have imputed ethnicity. In Burundi and Rwanda, I assign all individuals to Rundi and Ruanda, respectively, as these are the predominant ethnic groups in these countries. The DHS categorises ethnic groups in the Democratic Republic of Congo (DRC) according to their region of origin. I categorised ethnic groups originating in present-day DRC according to their observed locations in the ethnographic atlas and assigned inheritance rules according to that observed most often in these categories.[16] This could could cause country- or region-specific factors correlated with ethnic characteristics to affect the results. Panel B of appendix table B10 shows that the results are robust to excluding imputed ethnicities.

Data Validity

The ethnographic atlas has been criticised for its reliability. As put by Abad et al. (2021), when published, the atlas "was criticized widely and harshly by linguists, historians, and anthropologists in terms that make an economics seminar seem warm and welcoming". The inheritance rule for real property was even noted by Murdock, who compiled the ethnographic atlas, as being in need of revision (Kirby et al., 2016). Bahrami-Rad et al. (2021) show that, for variables that can be observed in both the atlas and the DHS, the atlas does in fact predict contemporaneous behaviours. While this underlines the validity of the ethnographic atlas, the variables they consider do not include the inheritance rule for real property. In appendix C2, I show that the ethnographic atlas is internally valid, in that practices associated with inheritance customs do possess strong correlations in the data. Yet as the ethnographic atlas describes pre-colonial inheritance rules, some may be inaccurately observed. Matrilineal ethnic groups observed to follow inheritance to matrilineal heirs other than sister's sons are likely to include ethnic groups whereby daughters inherit land. The practices of some groups may also have deviated from these pre-colonial rules.

I supplement the main data with data from the e-Human Relations Area Files (eHRAF), an online ethnography summarising information from various sources, largely reflecting the modern day practices of over 300 ethnic groups across the world, 64 of which are in Africa. 53.8% of my sample are from ethnic groups

[16] see appendix E1 for further details on the matching of ethnic groups in the Democratic Republic of Congo.

which are represented in eHRAF and of those, 74.7% have inheritance rules for real property which are validated by the information in eHRAF. I restrict my sample to only those ethnic groups whose inheritance rule for real property can be validated and, as shown in appendix table B11, the main results for boys are largely unchanged when doing so, albeit less precisely estimated given the smaller sample size. Deviations from ancestral inheritance rules are also likely driven by institutional changes or by geo-spatial factors, such as changes in land suitability for different types of agriculture. While the main results are estimated using country fixed effects, these do not necessarily reflect local factors that might drive deviations from the rules I observe. As seen in section 4.4, however, the results are robust to the inclusion of village fixed effects, which account for unobservable village-level factors, including geo-climactic conditions.

5 Support for the Main Hypothesis

The main results are consistent with the hypothesis that parents invest more in the education of sons and substitute between property inheritance and investment in education. These results do not prove that this substitution is the main mechanism underlying these effects, however. In this section, I present supporting evidence which adds significant weight to the main hypothesis outlined in section 2.

Using the World Bank's Women, Business and the Law data (WBL), I identify five countries which, during my period of observation, introduced inheritance laws guaranteeing that children, regardless of gender, can inherit a significant proportion of their parents' property.[17] Appendix A2 considers how such reforms should impact on the effect of sibling gender identified in section 4.1. For boys who previously did not inherit from their father, the negative effect of sibling gender should be reduced. This is because the diversion of resources toward a brother will be reduced when that brother will receive an inheritance. For boys who have always inherited from their father, as having a brother no longer leads to a loss of inheritance, these boys are no longer compensated for that loss. This should lead to a negative effect of having a brother for those boys. As parents in both groups are faced with the same optimisation problem after the introduction of these inheritance reforms, there should no longer be any heterogeneity in the effect of sibling gender according to traditional inheritance customs. This gives rise to the third testable prediction arising from the theoretical framework.

[17]For a detailed review of the Women, Business and Law data, see Hyland et al. (2020).

PREDICTION 3: *When inheritance is guaranteed to sons and daughters, the effect of sibling gender should converge across inheritance customs. The negative effect of having a brother will decrease in magnitude among no direct inheritance groups and increase in magnitude among direct inheritance groups.*

To test this, I exploit within-country across-birth cohort variation in exposure to the reform. Specifically, I restrict my sample to these five countries and re-estimate equation 2, introducing a further interaction with two reform variables, which reflect the age each individual would be at the time of the reform in their country.[18] As the results presented in section 4.1 are less robust for girls than than those for boys and because I can not rule out that son preference dominates the inheritance-education trade-off, I continue looking only at boys.

Identifying the impact of these reforms is complicated by the fact that everyone in my sample is affected. Individuals are affected to a greater or lesser extent according to their age at the time these reforms were introduced. As parents may make some decisions around educational investments quite early in life, the effect of the reforms may vary according to whether a child is above or below typical school-starting ages. As such, I define two reform variables, which are respectively equal to one if a child would be between the ages of seven and twelve and if a child would be aged six or younger at the time of the reform. This is because as children aged 6 or younger at the time of exposure to the reform are below typical school-starting ages, they can be considered as fully treated, while those aged 7-12 can be expected to already be enrolled in primary school and are therefore partially treated in comparison. The equation I estimate is as follows:

$$
\begin{aligned}
y_{iect} = {} & \beta_1 * Brother_i + \beta_2 * Brother_i * NDI_e \\
& + \beta_3 * Brother_i * NDI_e * I^{Age7-12} + \beta_4 * Brother_i * NDI_e * I^{Age\leq 6} \\
& + \mathbb{1}[Brother = 1] * (\pi_1 * I^{Age7-12} + \pi_2 * I^{Age\leq 6}) \\
& + \sum_{j=0}^{1} \mathbb{1}[NDI = j] * (\alpha + \lambda_{1j} * I_i^{Age7-12} + \lambda_{2j} * I_i^{Age\leq 6} + X_i'\rho + \gamma_c + \delta_t) + u_{iect}
\end{aligned}
\tag{6}
$$

where $I^{Age7-12}$ and $I^{Age\leq 6}$ represent whether a child was aged 7-12 and 6 or below at the time of the reform, respectively. Some of these reforms were introduced alongside other measures.[19] These reforms increased female empowerment and thus the incentives to educate daughters, which may bias the results

[18] Specifically, these are Benin, Mali, Rwanda, Sierra Leone and Zambia. Sources for each of these reforms and what the reforms entailed are provided in appendix E3.

[19] In Benin and Sierra Leone, these reforms coincided with the introduction of laws allowing

of the analysis. However, the reforms introduced in Mali, Rwanda and Zambia, which comprise 65% of this sample, did not coincide with any other reforms improving the legal rights of women relative to men, so the results should still be indicative of the effect of the reforms to inheritance law. Another concern is that families might not comply with these new laws. Mali does not respect customary law as a valid source of law and while Benin and Rwanda do, customary law is considered invalid if it leads to discrimination or inequality. In Sierra Leone and Zambia, who make up 30% of the sample, this is not the case. It is also possible that despite these legal provisions, some non-compliance will occur anyway, particularly in rural areas where customs and traditions are stronger and enforcement may be more difficult. As such, the effects of the introduction of these inheritance reforms may be attenuated.

The results of this analysis are presented in table 6. In columns (1), (3) and (5), I re-estimate the main results for the reduced sample, finding similar results for boys as those found in section 4.1. In columns (2), (4) and (6) I introduce the reform variables. Prior to the reforms, there was a much larger negative effect of having a brother relative to a sister for boys in no direct inheritance groups. The effects correspond to a net 0.186 decrease in highest grade completed and 2.3 and 3.6 percentage point decreases in the likelihood of currently attended or ever having attended school, relative to boys in direct inheritance groups. After the reform, the effect of having a brother does become negative for boys in direct inheritance groups. For boys in no direct inheritance groups, the effect of having a brother is positively affected by the reforms, bringing the difference in effects between groups very close to zero for children aged six or younger at the time of the reforms. While the coefficients on the interactions with the reform are not statistically significant with respect to highest grade completed and current attendance, the reduction in effects is significant at the 5% level for ever having attended school. Given that the pattern of results is similar across all three outcomes, these results do point toward the guarantee of an inheritance for children reducing the difference across groups with regard to the effect of sibling gender.

The convergence in effects across inheritance groups is presented graphically in an event study analysis. This involves re-estimating equation 6, replacing the interaction between sibling gender, inheritance customs and age at the time of the reform ($\beta_3 * Brother_i * NDI_e * I^{Age7-12} + \beta_4 * Brother_i * NDI_e * I^{Age \leq 6}$) with a series of interactions according to age, $\Sigma_{g, g \notin [7,9]} \tau_g * Brother_i * NDI_e * I_i^{age_i \in g}$,

women to get a job and open a bank account without their husband's consent. Benin simultaneously allowed women to be the head of a household and Sierra Leone simultaneously allowed women to sign contracts and register a business without their husband's permission.

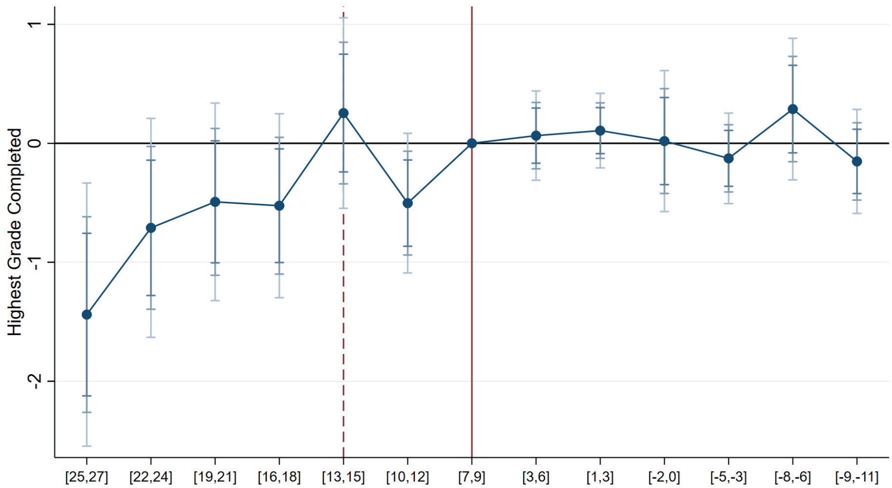
Table 6: Effect of Sibling Gender and Inheritance Customs, Reforms to Inheritance Law

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1)	(2)	(3)	(4)	(5)	(6)
Brother	0.008	0.129	–0.011	–0.000	–0.010**	0.016
	(0.020)	(0.214)	(0.008)	(0.010)	(0.005)	(0.017)
Brother*No direct inheritance	–0.126*	–0.315	–0.004	–0.023	–0.010	–0.052*
	(0.066)	(0.247)	(0.012)	(0.017)	(0.009)	(0.026)
Brother*Reform (Age 7-12)		–0.199		–0.020		–0.046
		(0.341)		(0.040)		(0.030)
Brother*Reform (Age ≤6)		–0.126		–0.010		–0.026*
		(0.248)		(0.014)		(0.015)
Brother*NDI*Reform (Age 7-12)		0.119		–0.003		0.051
		(0.391)		(0.045)		(0.045)
Brother*NDI*Reform (Age ≤6)		0.294		0.036		0.053**
		(0.278)		(0.023)		(0.024)
N	16,437	16,437	16,437	16,437	16,437	16,437
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Data on inheritance law is taken from the World Bank's *Women, Business and the Law* Database. More info on the specific reforms is provided in appendix E3. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*. Results including the coefficients on all variables included in the triple-interaction are presented in table B12.

where g represents age groups, categorised into three-year blocks. Age 7-9 is used as the reference age group as children aged 6 or younger can again be seen as being fully treated. Figure 6 shows the effect of sibling gender on highest grade completed for boys in no direct inheritance groups, relative to direct inheritance groups. A reference line is also added for age 13-15 as those aged 12 and younger at the time of the reform are partially treated. As can be seen, there exists a negative but decreasing relative effect of having a brother for boys born before the introduction of these inheritance reforms. Getting closer to the introduction of these reforms, the difference in effects starts to decrease since younger boys and/or their younger siblings would still have been exposed to the reform even if they are already at school-going ages. For boys aged 6 and below at the time of the reform, the difference in effects across inheritance customs is at or close to zero.

The results, including the coefficients on each of the variables included in the interaction terms are presented in appendix table B12. The reform had no effect on years of schooling for boys who could already inherit property and increased years of schooling for boys who previously could not. This finding disagrees with that of La Ferrara and Milazzo (2017), who examined a similar reform in Ghana, finding that guaranteeing inheritance to all children reduced educational attainment for boys who previously would not have inherited from their father. Parents may in fact view education and inheritance as complements in general, but as substitutes with regard to the intrahousehold division of resources. This



Notes: These figures plot the coefficients and 90, 95 and 99% confidence intervals on *Brother*No direct inheritance* by age at the introduction of inheritance reforms guaranteeing an inheritance to children, regardless of gender. Sources for reforms to inheritance law are detailed in appendix E3. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Figure 6: Event Study - Reform to Inheritance Law

is in line with the findings of Adhvaryu and Nyshadham (2016), who find that parents reinforce endowments on average but also mitigate against intrahousehold differences among children, a finding they attribute to inequality aversion.

Overall, the finding that guaranteeing inheritances to one's children removes any differences in the effect of having a brother according to inheritance customs points toward the mechanism outlined in section 2 as the main driver of sibling gender effects. Put simply, the intrahousehold substitution between educational investments and inheritance has a significant impact on the effect of sibling gender for boys.

6 Mitigating Sibling Gender Effects

In this section, I examine whether there is a role for educaiton policy to counteract the negative effect of having a brother. The intrahousehold allocation of resources is limited by the budget constraint of the household and one common barrier to education is that of the cost of schooling for parents. Reducing these costs could enable more children to both attend school and attain greater education. To examine the effect of reduced school costs on the effects identified, I identify 19 countries in my sample which, during my period of observation, introduced free primary education in the form of free tuition.[20]

I restrict my sample to these 19 countries and estimate an equation analogous to equation 6, now interacting sibling gender and inheritance customs with indicators for whether a child would be aged 7-12 and 6 or younger when free primary education was introduced. One issue with this analysis is that the introduction of free primary education has different meanings in different countries. For example, in Kenya and Uganda, free primary education removed tuition fees for all primary school grades. In Malawi, free primary education was introduced beginning only for first grade students, with subsequent grades being phased in over a period of four years (Kan and Klasen, 2020). In Malawi, free primary education covered tuition fees, uniforms and school books, whereas only tuition fees were removed in Ghana (Inoue and Oketch, 2008). While this means that there are likely to be heterogeneous impacts across countries, the results of this analysis are still informative about the average effect of free primary education on parental investments in education.

The results of this analysis are presented in table 7. In columns (1), (3) and (5), I re-estimate the main results for the reduced sample, finding similar results for boys as those found using the main sample. In columns (2), (4) and (6), I introduce the free primary education variables. As can be seen, prior to these reforms, there was a larger negative effect of having a brother relative to a sister for boys in no direct inheritance groups, corresponding to a net 0.205 decrease in highest grade completed. After the reforms, however, this effect is much smaller and closer to zero, significantly reducing the negative impact of having a brother for boys, particularly for boys aged 6 and below at the time of introduction. The results, including the coefficients on each of the variables included in the interaction terms are presented in appendix table B13. Free

[20] These are, in alphabetical order, Benin, Burkina Faso, Burundi, Cameroon, Democratic Republic of Congo, Ethiopia, Ghana, Ivory Coast, Kenya, Liberia, Malawi, Mozambique, Niger, Republic of Congo, Rwanda, Senegal, Sierra Leone, Togo and Zambia. Information on the sources used to identify the time of introduction is provided in appendix E4.

primary education significantly increases attendance for all boys. When it comes to highest grade completed, however, free primary education has no significant effect for boys in direct inheritance groups but has a negative effect for boys in no direct inheritance groups. This is consistent with studies showing that free primary education increases enrolment but has no effect or sometimes negative effects on achievement. This is because free education leads to increased class sizes and more marginal students enrolling in school (Lucas and Mbiti, 2012; Valente, 2019).

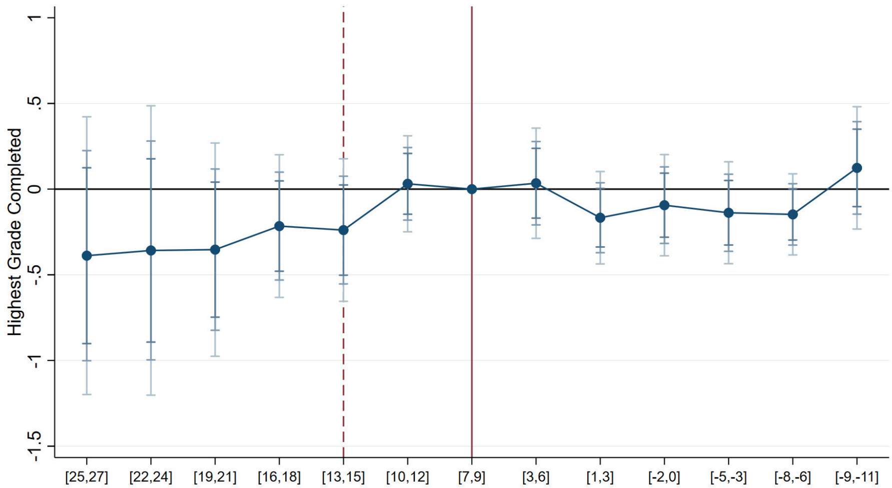
Table 7: Effect of Sibling Gender and Inheritance Customs, Introduction of Free Primary Education

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1)	(2)	(3)	(4)	(5)	(6)
Brother	0.021	0.086	–0.003	0.001	–0.000	0.017
	(0.029)	(0.091)	(0.006)	(0.013)	(0.005)	(0.016)
Brother*No direct inheritance	–0.129***	–0.291**	–0.011	–0.028	–0.012	–0.040*
	(0.047)	(0.121)	(0.009)	(0.021)	(0.009)	(0.021)
Brother*FPE (Age 7-12)		–0.073		–0.000		–0.025
		(0.145)		(0.024)		(0.023)
Brother*FPE (Age $\leq$6)		–0.086		–0.007		–0.019
		(0.106)		(0.013)		(0.016)
Brother*NDI*FPE (Age 7-12)		0.075		0.023		0.032
		(0.180)		(0.037)		(0.032)
Brother*NDI*FPE (Age $\leq$6)		0.255*		0.022		0.037
		(0.138)		(0.022)		(0.023)
N	41,007	41,007	41,007	41,007	41,007	41,007
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Sources for the introduction of free primary education are detailed in appendix E4. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*. Results including the coefficients on all variables included in the triple-interaction are presented in table B13.

I also perform an event study analysis for boys, again replacing the interaction terms for the age of a child at the time free primary education was introduced with a series of age-group interactions, categorised into three year groups. Figure 7 shows lags and leads for the coefficient on *Brother*No direct inheritance* around the introduction of free primary education. As with the analysis of the reforms to inheritance law, I use ages 7-9 as the baseline. Prior to the introduction of free primary education, while individual lags are not significantly different to zero, there is an average negative effect of having a brother for boys who cannot inherit from their fathers, relative to boys who can. After the introduction, the effects reach closer to zero on average.

These findings show that government policy has the power to affect cultural phenomena but also that school costs act as an inhibitor to gender equality in education. The use of free primary education as a means of reducing sibling gender effects is of relevance to this sample given that the average years of



Notes: These figures plot the coefficients and 95% confidence intervals on *Brother*No direct inheritance* by age at the introduction of inheritance reforms guaranteeing an inheritance to children, regardless of gender. Sources for reforms to inheritance law are detailed in appendix E4. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Figure 7: Event Study - Introduction of Free Primary Education

schooling of those I observe is 2.9 years. Going forward, as more and more children in Sub-Saharan Africa reach higher levels of education, these sibling gender effects may begin to manifest at secondary education.

7 Conclusion

Parents' investments in the human capital of their children respond to various endowments and incentives which can lead to intrahousehold inequalities in educational attainment. This paper provides novel evidence on how sibling gender and inheritance customs interact to determine education outcomes for children in 27 Sub-Saharan African countries and how policy can be leveraged to affect the allocation of resources.

I establish that sibling gender has a significant impact on education and that the effects of sibling gender vary significantly according to patrilineal and matrilineal inheritance customs. Having a brother negatively affects education outcomes for boys who are not in line to inherit property from their father and for girls in groups both with and without inheritance to sons. I supplement this finding with evidence that effects on parents' future fertility may mitigate or compound these effects depending on the gender composition of the first two children. In addition, I find no evidence that child labour, gender bias, differential returns to education or differential access to education act as mechanisms underlying the estimated effects. I show, using various alternative identification strategies and a series of robustness checks, that the estimated variation in sibling gender effects has a causal interpretation. These results underline the importance of customs and traditions in how parents' make decisions around their children's education. In particular, this paper adds to the growing evidence on mechanisms through which kinship determines education outcomes.

This paper also provides novel evidence on the power of policy to affect cultural variation in the intrahousehold allocation of resources. I show that legal reforms guaranteeing the inheritance of property by all children, regardless of gender, reduce differences in the effect of sibling gender according to traditional inheritance customs. Removing primary school tuition fees reduces the parental investment required for children to attain an education, allowing investments to be spread more evenly across children. I show that, in 19 countries, removing primary school tuition reduced the effect of sibling gender.

Overall, this paper provides interesting implications for policy makers, providing evidence that cultural traditions can act as a barrier to gender equality in education within households. These findings suggests a number of interesting avenues for further research, including how inheritance laws can be most efficiently designed to promote education and intergenerational mobility or how policies to reduce the costs of schooling can be leveraged to mitigate gender differences in education.

References

Abad, L. A., Maurer, N., et al. (2021). History never really says goodbye: A critical review of the persistence literature. *Journal of Historical Political Economy*, 1(1):31–68.

Adhvaryu, A. and Nyshadham, A. (2016). Endowments at birth and parents' investments in children. *The Economic Journal*, 126(593):781–820.

Aguwa, J. C. U. (2010). Culture summary: Mende. This culture summary is based on the article, "Mende" by Jude C. Aguwa, in the Encyclopedia of World Cultures, Supplement, Carol R. Ember, Melvin Ember, and Ian Skoggard, eds. New York: Macmillan Reference USA. 2002. Teferi Abate Adem wrote the synopsis and indexing notes in May 2008. The demography section was updated with information from the Sierra Leone census in June, 2008.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2):469–530.

Almond, D., Currie, J., and Duque, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, 56(4):1360–1446.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics*. Princeton university press.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2020). Bride price and female education. *Journal of Political Economy*, 128(2):591–641.

Åslund, O. and Grönqvist, H. (2010). Family size and child outcomes: Is there really no trade-off? *Labour Economics*, 17(1):130–139.

Atkin, D., Colson-Sihra, E., and Shayo, M. (2021). How do we choose our identity? A revealed preference approach using food consumption. *Journal of Political Economy*, 129(4):1193–1251.

Bahrami-Rad, D., Becker, A., and Henrich, J. (2021). Tabulated nonsense? Testing the validity of the Ethnographic Atlas. *Economics Letters*, 204:109880.

Barnes, S. T. (2009). Culture summary: Yoruba. This culture summary is from the article "Yoruba," by Sandra T. Barnes, in the Encyclopedia of World Cultures, Vol. 9, Africa and the Middle East, John Middleton, Amal Rassam, Candice Bradley, and Laurel L. Rose, eds. Boston, Mass.: G. K. Hall & Co. 1995. Population figures were updated in May 2008 with the advice of Sandra Barnes.

Bau, N. (2021). Can policy change culture? Government pension plans and traditional kinship practices. *American Economic Review*, forthcoming.

Becker, G. S. (1981). *A Treatise on the Family*. Harvard university press.

Berge, E., Kambewa, D., Munthali, A., and Wiig, H. (2014). Lineage and land reforms in malawi: Do matrilineal and patrilineal landholding systems represent a problem for land reforms in malawi? *Land Use Policy*, 41:61–69.

Birger, F. and Craissati, D. (2009). *Abolishing School Fees in Africa: Lessons Learned in Ethiopia, Ghana, Kenya and Mozambique*. The World Bank.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2):669–700.

Bongaarts, J. and Guilmoto, C. Z. (2017). How many more missing wonmen? Excess female mortality and prenatal sex selection, 1970-2050. *Population and Development Review*, 41(2):241–269.

Booth, A. L. and Kee, H. J. (2009). Birth order matters: The effect of family size and birth order on educational attainment. *Journal of Population Economics*, 22(2):367–397.

Brenøe, A. A. (2018). Brothers increase women's gender conformity.

Chen, S. H., Chen, Y.-C., and Liu, J.-T. (2019). The impact of family composition on educational achievement. *Journal of Human Resources*, 54(1):122–170.

Colson, E. (2014). Culture summary: Tonga. This cultural summary is from the article "Tonga" by Elizabeth Colson, in the Encyclopedia of World Cultures Vol. 9, Africa and the Middle East, John Middleton, Amal Rassam, Candice Bradley, and Laurel L. Rose, eds., Boston, Mass.: G. K. Hall & Co. 1995. Ian Skoggard updated the population figures in June 2012.

Congdon Fors, H., Houngbedji, K., and Lindskog, A. (2019). Land certification and schooling in rural Ethiopia. *World Development*, 115:190–208.

Cools, A. and Patacchini, E. (2019). The brother earnings penalty. *Labour Economics*, 58:37–51.

Demie, M. G. (2018). Cereals and gender roles: A historical perspective.

Duncan, B. A. (2010). Cocoa, marriage, labour and land in Ghana: Some matrilineal and patrilineal perspectives. *Africa*, pages 301–321.

Engel, J., Cossou, M., and Rose, P. (2011). Benin's progress in education: Expanding access and closing the gender gap. Technical report, Overseas Development Institute.

Enke, B. (2019). Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics*, 134(2):953–1019.

Evans, D. K., Akmal, M., and Jakiela, P. (2021). Gender gaps in education: The long view. *IZA Journal of Development and Migration*, 12(1):1637–1664.

Fuse, K. (2010). Variations in attitudinal gender preferences for children across 50 less-developed countries. *Demographic Research*, 23:1031–1048.

Gielen, A. C., Holmes, J., and Myers, C. (2016). Prenatal testosterone and the earnings of men and women. *Journal of Human Resources*, 51(1):30–61.

Gilbert, M., Lagacé, R. O., and Skoggard, I. (2000). Culture summary: Akan.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664.

Golsteyn, B. H. and Magnée, C. A. (2020). Does sibling gender affect personality traits? *Economics of Education Review*, 77:102016.

Government of Rwanda (2003). *LAW Nº 29/2003 OF 30/08/2003 ESTABLISHING THE ORGANISATION AND THE FUNCTIONING OF NURSERY, PRIMARY AND SECONDARY SCHOOLS*.

Government of Sierra Leone (2018). *Education Sector Plan 2018-2020 Getting it right - Service delivery, integrity and learning in Sierra Leone*. Ministry of Education, Science and Technology.

Hank, K. and Kohler, H.-P. (2000). Gender preferences for children in Europe: Empirical results from 17 FFS countries. *Demographic Research*, 2.

Hoogeveen, J. and Rossi, M. (2019). Primary education in Togo. In *Transforming Education Outcomes in Africa*, pages 9–29. Palgrave Pivot, Cham.

Hyland, M., Djankov, S., and Goldberg, P. K. (2020). Gendered laws and women in the workforce. *American Economic Review: Insights*, 2(4):475–90.

Inoue, K. and Oketch, M. (2008). Implementing free primary education policy in Malawi and Ghana: Equity and efficiency analysis. *Peabody Journal of Education*, 83(1):41–70.

Kamga, S. A. D. (2011). Realising the right to primary education in Cameroon. *African Human Rights Law Journal*, 11(1):171–193.

Kan, S. and Klasen, S. (2020). Evaluating universal primary education in Uganda: School fee abolition and educational outcomes. *Review of Development Economics*.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., et al. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PloS one*, 11(7):e0158391.

Kopytoff, I. (2016). Culture summary: Suku. This cultural summary is from the article "Suku" by Igor Kopytoff, in the Encyclopedia of World Cultures Vol. 9, Africa and the Middle East, John Middleton, Amal Rassam, Candice Bradley, and Laurel L. Rose, eds., Boston, Mass.: G. K. Hall & Co. 1995.

Kouraogo, P. and Dianda, A. Y. (2008). Education in Burkina Faso at horizon 2025. *Journal of International Cooperation in Education*, 11(1):23–38.

La Ferrara, E. and Milazzo, A. (2017). Customary norms, inheritance, and human capital: Evidence from a reform of the matrilineal system in Ghana. *American Economic Journal: Applied Economics*, 9(4):166–85.

Levine, D. and Kevane, M. (2003). Are investments in daughters lower when daughters move away? Evidence from Indonesia. *World Development*, 31(6):1065–1084.

Loper, J. (2019). Women's position in ancestral societies and female HIV: The long-term effect of matrilineality in Sub-Saharan Africa.

Lowes, S. (2020a). Kinship structure & women: Evidence from economics. *Daedalus*, 149(1):119–133.

Lowes, S. (2020b). Matrilineal kinship and spousal cooperation: Evidence from the matrilineal belt.

Lucas, A. M. and Mbiti, I. M. (2012). Access, sorting, and achievement: The short-run effects of free primary education in kenya. *American Economic Journal: Applied Economics*, 4(4):226–53.

Matlon, P. (1994). Indigenous land use systems and investments in soil fertility in Burkina Faso. In Bruce, J. W. and Migot-Adholla, S. E., editors, *Searching for Land Tenure Security in Africa*, chapter 3, pages 41–69. Kendall/Hunt Publishing Company.

Michalopoulos, S., Putterman, L., and Weil, D. N. (2019). The influence of ancestral lifeways on individual economic outcomes in Sub-Saharan Africa. *Journal of the European Economic Association*, 17(4):1186–1231.

Migot-Adholla, S. E., Place, F., and Oluoch-Kosura, W. (1994). Security of tenure and land productivity in Kenya. In Bruce, J. W. and Migot-Adholla, S. E., editors, *Searching for Land Tenure Security in Africa*, chapter 6, pages 119–140. Kendall/Hunt Publishing Company.

Ministry of Forreign Affairs, Liberia (2011). *AN ACT TO REPEAL AN ACT TO ADOPT THE EDUCATION LAW OF A. D. 2001, APPROVED JANUARY 8, 2002 AND ALL LAWS AMENDATORY THERETO, TO AMEND CERTAIN PROVISIONS OF CHAPTER 26 OF THE EXECUTIVE LAW, AND TO ENACT IN THEIR STEAD A NEW EDUCATION REFORM ACT OF 2011, TITLE 10, LIBERIAN CODE OF LAWS REVISED.*

Mogstad, M. and Wiswall, M. (2016). Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics*, 7(1):157–192.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.

Morduch, J. (2000). Sibling rivalry in Africa. *American Economic Review*, 90(2):405–409.

Morse, A. and Luke, N. (2021). Foetal loss and feminine sex ratios at birth in sub-Saharan Africa. *Population Studies*, pages 1–16.

Moscona, J., Nunn, N., and Robinson, J. A. (2020). Segmentary lineage organization and conflict in Sub-Saharan Africa. *Econometrica*, 88(5):1999–2036.

Murdock, G. P. (1959). *Africa: Its peoples and their culture history*. McGraw-Hill Book Company.

Murdock, G. P. (1967). Ethnographic Atlas: A summary.

Ndiaye, N. R. (2012). *The Road to Perpetual Stagnation: An Overview of the Senegalese Education System Since 1960*. PhD thesis, The Ohio State University.

Nunn, N. (2014). Gender and missionary influence in colonial Africa. In Akyeampong, E., Bates, R., Nunn, N., and Robinson, J. A., editors, *Africa's Development in Historical Perspective*, chapter 16, pages 489–512. Cambridge University Press.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–52.

Oyeniran, R. (2018). Education for all in developing countries: A critical analysis of Ivorian educational system. *Journal of Educational System*, 2(2):1–9.

Peter, N., Lundborg, P., Mikkelsen, S., and Webbink, D. (2018). The effect of a sibling's gender on earnings and family formation. *Labour Economics*, 54:61–78.

Pondorfer, A., Barsbai, T., and Schmidt, U. (2017). Gender differences in stereotypes of risk preferences: Experimental evidence from a matrilineal and a patrilineal society. *Management Science*, 63(10):3268–3284.

Quisumbing, A. R. and Otsuka, K. (2001). Land inheritance and schooling in matrilineal societies: Evidence from Sumatra. *World Development*, 29(12):2093–2110.

Rao, N. and Chatterjee, T. (2018). Sibling gender and wage differences. *Applied Economics*, 50(15):1725–1745.

Riddell, A. (2003). The introduction of free primary education in sub-Saharan Africa. *Background paper prepared for the Education for All Global Monitoring Report*, 4.

Roome, W. J. W. (1925). *Ethnographie survey of Africa. Showing the tribes and languages; Also the Stations of Missionary Societies.* E. Stanford Limited.

Sommeiller, E. and Wodon, Q. (2014). Enrolment gains from the elimination of primary school user fees in Burundi. *Washington, DC: World Bank.*

Teso, E. (2019). The long-term effect of demographic shocks on the evolution of gender roles: Evidence from the transatlantic slave trade. *Journal of the European Economic Association*, 17(2):497–534.

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

United Nations (2001). Committee on the rights of the child consideration of reports submitted by states parties under article 44 of the convention. CRC/C/3/Add.29/Rev.1.

United Nations (2012). Committee on the rights of the child consideration of reports submitted by states parties under article 44 of the convention. CRC/C/COG/2-4.

Valente, C. (2019). Primary education expansion and quality of schooling. *Economics of Education Review*, 73:101913.

Vogl, T. S. (2013). Marriage institutions and sibling competition: Evidence from South Asia. *The Quarterly Journal of Economics*, 128(3):1017–1072.

Wanyonyi, S. Z., Mariara, C. M., Vinayak, S., and Stones, W. (2017). Opportunities and challenges in realizing universal access to obstetric ultrasound in sub-Saharan Africa. *Ultrasound international open*, 3(2):E52.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

World Bank (2020). When I grow up, I'll be a teacher. Feature Story. available at: `https://www.worldbank.org/en/news/feature/2020/06/16/the-new-ambitions-of-congolese-schoolchildren-now-that-school-is-free`.

Appendix A: Theoretical Framework

In this section, I describe a simple theoretical framework, considering parental investment in education, incorporating gender differences in investment and the inheritance of real property. The first subsection outlines the set-up of the model. I then discuss the implications of the model with regard to how parents invest in the education of a given child according to their gender and expected inheritance. In the second subsection, I consider how the introduction of new inheritance laws allowing parents to transfer inheritance to their children would affect decision-making within this framework.

A1: The substitution between inheritance and educational investment

In order to identify the main mechanism of my application, I apply the following simplifying assumptions: (1) Households have parents and two children who differ only by gender; (2) Children are indexed by i and j but parents discriminate between children based only on gender (i.e. parents do not discriminate based on age, birth order or any other characteristics of the children, either observed or unobserved); (3) There exist no direct spillovers from one child to the other; and (4) Real property is divisible and can only be held by males.

Parents derive utility from their own consumption, C, and from what I term the 'preparedness' of their children to lead their own lives, Θ. Preparedness can be thought of more generally as being equivalent to one's earnings power but given the prevalence of subsistence farming in Sub-Saharan Africa, this could also be thought of as the ability to provide for themselves and their family. Parents may derive utility from their children's preparedness simply due to altruism or, for example, because they expect their children to provide for them later in life. Thus, the utility function of parents can be written as:

$$\mathrm{U} = U(C, \Theta_i, \Theta_j) \tag{7}$$

Preparedness takes the following form:

$$\Theta = \theta(e, p, g) \tag{8}$$

where e denotes parental investment in education, p denotes the inheritance of real property and g denotes gender. $g = 0$ for females and $g = 1$ for males, so comparing boys to girls can be considered as an increase in the value of g. Preparedness is increasing in both educational investment and property inheritance. As I discuss below, I am agnostic as to whether there exist diminishing, constant or increasing returns to educational investment.[21] In a world of perfect equality, preparedness is independent of gender but could be higher for males if, for example, men have greater autonomy in terms of their legal rights or if society is generally gender biased. Bias toward boys is perceived in many developing country settings. According to the results of rounds 3-7 of the Afrobarometer survey, among respondents in Sub-Saharan African countries, while only 15.6% stated that they believe boys should be prioritised in education, 41.2% believe that men have more of a right to a job than women and 51.6% believe that women should take care of the household. Parents also face the following budget constraint:

$$W = Y + P \tag{9}$$

where Y represents the sum of a household's liquid wealth and is divided between consumption and investment in education such that $Y = C + e_i + e_j$. P is the sum of inheritable property such that $P = p_i + p_j$. As inheritance can only be passed to sons and not daughters, among societies where inheritance to sons is not permitted due to traditional customs, $p_i = p_j = P = 0$. The distribution of real property is determined by the the sum of inheritable property and the gender of i and j and given by:

$$p^* = p(g_i, g_j, P) \tag{10}$$

Parents maximise their utility subject to their budget constraint and the preparedness functions of their children, choosing C and $e = (e_i, e_j)$. The optimal distribution of investment in education is determined by a household's liquid wealth, the gender of i and j and the expected property inheritance of i and j if parents see investments and inheritances as either complements or substitutes. This optimal distribution of investment in education is thus given by:

$$e^* = e(g_i, g_j, p_i, p_j, Y) \tag{11}$$

Using child i as the focal child (using child j provides identical results), I first

[21] For the purposes of this simple model of parental investment, assumptions on the marginal returns to property inheritance are not required, other than that the returns are positive.

consider how own gender affects preparedness, with a particular focus on how gender affects own educational investments:

$$\frac{d\Theta_i}{dg_i} = \frac{\partial\Theta_i}{\partial g_i} + \frac{\partial\Theta_i}{\partial p_i}\frac{\partial p_i}{\partial g_i} + \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial g_i} + \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial p_i}\frac{\partial p_i}{\partial g_i} \tag{12}$$

The term on the left hand side of the equation corresponds to the total effect of gender on preparedness. This total effect is divided on the right hand side into the direct effect of gender on preparedness and a series of indirect effects. The first term represents the direct effect. The second term identifies how preparedness is affected via the inheritance of real property, which is strictly positive. Inheritance is naturally increasing in gender as in this setting, only males may inherit property, which leads to increased preparedness.

The third term identifies how educational investment is affected by own gender, which in turn affects preparedness. In a world of perfect equality, this term would be equal to zero. Educational investment will be higher for boys if there exist higher returns to education for boys and parents invest more in children for whom there exist higher returns (Becker, 1981; Morduch, 2000). This term may also be positive if parent's possess a preference toward educating sons or negative in the case of a preference toward daughters. Essentially, if parents value the education of sons more than the education of daughters, boys will receive more educational investment than girls, and vice versa.

The fourth term outlines how educational investments are affected by the ability to inherit property, again affecting preparedness in turn. Overall, the sign of this term is ambiguous. As in the second term, $\frac{\partial p_i}{\partial g_i} > 0$, but the sign of $\frac{\partial e_i}{\partial p_i}$ is unknown as inheritance and education could be seen by parents as either complements or substitutes. Increasing property inheritance may lead to greater investment in education as owning more real property could increase the returns to education. On the other hand, parents may view inheritance and investment as substitutes, with the ability to transfer property to a son allowing parents to engage in greater consumption and/or greater investment in the other child's education. Substitution could occur if returns to education are diminishing with land inheritances. For example, returns might be lower for those working in agriculture than other sectors, leading parents to engage in greater consumption and/or greater investment in the other child's education. Alternatively, the returns to education may be increasing with inheritance but parents may be averse to intrahousehold inequality to the point that the negative effects of inequality on the utility of parents dominates the increasing preparedness of children who will inherit. This term is only relevant to sons in ethnic groups where inheritance from fathers to sons is permitted. The term essentially outlines that being a son rather than a daughter leads to the receipt of an inheritance, which could reduce educational investments if parents view educational investments and inheritance as substitutes.

Next, I consider the effect of sibling gender on outcomes, again using child i as the focal child:

$$\frac{d\Theta_i}{dg_j} = \frac{\partial\Theta_i}{\partial p_i}\frac{\partial p_i}{\partial p_j}\frac{\partial p_j}{\partial g_j} + \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial e_j}\frac{\partial e_j}{\partial g_j} + \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial e_j}\frac{\partial e_j}{\partial p_j}\frac{\partial p_j}{\partial g_j} + \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial p_i}\frac{\partial p_i}{\partial p_j}\frac{\partial p_j}{\partial g_j} \tag{13}$$

The term on the left hand side of the equation corresponds to the total effect of sibling gender on preparedness. This total effect is divided on the right hand side into four indirect effects, three of which work through the channel of educational investments. The first term identifies how sibling gender affects property inheritance, affecting preparedness in turn. This term is relevant only for children who stand to inherit property themselves, i.e. boys from ethnic groups practising inheritance from fathers to sons. As inheritance is limited to a maximum of P to be divided only among sons, inheritance must therefore decrease if one has a brother rather than a sister. Having a brother therefore has a negative effect on own preparedness.

The second term identifies how sibling gender directly affects educational investments, which in turn affects preparedness. Given the budget constraint, $\frac{\partial e_i}{\partial e_j}$ will be negative and, like $\frac{\partial e_i}{\partial e_j}$, $\frac{\partial e_j}{\partial g_j}$ is positive. If educational investments are higher for sons, then investments will be lower for those who have a brother, regardless of own gender. Having a brother thus means a larger share of educational resources will be invested in that brother, in comparison to a sister. For girls and boys, having a brother leads to a diversion of some resources away from parental consumption and from the education of the first child, causing a negative effect on that child's education.

The third term applies to both boys and girls in ethnic groups practising direct inheritance to sons. $\frac{\partial p_j}{\partial g_j}$ is positive, i.e. having a brother means that brother will receive an inheritance. $\frac{\partial e_j}{\partial p_j}$ is negative, however, meaning that that brother will receive less educational resources than if he were not in line to receive an inheritance. This implies that the diversion of resources towards brothers should be reduced in ethnic groups practising direct inheritance to sons.

The fourth term identifies how sibling gender affects educational investments via one's own inheritance, which again affects preparedness in turn. This term is relevant only for boys from ethnic groups practising inheritance from fathers to sons. As inheritance is lower for those with a brother, it is decreasing in sibling gender and I assume that $\frac{\partial p_i}{\partial p_j} < 0$. Again, like $\frac{\partial p_i}{\partial g_i}$, $\frac{\partial p_j}{\partial g_j} > 0$ How this reduction in inheritance affects educational investments is unclear. If inheritance and educational investment are complementary, the effect of reducing inheritance on educational investments will be negative, while if they are seen as substitutes, reducing inheritance will have a positive effect on educational investments, diminishing the negative effect of having a brother found in the second term. To sum up, for boys who can inherit real property from their father, having a brother leads to a reduced inheritance. If parents view inheritance and education as substitutes, they will compensate their son's reduced inheritance with increased educational investment, reducing the negative effect of having a brother on education.

With regard to sibling gender, under the assumption that any changes in educational investments manifest themselves in changes to observed education outcomes, this basic model of parental investment in education provides three testable implications: (1) For both girls and boys, if parents place a higher value on educating sons, then having a brother, rather than a sister, should lead to lower educational investment; (2) If parents view inheritance and education as substitutes, then for boys who are in line to inherit property from their father, this reduction in educational investments should be smaller than for boys who are not. This is because the initial reduction in investment is smaller and because parents compensate for lost inheritance by increasing educational investments; and (3) If parents view inheritance and education as substitutes, then for girls in ethnic groups where sons inherit from their father, this reduction in educational investments should be smaller than for girls in ethnic groups where boys do not inherit. This third implication, however, requires that condition that parents' preference for investing in sons will not dominate even when sons can inherit.

A2: Guaranteeing inheritance from parents to sons and daughters

This subsection considers how the introduction of inheritance laws guaranteeing inheritance to sons and daughters impacts on the effect of sibling gender. Such a law would impact the theoretical framework presented above in a number of ways. First, looking to how own gender affects education, as inheritance is divided among all children, gender no longer matters for inheritance and $\frac{\partial p_i}{\partial g_i}$ goes to zero. This means equation 12 collapses to:

$$\frac{d\Theta_i}{dg_i} = \frac{\partial\Theta_i}{\partial g_i} + \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial g_i} \tag{14}$$

This implies that now, for any given child, being a boy rather than a girl increases preparedness only through the direct effect of gender and by receiving a greater share of educational resources.

Looking to how sibling gender affects education, sibling gender no longer matters for sibling inheritance and $\frac{\partial p_j}{\partial g_j}$ goes to zero. Equation 13 therefore collapses to:

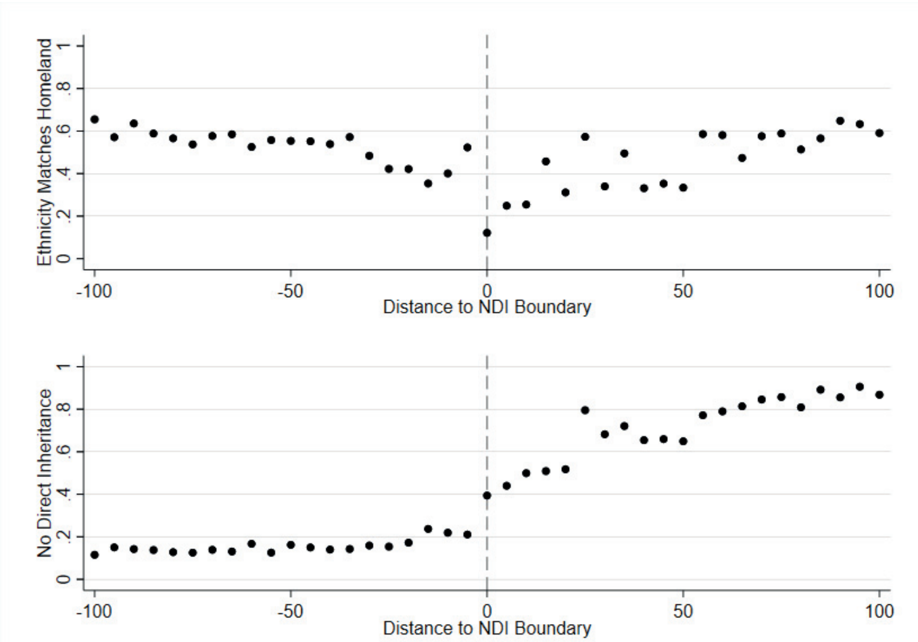
$$\frac{d\Theta_i}{dg_j} = \frac{\partial\Theta_i}{\partial e_i}\frac{\partial e_i}{\partial e_j}\frac{\partial e_j}{\partial g_j} \tag{15}$$

This implies that sibling gender now only affects education via the diversion of resources toward brothers. Looking only at equations 14 and 15, we might expect that the introduction of laws such as this will lead to negative effects of having a brother for all children. But this fails to take into account all aspects of the reform. First, consider boys who previously did not inherit from their parents. Laws such as these guarantee that those boys will receive an inheritance, which will reduce educational investments, assuming that parents view inheritance and education as substitutes. As child i is receiving less education as a result of the inheritance law, the magnitude of the effect of having a brother may be reduced. Moreover, as child j will also receive an inheritance, the educational resources diverted toward that sibling will be reduced. As boys receive a larger share of resources, this reduction in sibling education will be larger in absolute terms for brothers. This will reduce the negative effect of having a brother. Conversely, by guaranteeing inheritance to all children, for boys who could always inherit, having a brother no longer leads to a loss of expected inheritance. This means that those boys are no longer compensated for that loss of inheritance and we should expect to find negative effects of having a brother for boys who have always been able to inherit.

Most importantly, the guarantee of inheritance to sons and daughters means that parents from all ethnic groups will be faced with the same optimisation problem. This means that, after the introduction of these new inheritance laws, the effects of sibling gender should converge and there should be no differences in the effects of sibling gender according to traditional inheritance customs.

Considering the effect of the reform therefore provides three implications that, after guaranteeing inheritance to sons and daughters: (1) The negative effect of sibling gender should be reduced for boys who traditionally could not inherit from their fathers; (2) The negative effect of sibling gender should increase for boys who traditionally could not inherit from their fathers; and (3) The effect of sibling gender should not longer vary according to traditional inheritance customs.

Appendix B: Further Tables and Figures



Notes: The top panel plots on the y-axis, in 5km bins, the proportion of individuals whose observed ethnic group in the DHS matches the ethnic group whose homeland the individual's village is located in, according to the Murdock (1959) map of ethnic homelands. The bottom panel shows on the y-axis, in 5km bins, the proportion of individuals observed to belong to an ethnic group practising a no direct inheritance rule. In both panels, the x-axis shows the distance, in km, from an individual's village to the ancestral border of an ethnic group practising a no direct inheritance rule. The ethnic group classifications used in this map and those included in the ethnographic atlas differ. I therefore match ethnic groups to polygons in the map using the matchings of Teso (2019).

Figure B1: Regression Discontinuity Design

Table B1: Descriptive Statistics

	Full Sample			Direct Inheritance to Sons			No Direct Inheritance		
	All first-borns	First-born boys	First-born girls	All first-borns	First-born boys	First-born girls	All first-borns	First-born boys	First-born girls
Background characteristics									
Female	0.49	0.00	1.00	0.50	0.00	1.00	0.49	0.00	1.00
Age	11.07	11.14	11.00	11.08	11.13	11.03	11.04	11.14	10.94
Birth year	1999.42	1999.34	1999.50	1999.06	1998.97	1999.14	1999.94	1999.85	2000.03
Year of interview	2010.86	2010.84	2010.87	2010.51	2010.48	2010.54	2011.36	2011.37	2011.35
Rural	0.67	0.68	0.67	0.70	0.70	0.70	0.64	0.65	0.63
Interval to next sibling (in months)	36.97	36.96	36.98	36.05	36.13	35.98	38.28	38.14	38.43
Family size	4.01	4.01	4.00	4.06	4.05	4.06	3.93	3.95	3.91
Mother's age at birth	20.10	20.12	20.08	20.20	20.20	20.21	19.96	20.02	19.90
Mother married at birth	0.89	0.90	0.89	0.90	0.91	0.90	0.88	0.88	0.88
Mother is household head	0.15	0.15	0.15	0.17	0.17	0.17	0.12	0.12	0.13
Mother completed primary school	0.34	0.33	0.34	0.35	0.35	0.36	0.32	0.31	0.32
Mother completed secondary school	0.23	0.23	0.23	0.22	0.22	0.22	0.24	0.24	0.25
Treatment and Outcomes									
Next sibling is male	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Highest grade completed	2.97	2.98	2.97	2.99	2.98	2.99	2.95	2.97	2.93
Currently attending school	0.73	0.74	0.73	0.74	0.74	0.73	0.72	0.73	0.72
Ever attended school	0.79	0.80	0.79	0.80	0.80	0.79	0.78	0.79	0.78
Inheritance rule for real property									
Patrilineal by sons	0.59	0.59	0.59	1.00	1.00	1.00	0.00	0.00	0.00
Other patrilineal heirs	0.23	0.24	0.23	0.00	0.00	0.00	0.56	0.57	0.56
Matrilineal by sister's sons	0.02	0.02	0.02	0.00	0.00	0.00	0.06	0.05	0.06
Other matrilineal heirs	0.16	0.16	0.16	0.00	0.00	0.00	0.38	0.38	0.39
N	110,675	56,060	54,615	65,231	32,812	32,419	45,444	23,248	22,196

Notes: This table presents the mean of individual and family background characteristics, treatment and outcome variables and observed inheritance rules for the whole sample of first-born children who have at least one sibling.

Table B2: Effect of Sibling Gender and Inheritance Customs on Educational Outcomes

Dep. Var.:	Highest grade completed	Currently attending	Ever attended
	(1)	(2)	(3)
Brother	0.020	–0.000	0.003
	(0.025)	(0.006)	(0.006)
No direct inheritance	1.263*	–0.127	–0.193
	(0.683)	(0.164)	(0.172)
Female	0.415	–0.141	–0.131
	(0.473)	(0.137)	(0.128)
Brother*Female	–0.049	–0.008	–0.013*
	(0.032)	(0.009)	(0.008)
No direct inheritance*Female	–0.033	0.094	0.159
	(0.776)	(0.198)	(0.196)
Brother*No direct inheritance	–0.114***	–0.012	–0.015*
	(0.038)	(0.008)	(0.008)
Brother*No direct inheritance*Female	0.106**	0.019	0.024**
	(0.052)	(0.012)	(0.011)
N	110,675	110,675	110,675
Controls	X	X	X
Fixed Effects	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Point estimates in each column are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*, female and both.

Table B3: Family Outcomes and Educational Attainment

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls	(5) Boys	(6) Girls
Panel A: Total number of siblings						
Siblings	–0.213***	–0.246***	–0.012***	–0.014***	–0.007**	–0.008**
	(0.046)	(0.049)	(0.004)	(0.004)	(0.003)	(0.003)
Siblings*No direct inheritance	–0.050	–0.019	–0.008*	–0.007	–0.005	–0.004
	(0.055)	(0.060)	(0.005)	(0.005)	(0.004)	(0.005)
N	56,060	54,615	56,060	54,615	56,060	54,615
Panel B: Interval from 2nd to 3rd birth						
Interval	0.005***	0.007***	0.000*	0.000*	0.000	0.000
	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Interval*No direct inheritance	0.000	–0.001	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
N	46,792	45,678	46,792	45,678	46,792	45,678
Controls	X	X	X	X	X	X
Fied Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval between the births of the first and second-born siblings in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Table B4: Regression Discontinuity Design

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls	(5) Boys	(6) Girls
Panel A: Reduced Form RD Design						
Brother	0.027	–0.060	–0.002	–0.012**	–0.002	–0.012**
	(0.044)	(0.044)	(0.005)	(0.006)	(0.006)	(0.005)
Brother*No direct inheritance	–0.091	0.035	–0.013	0.001	–0.011	0.002
	(0.066)	(0.064)	(0.010)	(0.009)	(0.011)	(0.008)
Panel B: Fuzzy RD Design						
Brother	0.073	–0.058	0.004	–0.012	0.004	–0.011
	(0.060)	(0.061)	(0.008)	(0.008)	(0.008)	(0.007)
Brother*No direct inheritance	–0.186*	0.043	–0.026	0.000	–0.022	0.002
	(0.113)	(0.108)	(0.018)	(0.015)	(0.018)	(0.013)
N	48,265	46,915	48,265	46,915	48,265	46,915

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column in panel A, within each panel, are taken from the same regression. Regressions include a second order polynomial for distance to the boundary of an ethnic group practising no direct inheritance.

Table B5: Alternative Inheritance Variables

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls	(5) Boys	(6) Girls
Panel A: Inheritance Rule of Husband's Ethnic Group						
Brother	0.017	–0.028	–0.005	–0.010	–0.006	–0.022**
	(0.043)	(0.048)	(0.008)	(0.011)	(0.007)	(0.009)
Brother*No direct inheritance	–0.135*	0.069	–0.011	–0.010	–0.005	0.002
	(0.070)	(0.062)	(0.016)	(0.015)	(0.014)	(0.015)
N	17,720	17,670	17,720	17,670	17,720	17,670
Panel B: Inheritance Rule According to Location in Ethnic Group Homeland						
Brother	0.051	–0.023	0.006	0.006	–0.003	–0.001
	(0.038)	(0.039)	(0.005)	(0.009)	(0.005)	(0.009)
Brother*No direct inheritance	–0.159***	–0.012	–0.021**	–0.000	–0.008	0.009
	(0.054)	(0.051)	(0.010)	(0.011)	(0.008)	(0.011)
N	28,114	26,954	28,114	26,954	28,114	26,954
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Table B6: Controlling for Confounders - Highest Grade Completed

	Dep. Var. = Highest Grade Completed											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Boys only												
Brother	0.055	0.020	0.022	0.000	0.023	0.021	0.018	0.033	0.032	–0.033	0.070	0.052
	(0.037)	(0.024)	(0.024)	(0.042)	(0.027)	(0.026)	(0.029)	(0.023)	(0.023)	(0.059)	(0.044)	(0.066)
Brother*No direct inheritance	–0.151***	–0.114***	–0.108***	–0.110**	–0.121***	–0.115***	–0.145***	–0.136***	–0.135***	–0.139**	–0.160**	–0.149**
	(0.054)	(0.036)	(0.038)	(0.043)	(0.038)	(0.037)	(0.042)	(0.036)	(0.035)	(0.053)	(0.067)	(0.061)
N	56,060	56,060	54,173	50,541	48,094	52,077	46,638	52,223	52,596	34,236	47,988	34,046
Panel B: Girls only												
Brother	0.009	–0.029	–0.030	–0.107***	–0.029	–0.042*	–0.035	–0.036	–0.030	–0.085*	0.032	–0.055
	(0.039)	(0.023)	(0.022)	(0.033)	(0.025)	(0.025)	(0.028)	(0.023)	(0.023)	(0.051)	(0.039)	(0.050)
Brother*No direct inheritance	–0.088	–0.008	–0.009	0.022	0.003	0.002	0.005	–0.006	–0.009	0.017	0.004	0.048
	(0.059)	(0.032)	(0.035)	(0.032)	(0.034)	(0.034)	(0.034)	(0.033)	(0.032)	(0.048)	(0.067)	(0.052)
N	54,615	54,615	52,716	49,054	46,941	50,714	45,221	50,698	51,098	33,208	46,207	32,823
Standard controls		X	X	X	X	X	X	X	X	X		X
Child-carer present			X							X		X
Foster children			X							X		X
Religion				X						X		X
Mission distance				X						X		X
Ethnographic controls					X					X		X
Ethnic fractionalisation						X				X		X
Ethnic polarisation						X				X		X
Ethnic-group wealth						X				X		X
Geo-spatial controls							X			X		
Slave trade exposure								X		X		X
Crop yields									X	X		X
Fixed effects	None	Country	Country	Country	Country	Country	Country	Country	Country	Country	Village	Village

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Standard controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Information on other controls included is provided in appendix D. Each of these controls are also interacted with *No direct inheritance*.

Table B7: Controlling for Confounders - Currently Attending

	Dep. Var. = Currently Attending											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Boys only												
Brother	−0.005	−0.000	0.001	0.005	−0.004	0.002	0.003	0.002	0.001	−0.003	0.001	0.004
	(0.006)	(0.005)	(0.005)	(0.008)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.010)	(0.006)	(0.014)
Brother*No direct inheritance	−0.011	−0.012	-0.015^{*}	-0.015^{**}	−0.009	−0.013	-0.018^{**}	-0.014^{*}	-0.014^{*}	−0.013	−0.014	−0.017
	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.011)	(0.010)	(0.015)
N	56,060	56,060	54,173	50,541	48,094	52,077	46,638	52,223	52,596	34,236	47,988	34,046
Panel B: Girls only												
Brother	−0.008	−0.009	-0.010^{*}	-0.023^{***}	−0.008	-0.011^{**}	-0.011^{*}	-0.010^{*}	-0.009^{*}	-0.026^{**}	−0.005	−0.006
	(0.006)	(0.005)	(0.005)	(0.008)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.012)	(0.007)	(0.013)
Brother*No direct inheritance	−0.001	0.006	0.005	0.013	0.003	0.007	0.007	0.009	0.008	0.002	0.013	0.003
	(0.009)	(0.008)	(0.008)	(0.010)	(0.008)	(0.009)	(0.010)	(0.008)	(0.008)	(0.012)	(0.012)	(0.014)
N	54,615	54,615	52,716	49,054	46,941	50,714	45,221	50,698	51,098	33,208	46,207	32,823
Standard controls		X	X	X	X	X	X	X	X	X		X
Child-carer present			X							X		X
Foster children			X							X		X
Religion				X						X		X
Mission distance				X						X		X
Ethnographic controls					X					X		X
Ethnic fractionalisation						X				X		X
Ethnic polarisation						X				X		X
Ethnic-group wealth						X				X		X
Geo-spatial controls							X			X		
Slave trade exposure								X		X		X
Crop yields									X	X		X
Fixed effects	None	Country	Country	Country	Country	Country	Country	Country	Country	Country	Village	Village

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Standard controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Information on other controls included is provided in section appendix D. Each of these controls are also interacted with *No direct inheritance*.

Table B8: Controlling for Confounders - Ever Attended

	Dep. Var. = Ever Attended											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Boys only												
Brother	−0.001	0.003	0.003	0.007	−0.001	0.003	0.004	0.004	0.003	−0.007	0.003	0.008
	(0.006)	(0.005)	(0.005)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.010)	(0.007)	(0.012)
Brother*No direct inheritance	−0.013	−0.015*	−0.017**	−0.017**	−0.014*	−0.012	−0.023***	−0.016*	−0.015*	−0.020**	−0.020*	−0.022*
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.011)	(0.011)
N	56,060	56,060	54,173	50,541	48,094	52,077	46,638	52,223	52,596	34,236	47,988	34,046
Panel B: Girls only												
Brother	−0.010*	−0.010**	−0.011**	−0.016**	−0.011**	−0.013***	−0.012**	−0.011**	−0.010**	−0.020**	−0.007	−0.003
	(0.006)	(0.005)	(0.005)	(0.008)	(0.005)	(0.004)	(0.006)	(0.004)	(0.005)	(0.010)	(0.005)	(0.010)
Brother*No direct inheritance	0.004	0.009	0.009	0.012	0.010	0.011	0.010	0.012*	0.010	0.011	0.014	0.008
	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.009)	(0.010)	(0.009)
N	54,615	54,615	52,716	49,054	46,941	50,714	45,221	50,698	51,098	33,208	46,207	32,823
Standard controls		X	X	X	X	X	X	X	X	X		X
Child-carer present			X							X		X
Foster children			X							X		X
Religion				X						X		X
Mission distance				X						X		X
Ethnographic controls					X					X		X
Ethnic fractionalisation						X				X		X
Ethnic polarisation						X				X		X
Ethnic-group wealth						X				X		X
Geo-spatial controls							X			X		
Slave trade exposure								X		X		X
Crop yields									X	X		X
Fixed effects	None	Country	Country	Country	Country	Country	Country	Country	Country	Country	Village	Village

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Standard controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Information on other controls included is provided in section appendix D. Each of these controls are also interacted with *No direct inheritance*.

Table B9: Selection Into Having a Sibling, Being Observed

Dep. Var.:	Any Sibling	Survival to Survey		Observed by Survey	
	All	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)	(5)
Female	0.001				
	(0.002)				
Brother		–0.000	0.002	0.006	0.001
		(0.003)	(0.003)	(0.005)	(0.005)
Female*No direct inheritance	–0.003				
	(0.003)				
Brother*No direct inheritance		–0.001	–0.004	–0.004	0.001
		(0.005)	(0.005)	(0.008)	(0.009)
N	162,854	92,163	89,365	76,803	76,647
Controls	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Table B10: Selection of Ethnic Groups

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1)	(2)	(3)	(4)	(5)	(6)
	Boys	Girls	Boys	Girls	Boys	Girls
Panel A: Excluding Matrilineal Groups						
Brother	0.020	–0.029	–0.000	–0.009	0.003	–0.010**
	(0.024)	(0.024)	(0.006)	(0.006)	(0.006)	(0.005)
Brother*No direct inheritance	–0.141***	–0.005	–0.010	0.005	–0.018*	0.012*
	(0.047)	(0.035)	(0.009)	(0.009)	(0.010)	(0.007)
N	46,031	44,775	46,031	44,775	46,031	44,775
Panel B: Excluding Imputed Ethnicities						
Brother	0.000	–0.037	–0.001	–0.010	0.007	–0.007
	(0.030)	(0.025)	(0.007)	(0.006)	(0.006)	(0.006)
Brother*No direct inheritance	–0.089**	0.001	–0.013	0.006	–0.019**	0.006
	(0.042)	(0.034)	(0.009)	(0.009)	(0.009)	(0.007)
N	46,159	44,477	46,159	44,477	46,159	44,477
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Table B11: Inheritance Rules Validated by eHRAF

Dep. Var.:	Highest Grade Completed		Currently Attending		Ever Attended	
	(1) Boys	(2) Girls	(3) Boys	(4) Girls	(5) Boys	(6) Girls
Brother	0.040	–0.013	0.007	–0.005	0.004	–0.005
	(0.024)	(0.029)	(0.006)	(0.007)	(0.008)	(0.007)
Brother*No direct inheritance	–0.107	–0.059	–0.008	0.000	–0.021*	0.001
	(0.066)	(0.040)	(0.011)	(0.013)	(0.010)	(0.011)
N	22,246	21,881	22,246	21,881	22,246	21,881
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Table B12: Effect of Sibling Gender and Inheritance Customs, Reforms to Inheritance Law

Dep. Var.:	Highest Grade Completed (1)	Currently Attending (2)	Ever Attended (3)
Brother	0.129	–0.000	0.016
	(0.214)	(0.010)	(0.017)
No direct inheritance	0.427	–0.060	–0.152
	(1.436)	(0.251)	(0.312)
Brother*No direct inheritance	–0.315	–0.023	–0.052*
	(0.247)	(0.017)	(0.026)
Reform (Age 7-12)	0.079	0.099***	0.079**
	(0.342)	(0.025)	(0.033)
Reform (Age $\leq$6)	–0.128	0.140***	0.083
	(0.421)	(0.046)	(0.071)
Brother*Reform (Age 7-12)	–0.199	–0.020	–0.046
	(0.341)	(0.040)	(0.030)
Brother*Reform (Age $\leq$6)	–0.126	–0.010	–0.026*
	(0.248)	(0.014)	(0.015)
NDI*Reform (Age 7-12)	0.810*	–0.011	0.037
	(0.462)	(0.044)	(0.052)
NDI*Reform (Age $\leq$6)	1.094*	–0.042	0.070
	(0.623)	(0.073)	(0.095)
Brother*NDI*Reform (7-12)	0.119	–0.003	0.051
	(0.391)	(0.045)	(0.045)
Brother*NDI*Reform (Age $\leq$6)	0.294	0.036	0.053**
	(0.278)	(0.023)	(0.024)
N	16,437	16,437	16,437
Controls	X	X	X
Fixed Effects	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from the same OLS regression. Data on inheritance law is taken from the World Bank's *Women, Business and the Law* Database. More info on the specific reforms is provided in appendix E3. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance.*

Table B13: Effect of Sibling Gender and Inheritance Customs, Introduction of Free Primary Education

Dep. Var.:	Highest Grade Completed (1)	Currently Attending (2)	Ever Attended (3)
Brother	0.086	0.001	0.017
	(0.091)	(0.013)	(0.016)
No direct inheritance	0.941	–0.085	–0.349
	(1.003)	(0.235)	(0.223)
Brother*No direct inheritance	–0.291**	–0.028	–0.040*
	(0.121)	(0.021)	(0.021)
FPE (Age 7-12)	0.420***	0.070***	0.045
	(0.149)	(0.026)	(0.031)
FPE (Age ≤6)	0.429**	0.050	0.037
	(0.181)	(0.034)	(0.041)
Brother*FPE (Age 7-12)	–0.073	–0.000	–0.025
	(0.145)	(0.024)	(0.023)
Brother*FPE (Age ≤6)	–0.086	–0.007	–0.019
	(0.106)	(0.013)	(0.016)
NDI*FPE (Age 7-12)	–0.614***	–0.070*	–0.032
	(0.215)	(0.040)	(0.040)
NDI*FPE (Age ≤6)	–0.909***	–0.045	–0.046
	(0.263)	(0.051)	(0.055)
Brother*NDI*FPE (7-12)	0.075	0.023	0.032
	(0.180)	(0.037)	(0.032)
Brother*NDI*FPE (Age ≤6)	0.255*	0.022	0.037
	(0.138)	(0.022)	(0.023)
N	41,007	41,007	41,007
Controls	X	X	X
Fixed Effects	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from the same OLS regression. Sources for the introduction of free primary education are detailed in appendix E4. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with *No direct inheritance*.

Appendix C: Additional Results

C1: Mechanisms

In section 4.1, I identified a negative effect of having a brother rather than a sister for boys who cannot inherit from their father and for girls, regardless of inheritance customs, and in section 4.2 I showed that family size outcomes contribute to the difference in effects between boys and girls. In this section, I investigate various additional outcomes that may be affected by sibling gender and/or inheritance customs, with consequences for education outcomes.

One mechanism which may contribute to sibling gender effects is that of child labour. If some household tasks are assigned based on gender, then the gender of one's sibling could have a significant impact on the amount and type of work conducted by children. If children can inherit property, they may be expected to engage more in family work as a child in order to learn skills relevant to the land or business they inherit or may simply be expected to do more work during the time they would otherwise be in school. A subset of DHS surveys include modules on the performance of child labour. For this subset of surveys, I re-estimate equation 2 using as outcomes whether, in the week prior to the survey, a child is reported to have engaged in any work, paid work, fetched wood or water, engaged in household chores or other unpaid work for family. The results are presented in table C1. For boys in no direct inheritance groups, having a brother appears to reduce th likelihood of fetching firewood or water, while for girls, having a brother appears to reduce the likelihood of having done unpaid work for family. There is no evidence of any effect of sibling gender on other measures of child labour, however. Overall, these results do not provide any conclusive evidence that sibling gender and inheritance affect the likelihood of engaging in child labour. As these variables measure the extensive margin of any child labour in the week prior to the survey, I cannot rule out effects of sibling gender at the intensive margin, i.e. the amount of labour engaged in.

I now turn my attention to the possibility that the effects I find are partly due to other factors that are correlated with inheritance customs and also affect educational investments. If groups who do not practise direct inheritance to sons are also more gender biased they may invest more in the education of sons. I test for this using a similar model to that used above but including only inheritance customs as an explanatory variable. As an outcome, using data from the woman's and the man's questionnaires from the DHS, I create an index based on a question which asks respondents if it is justified for a husband to beat his wife in five different scenarios. The index represents the average response to the five scenarios, with 1 representing five yes answers and 0 five no answers. I supplement the DHS with data from Afrobarometer. Afrobarometer conducts public attitude surveys on democracy, governance, the economy and society on a repeated basis in over 30 African countries. Afrobarometer rounds 3-7 include questions on the ethnicity of respondents, which allows me to match data on individuals to the ethnographic atlas in the same manner as with the DHS. As outcomes, I create binary variables reflecting whether a respondent has experienced gender discrimination, whether a respondent believes men make more suitable leaders, whether men have a greater right to a job, whether women should take care of the household and whether boys should be prioritised in education. The results of this analysis are presented in table C2. I do not find any significant evidence that individuals in no direct inheritance groups are more gender biased than those in direct inheritance groups. If anything, the opposite may be true, as individuals from no direct inheritance groups are less likely to agree that men have more of a right to a job than women or that boys should be prioritised in education.

Even if individuals' gender biases are the same in both groups, the returns to education could be relatively larger for males in no direct inheritance groups, relative to direct inheritance groups. To test this hypothesis, I take six outcomes reflecting returns to education. From the DHS, I take indicators for whether a respondent is currently working, if they work year-round and the household wealth index. From Afrobarometer, I use indicators for whether a respondent is employed, employed full-time and whether they responded positively about their living conditions relative to others. As my main explanatory variables, I include a triple-interaction between inheritance customs, gender and indicators for whether an individual has completed primary and secondary school, respectively. As the likelihood of completing primary or secondary school is not exogenous to inheritance customs, the results of this analysis should be considered as descriptive of the likely returns to schooling rather than a causal effect. The results are presented in table C3. If returns to education are larger for men relative to women in no direct inheritance groups, relative to men and women in direct inheritance groups, then we should expect to find negative and significant coefficients on *Female*Primary*No direct inheritance* and *Female*Secondary*No direct inheritance*. While there is a negative and significant coefficient on the interaction with primary school completion with regard to household wealth, none of the other interaction terms are not statistically significant. This analysis does not therefore provides much evidence in support of differential returns to education as a key mechanism.

Another alternative explanation for my findings is that groups not practising direct inheritance to sons may have differential access to education. If educational opportunities are lacking, parents may choose to focus more of their investment in boys if they perceive a greater return to investment. In addition, if attending school requires a lot of travel, parents may feel that it is less safe for girls to travel to school than for boys, leading to greater investment in sons' education. To test this hypothesis, I examine whether there are differences across inheritance groups using data from Afrobarometer on attitudes and access to education. I take as outcomes whether an individual has a school in their local public services area, whether they think school is too expensive, if they think their nearest school has poor teaching or poor facilities, whether a respondent thinks the government should prioritise education and whether they have a positive view of free education. The results of this analysis are presented in table C4. These results do not provide any evidence of differences in attitudes or access to school across inheritance groups.

Table C1: Child Labour

Dep. Var.:	Any work (1)	Paid work (2)	Fetching wood or water (3)	Household chores (4)	Other unpaid work for family (5)
Panel A: Boys only					
Brother	–0.001	–0.001	0.003	–0.004	0.002
	(0.009)	(0.005)	(0.011)	(0.011)	(0.010)
Brother*No direct inheritance	0.019	0.008	–0.031*	0.014	0.027
	(0.013)	(0.007)	(0.017)	(0.017)	(0.018)
Sample mean	.128	.0247	.565	.645	.278
N	12,216	12,216	6,871	12,026	12,194
Panel B: Girls only					
Brother	–0.001	0.001	0.019	0.010	–0.021**
	(0.009)	(0.004)	(0.016)	(0.011)	(0.010)
Brother*No direct inheritance	–0.018	–0.008	–0.026	0.014	0.016
	(0.012)	(0.005)	(0.025)	(0.017)	(0.013)
Sample mean	.129	.0178	.688	.799	.236
N	12,222	12,222	7,006	12,055	12,184
Controls	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Point estimates in each column, within each panel, are taken from the same OLS regression. Controls include dummies for age, birth year, DHS phase, interview year, country fixed effects, rural and whether the mother was married at birth, in addition to trends for the interval to sibling birth in months and mother's age at birth. Each of these controls are also interacted with gender and *No direct inheritance*.

Table C2: Gender Bias

	DHS	Afrobarometer				
Dep. Var.:	Dom. Violence Justified (1)	Experienced Gender Disc. (2)	Men as Leaders (3)	Men Have More Right to Job (4)	Women Take Care of Household (5)	Prioritise Boys in Education (6)
No direct inheritance	–0.007	0.001	–0.016	–0.032*	–0.026	–0.021**
	(0.016)	(0.009)	(0.010)	(0.019)	(0.020)	(0.010)
Sample mean	.257	.115	.369	.412	.516	.156
N	561,513	18,289	72,938	18,127	18,124	22,675
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from separate OLS regressions. Where included in the data, controls include dummies for DHS phase or Afrobarometer round, interview year, country fixed effects and rural location, in addition to trends for the age and age squared.

Table C3: Returns to Education

	DHS			Afrobarometer		
Dep. Var.:	Currently Working (1)	Working All Year (2)	Household Wealth (3)	Employed (4)	Employed Full-Time (5)	Relative Living Cond. (6)
No direct inheritance	0.057	0.012	–0.140**	–0.005	–0.015	0.041
	(0.038)	(0.033)	(0.055)	(0.013)	(0.011)	(0.048)
Female	0.116**	0.113***	0.055***	–0.107***	–0.069***	–0.027
	(0.055)	(0.032)	(0.015)	(0.016)	(0.012)	(0.019)
Female*No direct inheritance	0.001	–0.032	0.033*	0.029*	0.022*	–0.032
	(0.060)	(0.035)	(0.019)	(0.017)	(0.013)	(0.028)
Primary	–0.084***	–0.020	0.168***	0.054***	0.048***	0.149***
	(0.025)	(0.019)	(0.022)	(0.011)	(0.009)	(0.027)
Secondary	0.072***	0.097***	0.233***	0.122***	0.117***	0.244***
	(0.013)	(0.012)	(0.031)	(0.014)	(0.012)	(0.029)
Female*Primary	0.121***	0.050**	–0.036***	0.016	–0.006	0.065**
	(0.039)	(0.024)	(0.013)	(0.010)	(0.010)	(0.025)
Female*Secondary	–0.079***	–0.036**	0.017	–0.011	0.005	0.007
	(0.019)	(0.016)	(0.023)	(0.011)	(0.009)	(0.026)
Primary*No direct inheritance	–0.043	0.005	0.192**	0.005	0.011	–0.026
	(0.030)	(0.028)	(0.077)	(0.015)	(0.013)	(0.047)
Secondary*No direct inheritance	–0.015	–0.008	0.043	–0.001	0.001	–0.028
	(0.013)	(0.024)	(0.045)	(0.033)	(0.032)	(0.039)
Female*Primary*No direct inheritance	0.030	0.051	–0.070**	–0.021	–0.005	–0.020
	(0.053)	(0.035)	(0.035)	(0.014)	(0.013)	(0.040)
Female*Secondary*No direct inheritance	0.009	–0.023	–0.044	0.016	–0.015	0.007
	(0.023)	(0.020)	(0.037)	(0.019)	(0.016)	(0.048)
Sample mean	.741	.616	–3.36e-06	.337	.237	2.85
N	618,641	462,999	592,552	85,416	85,416	84,756
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from the same OLS regression. Where included in the data, controls include dummies for DHS phase or Afrobarometer round, interview year, country fixed effects, rural, in addition to trends for the age and age squared. Each of these controls are also interacted with gender and *No direct inheritance.*

Table C4: Attitudes and Access to Education

Dep. Var.:	School in local area (1)	School too expensive (2)	School has poor teachers (3)	School has poor facilities (4)	Govt. should prioritise ed. (5)	Attitude to free education (6)
No direct inheritance	0.017	–0.007	–0.025	–0.019	–0.015	0.004
	(0.011)	(0.009)	(0.021)	(0.017)	(0.014)	(0.021)
Sample mean	.873	.278	.378	.419	.372	.375
N	87,068	23,967	23,031	23,366	20,870	11,520
Controls	X	X	X	X	X	X
Fixed Effects	Country	Country	Country	Country	Country	Country

Notes: Robust standard errors, clustered by DHS sampling cluster and ethnic group in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column are taken from separate OLS regressions. Controls include dummies for age, age squared, Afrobarometer round, interview year, country fixed effects, rural, whether the respondent has primary education and whether the respondent has secondary education.

C2: Internal validity of the ethnographic atlas data

In this section, I test the internal validity of the ethnographic atlas data. Bau (2021) and Lowes (2020b) discuss various theories explaining the rise of matriliny in Africa, which both authors are able to verify using the sample of ethnic groups they examine in the ethnographic atlas. They discuss how societies relying more on animal husbandry and where bride prices are practised are more likely to practise patriliny while societies practising hoe agriculture, as opposed to plow agriculture, are more likely to practise matriliny. As discussed in section 2, matriliny is also positively correlated with extensive agriculture and hunting and gathering. It should also be expected that matrilineal kinship and matrilocal residence are correlated with matrilineal inheritance and, in turn, the no direct inheritance variable. To test these predictions, I examine the relationships between each of these characteristics and my explanatory variable. Table C5 shows the results of this analysis for the full sample of ethnic groups observed in Africa in the ethnographic atlas and for those groups present in my sample. As can be seen, the predicted relationships exist in my data, emphasising the internal validity of the variables included in the ethnographic atlas.

Table C5: Internal Validity

	Dep. Var. = No direct inheritance									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: All of Africa										
Patrilineal kinship	–0.364***									–0.064
	(0.054)									(0.074)
Matrilineal kinship		0.619***								0.470***
		(0.041)								(0.099)
Patrilocality			–0.493***							–0.051
			(0.056)							(0.124)
Matrilocality				0.597***						0.005
				(0.042)						(0.137)
Plow use					–0.364***					–0.191***
					(0.064)					(0.047)
Bride price						–0.216***				–0.005
						(0.070)				(0.069)
Animal husbandry							–1.199***			–0.751***
							(0.150)			(0.156)
Hunting, gathering								0.668		–0.128
								(0.448)		(0.265)
Extensive agriculture									0.221***	0.075
									(0.052)	(0.057)
N	346	346	346	346	321	346	346	346	346	321
Panel B: Ethnic Groups in Main Sample										
Patrilineal kinship	–0.479***									–0.117
	(0.073)									(0.111)
Matrilineal kinship		0.702***								0.523***
		(0.039)								(0.120)
Patrilocality			–0.510***							–0.148
			(0.078)							(0.292)
Matrilocality				0.564***						–0.163
				(0.071)						(0.313)
Plow use					–0.427***					–0.239***
					(0.040)					(0.084)
Bride price						–0.222**				0.008
						(0.088)				(0.084)
Animal husbandry							–1.629***			–0.884***
							(0.317)			(0.310)
Hunting, gathering								0.268		–0.387
								(0.484)		(0.270)
Extensive agriculture									0.230***	0.106
									(0.074)	(0.074)
N	170	170	171	171	164	173	173	173	173	160

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Point estimates in each column, within each panel, are taken from the same OLS regression.

Appendix D: Controlling for Confounders

In addition to the controls included in the main analysis, I control for a series of potential confounders:

Child-carer Controls

I control for whether, in addition to a child's mother, an additional potential child-carer is present in the household. As the analysis assumes parents have limited resources to invest in children's education, an additional child-carer may relax the budget constraint with regard to time investments by parents, in turn affecting the results. Specifically, for individuals who are the children, foster children or grandchildren of the household head, I designate a potential child-carer as being present if I can identify an aunt or grandmother as being present in the household.

Foster Children

I include indicators for whether a household contains a foster brother or sister who is either older than the child in my sample or born before the birth of that child's next biological sibling. Even if foster children do not affect the expected division of inheritance, it is likely that they require time investment from parents which could affect the time invested in the first-born child.

Religion

I include indicators for the largest religious groups in my sample, Christianity and Islam. It is possible that families with different religious beliefs may treat siblings of different genders differently and/or place a different emphasis on education. If religion is correlated with ethnicity, then this may be driving the results.

Mission distance

I control for the distance from each individual's village to the nearest Christian, Protestant or any religious mission.[22] Missions of all denominations enacted change in culture and customs. Protestant missions, in particular, placed greater emphasis on the education of women (Nunn, 2014). If protestant missions were more common near direct inheritance groups, this could explain the smaller

[22]The location of religious missions was mapped by Roome (1925) and was digitised by Nunn (2014).

effects of sibling gender found among direct inheritance groups.

Ethnographic Controls

As discussed in section 4.4, inheritance customs stem from kinship, which is correlated with various other characteristics of pre-colonial ethnic groups. Other research has also shown that various ethnographic characteristics affect the distribution of education resources and investment across gender directly and/or affect female empowerment today, which in turn affects education. Practising patrilineal or matrilineal kinship is highly correlated with practising direct inheritance to sons. The same can be said for patrilocal and matrilocal residence, whereby married couples and their families reside with the extended family of the husband or wife, respectively, which has been shown by Bau (2021) to affect educational investment across gender. Similarly, kinship tightness also varies with unilineal descent systems and has recently been shown to affect beliefs and culture (Enke, 2019). To test that the results are not driven by these channels, rather than by inheritance customs, I control for patriliny, matriliny, patrilocality, matrilocality and kinship tightness.[23] I control for whether ethnic groups practise a bride price, which affects female education (Ashraf et al., 2020), and whether ethnic groups practised plow agriculture before colonisation, which affect present-day gender roles (Alesina et al., 2013), along with indicators for whether females were the main performers of an ethnic group's primary economic activity, if the performance of that main activity was mixed between men and women, whether a group's main agricultural crop was a cereal grain and whether ethnic groups have a tradition of polygyny and continuous measures of community size, the number of hierarchical segments in a group's historical societies and pre-colonial reliance on animal husbandry.

Ethnic fractionalisation and polarisation

It is possible that the degree of ethnic homogeneity in an area can affect the extent to which families maintain traditional ethnic customs. For example, families who wish to break from the traditions of their ethnic group may be more willing to do so in more heterogeneous societies, where traditions may be less strong (Atkin et al., 2021). Similarly, families who wish to break from tradition but who live in more ethnically polarised societies may perceive a greater cost

[23]I follow Enke (2019) in defining kinship tightness as the average of four variables defined using data from the ethnographic atlas. These are indicators equalling one if an ethnic group is not observed to practise bilateral kinship (which applies to everyone in my sample), equalling one if an ethnic group is observed practising either matrilocality or patrilocality, equalling one if an ethnic group is not observed to practise residence in the form of nuclear families and equalling one if an ethnic group is observed to live in segmented communities or clan barrios.

to doing so as to break from tradition may be seen as turning one's back on their group. This means that the strength of the relationship between ancestral inheritance rules and outcomes may be correlated with ethnic homogeneity in a society. I thus control for ethnic fractionalisation and polarisation at both the village and country-year level. These are defined by Montalvo and Reynal-Querol (2005). Ethnic fractionalisation is measured using a Herfindahl index, which estimates the probability that any two individuals selected at random from the same sampling cluster will not belong to the same ethnic group. It is calculated using the formula $1 - \Sigma_{i=1}^{N} \pi_e^2$ where π represents the proportion of a population from ethnicity e and N the number of ethnic groups present in a population. Ethnic polarisation measures how far the distribution of ethnic groups in a population deviates from a (0.5,0.5,0,0,...,0) distribution, which represents the maximum level of polarisation. It is calculated using the formula $1 - \Sigma_{i=1}^{N} (\frac{0.5-\pi_i}{0.5})^2 \pi_i^2$.

Ethnic-group wealth

I control for mean household wealth at the ethnic group level (excluding the focal household), as a group's relative affluence may affect social customs with regard to investment in children's education.

Geo-spatial controls

As can be seen in figure 4, there exists apparent spatial correlation in the inheritance rules observed among different ethnic groups. In an attempt to ensure that inheritance rules are not correlated with some other spatially correlated factors affecting education outcomes, both on average and via interactions with sibling gender, I include a battery of geographic and economic covariates. These are provided at the village level by the DHS. Specifically, I control for latitude, longitude, altitude, ground slope, length of the growing season (in months), malaria prevalence (in 2010), mean temperature (2010), rainfall (2010), population density (in 2010), a nightlight composite to proxy for economic activity, travel time to the nearest large city (in 2015), distance to an international border and distance to the nearest body of water.

Slave trade exposure

Recent research has pointed out further channels affecting gender roles along ethnic lines. Specifically, Teso (2019) shows how exposure to the transatlantic slave trade led to increased female empowerment today. I thus control for exposure to the transatlantic slave trade and its interaction with sibling gender. Using data from Nunn and Wantchekon (2011), I measure slave trade exposure as the natural log of the number of males taken from an ethnic group divided by its ancestral land area.

Crop Yields

Demie (2018) shows how societies producing cereal grains as their main crop exhibit less female empowerment today. Using data from the Global Agro-Ecological Zones project, I control for the difference in the potential yield (measured in millions of kilo-calories per hectare per year) of the best cereal crop and best root or tuber crop that can be harvested on an ethnic groups ancestral homeland, as defined by the Murdock (1959) map.

Appendix E: Data Appendix

E1: Matching Ethnicities in the Democratic Republic of Congo

The Demographic and Health Survey (DHS) categorises ethnic groups in the Democratic Republic of Congo into large geographic groups, based on shared cultures. These are "Kasaï-Katanga-Tanganyika", "Basele-Komo, Maniema et Kivu", "Bas-Kasaï et KwiluKwango", "Ubangi et Itimbiri-Ngiri", "Cuvette Centrale", "Bakongo du Nord et du Sud", "Uele-Lac Albert", "Lunda" and "Pygmèes".

While I am unable to match the "Lunda" and "Pygmèes" groups to the ethnographic atlas, I match "Bakongo du Nord et du Sud" directly to the "Kongo" group in the ethnographic atlas. As the ethnographic atlas provides geographic co-ordinates of ethnic groups, I categorise groups located within the modern-day national border by location to match the categories used by the DHS and assign inheritance rules for real property according to the majority rule within each category. These are assigned as follows:

1. To the "Basele-Komo, Maniema et Kivu" category, I assign all groups located within the provinces of Maniema, North Kivu and South Kivu.
2. To the "Cuvette Centrale" category, I assign all groups located in the Cuvette Central region, which is the region bordered to the west, north and south by the Congo river.
3. To the "Ubangi et Itimbiri-Ngiri" category, I assign all groups located north of the Congo river, to the south and east of the Ubangi river and to the west of both the point where the Itimbiri and Congo rivers meet and the point where the Ubangi and Uele rivers meet.
4. To the "Uele-Lac Albert" category, I assign all groups located east of the point where the Ubangi and Uele rivers meet, to the west of Lake Albert and to the north of the provinces of Maniema and North Kivu.
5. To the "Kasaï-Katanga-Tanganyika" category, I assign all groups located in the provinces of Kasaï, Kasaï-Central, Kasaï-Oriental, Katanga and Tanganyika and all areas in between these provinces which are not part of the provinces of Maniema and South Kivu.
6. To the "Bas-Kasaï et KwiluKwango" category, I assign all groups located in the provinces of Kwilu and Kwango and all groups located in the vicinity

of the lower Kasaï river south of the Cuvette Central region and any remaining groups to the south-west of the country.

This will naturally lead to some misallocation of ethnic groups in the ethnographic atlas to categories in the DHS. Using majority rule to classify inheritance rules within each category should reduce the likelihood of errors in classifying inheritance rules for DHS categories. The geographic locations of these groups and the categories to which they are assigned are presented graphically in figure E1

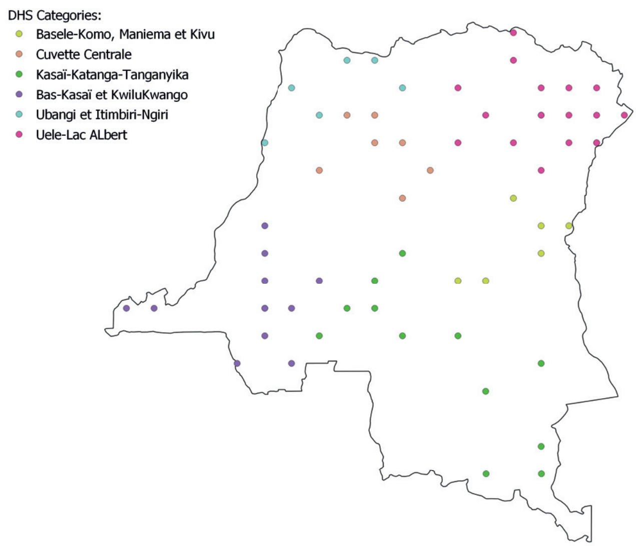


Figure E1: Ethnic Groups in the Ethnographic Atlas, matched to DHS categories

0.1 E2: Surveys Included in Sample

Survey data on educational outcomes for individuals is provided from various Demographic and Health Surveys (DHS). I use data from all surveys from which children observed in the household questionnaire could be matched to their mother's birth history from the women's questionnaire and from which their mother reported being from an ethnic group which is observed to practive a matrilineal or patrilineal inheritance rule in the Ethnographic Atlas. Table E1 thus provides information on the countries, DHS phases and interview years of each survey used.

Table E1: Surveys included in estimation sample

Country	Phase	Interview Year	Country	Phase	Interview Year
Benin	4	2001	Malawi	4	2000
Benin	5	2006	Malawi	4	2004 - 2005
Benin	6	2011 - 2012	Malawi	6	2010
Benin	7	2017 - 2018	Malawi	7	2015 - 2016
Burkina Faso	4	2003	Mali	4	2001
Burkina Faso	6	2010	Mali	5	2006
Burundi	6	2010 - 2011†	Mali	6	2012 - 2013
Burundi	7	2016 - 2017†	Mali	7	2018
Cameroon	4	2004	Mozambique	6	2011
Cameroon	6	2011	Namibia	4	2000
Cameroon	7	2018 - 2019	Niger	5	2006
Chad	7	2014 - 2015	Nigeria	5	2008
Republic of Congo	5	2005	Nigeria	6	2013
Republic of Congo	6	2011 - 2012	Nigeria	7	2018
Congo Democratic Republic	5	2007	Rwanda	4	2000†
Congo Democratic Republic	6	2013 - 2014	Rwanda	5	2005†
Ethiopia	4	1992	Rwanda	6	2010 - 2011†
Ethiopia	5	1997	Rwanda	7	2014 - 2015†
Ethiopia	6	2003	Senegal	4	2005
Ethiopia	7	2008	Senegal	6	2010 - 2011
Gabon	6	2012	Senegal	6	2012 - 2013
The Gambia	6	2013	Senegal	7	2014
The Gambia	8	2019 - 2020	Senegal	7	2015
Ghana	4	2003	Senegal	7	2016
Ghana	5	2008	Senegal	7	2017
Ghana	7	2014	Senegal	8	2018
Guinea	4	1999	Senegal	8	2019
Guinea	5	2005	Sierra Leone	5	2008
Guinea	6	2012	Sierra Leone	6	2013
Guinea	7	2018	Sierra Leone	7	2019
Ivory Coast	6	2011 - 2012	Togo	6	2013 - 2014
Kenya	4	2003	Uganda	6	2011
Kenya	5	2008 - 2009	Uganda	7	2016
Kenya	7	2014	Zambia	4	2001 - 2002
Liberia	6	2013	Zambia	5	2007
			Zambia	6	2013 - 2014

Notes: †Ethnic information for Burundi and Rwanda was imputed to the rundi and ruanda ethnic groups, respectively, as these are the predominant groups in each country.

E3: Reforms to Inheritance Law

This section outlines the national laws used to identify reforms to land inheritance rules, as discussed in section 5.[24]

Benin

Loi N° 2002-07: Portant Code des personnes et de la famille, Art. 619

Article 619

Les enfants ou leurs descendants succèdent à leurs père et mère ou autres ascendants sans distinction de sexe ni d'âge encore qu'ils soient issus de différents mariages, sous réserve des dispositions prévues au présent code relativement aux enfants incestueux.

[Translated to English:

Children or their descendants succeed their father and mother or other ascendants without distinction of sex or age, even if they are from different marriages, subject to the provisions of this code relating to incestuous children.]

Mali

Loi No. 2011-087 Portant Code des Personnes et de la Famille, Arts. 772, 773

Article 772

Les parents en l'absence de conjoint successible, sont appelés à succéder ainsi qu'il suit:

1. les enfants et leurs descendants;
2. les père et mère; les frères et sœurs et les descendants de ces derniers;
3. les ascendants autres que les père et mère;

[24] For laws that are not originally written in English, translations are made by the author and presented in square parentheses.

4. les collatéraux autres que les frères et sœurs et les descendants de ces derniers.

A l'exception des père et mère qui héritent du dixième, chacune de ces quatre catégories constitue un ordre d'héritiers qui exclut les suivants.

[Translated to English:

Parents in the absence of a succeeding spouse are called upon to succeed as follows:

1. the children and their descendants;
2. the father and mother; siblings and descendants of the latter;
3. the ascendants other than the father and mother;
4. the collaterals other than siblings and their descendants.

With the exception of the father and mother who inherit one tenth, each of these four categories constitutes an order of heirs which excludes the following.]

Article 773

Les enfants ou leurs descendants succèdent à leurs père et mère ou autres ascendants, sans distinction de sexe, ni de primogéniture, même s'ils sont issus d'unions différentes.

[Translated to English:

Children or their descendants succeed their father and mother or other ascendants without distinction of sex or primogeniture, even if they come from different unions.]

Rwanda

Law N°27/2016 of 08/07/2016 Governing Matrimonial Regimes, Donations and Successions, Arts. 54, 73

Article 54: Equal treatment of children in succession

Legitimate children of the de cujus succeed in equal portions without any discrimination between male and female children.

Article 73: Order of regular heirs

Heirs are entitled to inherit in the following order:

1° children of the de cujus;

2° father and mother of the de cujus;

3° full-blood brothers and sisters of the de cujus;

4° half-brothers and half-sisters of the de cujus;

5° grandparents of the de cujus;

6° paternal and maternal uncles and aunts of the de cujus;

Subject to provisions of Article 41 of this Law, each category of successors excludes others in the order of succession.

Full-blood children of the de cujus inherit from both the paternal and maternal sides, while consanguineous and uterine children inherit only from the side of the parent to whom they are related.

Sierra Leone

The Devolution of Estates Act, 2007, Secs. 7, 8

Section 7: Intestate survived by child only

(1) Subject to subsection (2) and section 15, where an intestate is survived by one child and no spouse, parent or grandchild, the whole of the estate shall devolve to the surviving child.

(2) Where an intestate is survived by two or more children and no spouse, parent or grandchild, the estate shall devolve to the children in equal shares.

Section 8: Intestate survived by spouse, child and parent

Where the intestate is survived by a spouse, child and parent, the estate shall devolve in the following manner:-

(a) thirty five percent to the surviving spouse;

(b) thirty five percent to the surviving child;

(c) fifteen percent to the surviving parent;

(d) fifteen percent in accordance with customary law or Muslim law, as applicable.

Zambia

The Intestate Succession Act, Sec. 5, 6(a), 7

Section 5: Distribution of estate

(1) Subject to sections eight, nine, ten and eleven the estate of an intestate shall be distributed as follows:

(a) twenty per cent of the estate shall devolve upon the surviving spouse; except that where more than one widow survives the intestate, twenty per cent of the estate shall be distributed among them proportional to the duration of their respective marriages to the deceased, and other factors such as the widow's contribution to the deceased's property may be taken into account when justice so requires;

(b) fifty per cent of the estate shall devolve upon the children in such proportions as are commensurate with a child's age or educational needs or both;

(c) twenty per cent of the estate shall devolve upon the parents of the deceased;

(d) ten per cent of the estate shall devolve upon the dependants, in equal shares:

Provided that a priority dependant whose portion of the estate under this section is unreasonably small having regard to his degree of dependence on the deceased shall have the right to apply to a court for adjustment to be made to the portions inherited and in that case, Part III of the Wills and Administration of Testate Estates Act shall apply, with the necessary changes, to the application.

(2) In respect of a minor, the mother, father or guardian shall hold his share of the estate in trust until he ceases to be a minor.

Section 6: Distribution where intestate survived by no spouse, etc.

Where an intestate leaves-

(a) no spouse, the portion of the estate which the spouse would have inherited shall be distributed to the children in such proportions as are commensurate with a child's age or educational needs or both; Distribution where intestate survived by no spouse, etc.

(b) no spouse or children; the aggregate portion of the estate which the spouse and children would have inherited shall be distributed equally to the parents of the deceased;

(c) no spouse, children or parents, the estate shall be distributed to dependants in equal shares;

(d) no spouse, children, parents, or dependants, the estate shall be distributed to near relatives in equal shares;

(e) no spouse, children, parents, dependants or near relatives, the estate shall be bona vacantia and shall devolve upon the State;

Section 7: Distribution where intestate survived by spouse, etc.

Where an intestate leaves-

(a) a spouse, children, dependants but no parents, the proportion of the estate which the parents would have inherited shall be shared equally between the surviving spouse and children on the one hand and the dependants on the other;

(b) a spouse, parents, dependants but no children, the portion of the estate which the children would have inherited shall be distributed to the surviving spouse, parents and dependants in proportion to their shares of the estate as specified in section five;

(c) a spouse, children, parents but no dependants, the portion which the dependants would have inherited shall be distributed equally to the parents;

(d) a spouse and dependants but no children or parents, the portion of the estate which the children and parents would have inherited shall be distributed to the surviving spouse and the dependants in proportion to their shares of the estate as specified in section five;

(e) a spouse and children but no parents or dependants, the portion of the estate which the parents and dependants would have inherited shall be shared equally among the surviving spouse on the one hand and the children on the other;

(f) a spouse but no children, parents or dependants, the portion of the estate which the children, parents and dependants would have inherited shall be distributed equally between the surviving spouse on the one hand and the near relatives on the other.

E4: Introduction of Free Primary Education

This section outlines the sources used to identify the introduction of free primary education, as discussed in section 6.

Benin

According to Engel et al. (2011):

"The gross admissions rate for Grade 1 was over 150% in 2007/08, a result of the abolition of school fees in the previous year", implying that tuition fees were removed for the 2006/07 school year.

Burkina Faso

According to Kouraogo and Dianda (2008):

"In 2007 the government launched a general education reform that should bring among other innovations an extension of basic education from the current six years to ten years, and generalise progressively free education for children aged 6-16."

Burundi

According to Sommeiller and Wodon (2014):

"Following two decades of conflict and after a process of reconciliation that lasted several years, the newly elected President of Burundi declared in 2005 that primary education in public schools would be provided for free. The policy became effective starting with the 2005-06 school year."

Cameroon

According to Kamga (2011):

"In assessing free primary education, the General State of Education Workshop held in May 1995 in Yaoundé, Cameroon, provided a general consensus calling for free and compulsory basic education for all. As a result, the principle of free primary education was underlined by the government's order of February 1996 that organises education in the country, and was translated into the Finance Law 2000/8"

Democratic Republic of Congo

According to World Bank (2020):

"To tackle these challenges, the DRC launched a sweeping reform, introducing free primary education throughout the country as of September 2019."

Republic of the Congo

According to United Nations (2012):

"Following the announcement made by the head of State in his end-of-year speech in 2007, an order signed jointly by the Ministers of Finance and Budget, Technical and Vocational Education and Primary, Secondary and Literacy Education (No. 278/MEFB/METP/MEPSA of 20 March 2008) put into effect the constitutional provisions on free primary and secondary education."

Ethiopia

According to Birger and Craissati (2009):

"Approach and year of fee abolition: "Big Bang" in 1994. Instructions to schools provided one year after decision."

Ghana

According to Birger and Craissati (2009):

"Approach and year of fee abolition: 2005; scaling up of pilot started in 2003 for deprived districts"

Ivory Coast

According to Oyeniran (2018):

“Just recently in 2015-16, the state enacted grants free tuition in the basic education and colleges in the country. ‘Free’ education denotes that tuition fees will be waived through government funding.”

Kenya

According to Birger and Craissati (2009):

“Approach and year of fee abolition: “Big Bang” in January 2003 followed December 2002 election.”

Liberia

According to Ministry of Forreign Affairs, Liberia (2011):

“This level of education, which consists of full-time formal schooling that is provided for children from age six (6) to age twelve (12), and constituting grades 1-6, shall be free and compulsory for all children of the age range for such school level, and shall be free for all pupils within the public school system;”

Malawi

According to Birger and Craissati (2009):

“Approach and year of fee abolition:“Big Bang” in 1994 followed pledge during first multiparty election, although partial fee removal was introduced in 1991 and 1992.”

Mozambique

According to Birger and Craissati (2009):

“Approach and year of fee abolition: Decision in 2003 became effective in 2004 after testing. Phased implementation of direct support to schools 2004–06.”

Niger

According to United Nations (2001):

“The Constitution is complemented by legislative and regulatory instruments (decrees, laws, orders) which constitute the legal framework for education. Act No. 98-12 of 1 June 1998, which sets out the aims of the education system, states that formal education is a means of acquiring education and vocational training in a school setting. The Act sets forth the right of the child to education and the obligation of the State to make primary education compulsory and free. ”

Rwanda

According to Government of Rwanda (2003):

“Primary school education is compulsory and free both in public and government aided schools. Free education refers to free access to learning, teaching aid as well as basic textbooks needed by pupils and teachers.”

Senegal

According to Ndiaye (2012):

“Although, the first president’s educational policies were strongly based on the importance of French as the language of instruction, he attempted to reform the colonial system under the guideline of l’enracinement and l’ouverture (Sylla, 1993). This reform was presented as a multicultural policy that would “ensure an awareness of firm roots [and] simultaneously incorporate universal values and civilization” (Sylla, 1993:376). This new law was also meant to be democratic in the sense that it was meant to provide all citizens with free education, recognize their rights to equal educational opportunities (Sylla, 1993) and ensure that the curriculum would be relevant to the citizens’ lived experiences.”

Sierra Leone

According to Government of Sierra Leone (2018):

“To attain the set goals, the Government has decided that basic education in Sierra Leone should be ‘free and compulsory’ to the extent stated in the Education Act of 2004.”

Togo

According to Hoogeveen and Rossi (2019):

"Encouraged by the success of the 2007 elections and the new government's reform platform, which included the abolition of school fees starting in the 2008/2009 school year and the gradual integration of EDIL schools in the public school system, donors reengaged with the country after more than 15 years of providing limited assistance."

Zambia

According to Riddell (2003):

"Country: Zambia; FPE Provision: February 2002. user fees abolished. uniforms not compulsory. fees can be levied by PTAs and boards, but no student can be denied an education because of cost"

Paper III

The Long-Term Consequences of More Informative Grading

Co-authored with Jonas Lundstedt[25]

Abstract

We study the effect of more informative feedback on student performance. Using data on the population of Swedish school children, we exploit a reform to the grading system in compulsory school which introduced a more granular grading scale and thus provided students with more informative feedback on their academic performance. Exploiting a difference-in-discontinuity research design, we find that students exposed to more informative grading were less likely to graduate from high school and from an academic high school track. The likelihood of a student graduating from a STEM high school track or enrolling in a STEM track in university also decreased as a result of more informative grading. We estimate that this caused average yearly income to decrease between the ages of 28 and 30. These results appear to be driven by a negative shock to students' self-belief and increased stress levels.

Keywords: grading; feedback; educational outcomes; natural experiment; HBSC
JEL Classifications: I21, I28, I12

[25]Lund University, Centre for Economic Demography

1 Introduction

Performance feedback is prevalent in many different settings, including in the workplace and throughout education, and may affect individuals' incentives to exert effort (Bénabou and Tirole, 2002; Thaler et al., 2013). In education, it is commonplace for students to receive feedback in the form of grades. A small but growing literature studies whether receiving any feedback improves educational performance, with varying results (Bandiera et al., 2015; Sjögren, 2010). Yet, we have little knowledge about how the type and, in particular, the precision of feedback affects performance, despite the widespread use of grading feedback in today's schools.

There are many reasons why more informative grading could matter for an individual's educational outcomes. If people are overconfident, like a large literature suggests, more informative grading may lead to reduced self-belief (Möbius et al., 2014), lower effort (Fischer and Sliwka, 2018) and increased dropout rates (Stinebrickner and Stinebrickner, 2014).[1] On the other hand, more informative grading should provide students with more accurate information on their comparative advantages, helping students to find better matches in education and the labour market (Bobba and Frisancho, 2016; Chevalier et al., 2009). Moreover, more informative feedback reduces information frictions not only for students but also their parents, allowing parents to make more appropriate investments in their children's human capital (Cobb-Clark et al., 2021; Dizon-Ross, 2019).

Studying the effect of more informative grading is challenging. Comparing school systems with different types of grades is not informative, as there is a myriad of institutional, cultural and demographic factors affecting outcomes that vary and may be difficult to account for. We contribute to the literature by exploiting a natural experiment, where a reform to the compulsory school grading system in Sweden in 2011 led to the introduction of more informative grades. This allows us to examine the causal effect of receiving more informative feedback on an absolute scale in 8th and 9th grade on education and labour market outcomes.

The reform affected the information content of grades by replacing the previous scale, which included three passing grades, with a more granular scale with five passing grades. As children in Sweden are assigned to school cohorts based on their date of birth, we are able to exploit this reform in a difference-in-discontinuity research design. Children born just after January 1st, 1997 were

[1] See Zimmermann (2020) for a detailed review of literature on overconfidence.

exposed to more informative grading, while those born just before were not. Those born just after the admissions cut-off date are also subject to school-starting age effects which arise across the January 1st assignment cut-off in every year (Black et al., 2011; Fredriksson and Öckert, 2014). By comparing the difference in outcomes between those born just before and just after the admissions cut-off of January 1st, 1997 to the difference in outcomes of those born just before and just after the January 1st admissions cut-offs from 1992 to 1996, we are able to separate the confounding effect of school starting age and identify the effect of exposure to more informative grading.

In our analysis, we use administrative data covering the universe of compulsory school students in Sweden born between 1991 and 1997, who are tracked from birth until 2017. With these data we can follow students up to the age of 21, regardless of whether they remain in full-time education or not. At high school level, our main outcome variables include the likelihood of graduating and of graduating from specific tracks. We also study the university level and examine the likelihood of enrolment by age 21, type of university enrolled in, credits enrolled in and passed and field of study enrolled in.

We find that exposure to more informative grading has negative consequences for educational outcomes, with treated students being 3.3% less likely to graduate from high school and 7.6% and 11.9% less likely to complete an academic track or an academic STEM track in high school, respectively. While we do not find evidence of reduced university enrolment, we do find a 16.7% decrease in the likelihood of continuing to a STEM track in university. We do not find evidence of heterogeneity by gender, family income, parental education or immigrant status. Using the surrogate index approach proposed by Athey et al. (2019), we find that exposure to more informative grading leads to an approximate 1.8% reduction in average yearly earnings between ages 28-30.

To investigate potential mechanisms underlying our results, we use survey data from an internationally representative sample of school children to compare treated Swedish students to their counterparts in Denmark and Norway, finding that treated children are less likely to see themselves as performing well or very well in school. Treated children are also more likely to report feeling pressure from their school work and to report experiencing sleeping difficulties. This suggests that the reductions in high school graduation and STEM participation are due to reduced confidence in one's own academic ability and increased stress.

We examine possible threats to our identification strategy. As our identification is based on school cohort-specific exposure to more informative grading, we may be picking up a simultaneous cohort-specific idiosyncratic shock that affects outcomes. If so, it is likely that similar shocks also occur at other times, affecting other school cohorts. This should lead us to identify significant differences in outcomes across other school cohort assignment cut-offs, relative to other control years. We therefore run a set of placebo tests, using other admission cut-offs as the treatment. We also re-estimate the p-values of our estimated treatment coefficients using randomisation inference, using each possible month in our data as placebo cut-offs. These tests do not provide evidence that our results are confounded by any simultaneous shocks. In addition, we show that our results are robust to alternative bandwidths, functional forms, kernel weights and to the inclusion of a battery of covariates.

Our study contributes to the empirical literature on absolute performance feedback in education. At the extensive margin, the provision of absolute grading feedback at university is found to have a positive effect on future performance (Bandiera et al., 2015). Examining the timing of feedback, Fischer and Wagner (2018) find that receiving feedback within a couple of days of an exam led to improved results, relative to receiving feedback just before the next exam. Regarding the granularity of grading scales, Jalava et al. (2015) examine, in an experiment involving 6th grade students, how the grading scale used to grade a test affects performance. Performance was measured by testing students immediately after informing them of the grading scale on which they would be evaluated, meaning that the treatment examined was that of the intrinsic motivational power of the grading scale. Relative to a continuous score ranging from 0 to 22, students graded using A-F letter grades performed better, but not statistically significantly so. The main contribution of this paper is to provide the first evidence on the effects of receiving more informative grades in school on longer-term educational and labour market outcomes.

This paper also contributes to the broader empirical literature on how grades can be leveraged to improve educational performance. More recently, this literature largely examines the benefits of relative grading systems, finding mixed results. At the university level, relative grading is found to have a positive effect on performance (Azmat et al., 2019; Dobrescu et al., 2021). At the high school level, Azmat and Iriberri (2010) find positive effects that dissipate after relative grading feedback was no longer provided, while Goulas and Megalokonomou (2015) find that relative grading improved the performance of higher ability students but had a negative effect on lower ability students. Also at high school level, Fischer and Wagner (2018) find no difference between absolute and relative grading. In adult education, Tran and Zeckhauser (2012) find that public rank information improved outcomes for higher ability students. We contribute to this literature by examining effects on individuals who are treated between the ages of 13 and 15, ages at which individuals' responses to treatment may be quite different to those treated at the high school and university levels.

At a more general level, our paper contributes to the broader literature on performance feedback. This literature spans many aspects of feedback, including whether feedback is provided at all (Bradler et al., 2016), whether feedback is presented on an absolute or a relative basis (Ashraf et al., 2014; Blanes i Vidal and Nossol, 2011; Eriksson et al., 2009; Gill et al., 2019), whether feedback is presented publicly or privately (Hannan et al., 2013; Tafkov, 2013), how the effects of feedback vary according to whether workers are compensated based on performance (Azmat and Iriberri, 2016) and how effects vary according to the likelihood of receiving feedback and benchmarks used for relative performance (Kuhnen and Tymula, 2012). We contribute to this literature by using administrative data to examine the long-term effects of a nationwide reform.

The structure of the paper is as follows: in the next section, we describe the institutional context of the Swedish education system and the natural experiment we exploit for identification. In section 3, we describe the data and in section 4 we outline the empirical strategy we employ. Section 5 provides our estimates for the effects of more informative grading, where we examine outcomes at high school, university and earnings later in life. In section 6 we examine possible mechanisms for our effects, while section 7 then concludes.

2 Institutional Context

2.1 The Swedish education system

The Swedish school system consists of nine years of compulsory school covering grades one through nine. During the period of interest for this paper, students received grades from eighth grade onward. Students enter first grade in the autumn of the year they turn seven and thus graduate in the year they turn 16. Parents with strong preferences toward school starting age may enrol children in school with a different birth cohort, although in practice this is rare. It is also possible that a student will enrol in education in the assigned year but not pass through the entirety of compulsory school in that cohort. This could be due to, for example, students repeating a year due to slower development or personal circumstances. The decision to move a child to a different school cohort can only be made by the principal of a school, following discussions with a child's primary caregivers. Just over 6% of the individuals in our sample graduate in a different school cohort.

After nine years of compulsory school, students can apply to high school, which comprises grades 10 to 12. Approximately 99% of students continue to high school although only around 79% of students graduate, with 70% doing so within three years. Students can choose between 18 national programs of which six are academic and twelve vocational. All programs comprise three years of courses and admission is based on students' grades in compulsory school. Various preparatory programs also exist with the aim of preparing non-eligible students for these national programs. While municipalities are responsible for the funding and provision of compulsory school and high school education, the government decides on the overall goals of the educational system.

The admissions system to higher education in Sweden is centrally organised by the Swedish Council for Higher Education (SCHE). Students apply by submitting a rank-ordered preference list over desired school and program combinations. Admission is then based on two distinct quotas, one based on students' high school grade point averages and another based on points received on the Swedish Scholastic Assessment Test (SweSAT). Roughly 30% of students accepted to higher education are accepted via the SweSAT quota. Taking the SweSAT is optional and is administered outside schools by SCHE. Higher education is free of charge and students have the possibility to receive generous maintenance loans and grants from the state.

2.2 Reform to the grading scale

In 2011, the "Curriculum for the compulsory school, preschool class and the recreation centre" (Lgr 11) was implemented, affecting all students in compulsory school in Sweden.[2] A more informative grading scale in compulsory school was implemented as a part of the new curriculum. Prior to the reform, students were graded on a scale with three passing grades, namely *godkänt* (≈ pass), *väl godkänt* (≈ pass with distinction) or *mycket väl godkänt* (≈ pass with special distinction). After the reform, students were graded on an A-F scale with five passing grades, i.e. A-E. The change in the grading scale was implemented in the academic year of 2011/12 and affected students entering eighth grade that year while the students entering ninth grade continued to receive grades on the old scale throughout their last year in compulsory school (SFS, 2010). Since students are assigned to school cohorts according to birth year, this implied that students born in 1996 and earlier received grades on the old scale throughout compulsory school, while students born in 1997 and later received letter grades.

[2]In Swedish: *Läroplan för grundskolan, förskoleklassen och fritidshemmet*

The introduction of the new curriculum instituted a larger reform of the Swedish compulsory school system, including the introduction of a new syllabus for each course, the introduction of more informative grades and the introduction of grading from an earlier age. These distinct components of the new curriculum were implemented in a stepwise manner, allowing us to isolate the effect of more informative grading.

The phasing in of the new curriculum began in 2011. First, the new syllabi were introduced for all students in compulsory school during the 2011/12 academic year, which corresponds to students born in 1996 or after. This implies that some of our comparison group (i.e. those born during 1996) were affected by the new syllabi. The new syllabi did not affect student learning, however. The goal of the introduction of new syllabi was to ensure greater clarity, both for students and teachers, and to make the syllabi more comparable across courses. In the official report that preceded Lgr 11, the old syllabi were criticised for containing too many "Imprecise and generally formulated expressions...", with the report emphasising that any future syllabi should be more easily understood (SOU, 2007). This means that the syllabi did not change in practice. In appendix tables B1 to B3, we present a comparison between various knowledge requirements from the old and new syllabi for 9th grade Swedish to show the similarities between the two. The Swedish syllabi do not regulate what textbooks are to be used in schools. Furthermore, the new syllabi did not affect teaching styles. From a survey of 1,887 teachers of 6th and 9th grade classes, Wahlström and Sundberg (2015, p. 50) conclude that the new syllabi encouraged forms of teaching that, according to teachers' interpretations, corresponded to the teaching practices they already used. As a robustness check, we use a placebo test to show that the new syllabi did not have any affect on the education outcomes we examine.

Second, the more informative grading scale introduced a scale with more grade thresholds, with the goal of motivating more students to work hard and reach those thresholds (Utbildningsdepartmentet, 2008). The old and new scales both specified knowledge requirements for different grades, one for each of the grades in the old system and one for each of the grades E, C and A in the new system. The grades D and B did not have any specific requirements of their own but were awarded to students whose performance was in between the specified steps. The formulations of the requirements are similar in both the new and the old system where pass corresponds to E, while pass with distinction and pass with special distinction possess similar knowledge requirements to C and A, respectively. However, under the new grading scale, students must fulfil all individual knowledge requirements in order to reach a certain grade level, while under the old grading scale, excellence in some requirements could compensate for deficiencies in others. This means that it is more difficult to achieve an A or C grade under the new scale, in comparison to a pass with special distinction or pass with distinction under the old scale, respectively. Consider an example where a course were to include four knowledge requirements and in which a student would receive three A grades and one C grade for those requirements. Assuming that student performed excellently in some or all of the three requirements in which they achieved an A standard, that student would likely receive a pass with special distinction under the old scale but a B grade under the new scale. According to the Swedish national agency for education, the passing requirements remained unchanged (Utbildningsdepartementet, 2009).

A comparison of the standard of achievement required for each grade level under the old and new scales is presented in appendix figure B1. As can be seen, while the standard required to achieve pass and E grades are equal, D and C grades are at lower and higher standards than pass with distinction, respectively, and the same for B and A grades, relative to pass with special distinction. As the post-reform distribution of grades would reflect both mechanical changes in the distribution caused by the change of scale and any treatment effect of more informative grading, it is not possible to compare achievement in the form of compulsory school grades before and after the reform.

Third, grading from an earlier age was introduced in the 2012/13 academic year, when the cohort born in 1999 started receiving grading feedback from seventh grade and all cohorts born in 2000 and onward received grading feedback from sixth grade. Students born in these years are not included in our sample.

In summary, students born up and including 1995 received grading feedback on the old scale in 8th and 9th grade and were subject to the old syllabi throughout compulsory school. Students born in 1996 switched from the old to the new syllabi in 9th grade but continued to receive grading feedback on the old scale in 8th and 9th grade. Students born in 1997 switched from the old to the new syllabi in 8th grade and received grading feedback on the new scale in 8th and 9th grade. Appendix table B4 details the introduction and roll-out of the new grading scale across cohorts.

3 Data

We rely on data from the Swedish Interdisciplinary Panel, which is administered by the Centre for Economic Demography.[3] Our empirical analyses are based on a comprehensive dataset covering all Swedish residents who attended compulsory school during the reform period. This dataset was created by merging a number of administrative registers using anonymised individual identifiers. To construct our database, we begin by merging the Total Population Register to information on compulsory school grades in the National Agency for Education's Pupil Register to identify all compulsory school students in Sweden born between 1991 and 1997, who are tracked from birth until 2017. This is further merged with information on high school leavers and high school grades in the National Agency for Education's Pupil Register, information on university enrolment and university credit from the University and Higher Education Register and outpatient and inpatient health records from the National Patient Register. In addition, the Multi-Generation Register is used to match students to their siblings and parents, both biological and/or adoptive. The Education register and information on earned income from the Income and Taxation Register are merged to parental identifiers to attain social and demographic characteristics of each student's family.

The high school leavers register provides us with information on track choice for all students who either graduate from high school or leave with formal documentation of the education they have completed but did not achieve the requirements for graduation. This includes the specific academic or vocational track undertaken. We assign academic tracks to three groups, the first being STEM tracks, which comprise the technical and natural science tracks. The second group we define as the grouping of the humanities, economics and social science tracks, leaving the aesthetics track as the only art track.[4] As students are recorded in this register only upon leaving high school, for those students who do not complete high school, we do not know if they chose not to attend or if and when they dropped out of high school. In the case of students who change school or track, we observe only the track from which they graduated. Therefore, high school outcomes represent whether a student completed high

[3]The data used in this paper come from the Swedish Interdisciplinary Panel approved by the Regional Ethics Committee in Lund (2012/03). Data are available to researchers affiliated to the Centre for Economic Demography at Lund University working with issues falling under the research program "Sambandet mellan kritiska perioder tidigt i livet och socioekonomiska samt demografiska utfall" (cleared by the Regional Ethics Committee).

[4]Individuals in vocational tracks as well as those not observed in the high school leavers register are assigned values of zero for these outcomes, while those in preparatory programs and special education programs are excluded from our analysis.

school or completed a specific high school track, not whether a student attended. While the high school leavers register does include information on grades, we do not include these in our analysis. As we discuss below, the grading scale in high school was reformed for cohorts born in 1995 and onward, meaning that we only have comparable grades for one birth cohort prior to our treatment cohort, which prevents us from testing the robustness of our results to simultaneous shocks. In addition, we find effects on both high school completion, which determines selection into our sample, and track composition, which affects the courses taken by students. Taking these effects into account, the distribution of high school grades before and after our treatment are not comparable.

From the university enrolment register and the university credit register, we have information on the time of enrolment in university, the university in which a student has enrolled and the credit values of all courses in which a student has enrolled. The university enrolment register provides, for each course a student has enrolled in, its name and course code, which we supplement with corresponding field of study classifications for each course code. We assign students to one of STEM, humanities/social science/economics, art or other tracks based on the field of study in which a student has taken the most credits. In the case of ties, we assign students to several tracks although this is a rare occurrence. The specific codes which we assigned to each grouping are detailed in appendix table D1. As the treatment group in our data was born in 1997 and the university enrolment register tracks students only until 2018, we are able to identify university enrolment outcomes for our treatment group up to and including the spring of the year in which they turn 21. The university credit register records data up to as far as one semester earlier, that is up to and including the autumn semester of the year the treatment group turns 20. In our sample, 28% of students had enrolled in university by the spring semester of the year in which they turned 21. In comparison 41.7% of the Swedish population had enrolled in university by the age of 25.[5]

As our empirical strategy involves re-centering years around January 1st, we include all students born between July 1991 and June 1994 and between July 1995 and June 1997.[6] Our final sample includes five full birth cohorts, amounting to just under 550,000 individuals for whom we have background characteristics, high school records and university registration records. The final sample consists of 97% of the total number of Swedish compulsory school students. We are unable to assign 6% of the students to university tracks due to missing

[5]This number is based on the fraction of people born in 1991, 1992 and 1993 enrolled in university by the age of 25.

[6]Students born in the year centered around January 1st, 1995 are excluded as they were affected by a reform to the high school system. This is discussed in greater detail in section 4.

course codes, although the proportion of missing data does not vary over time, so this does not pose a threat to the validity of our results. Our sample size remains over 500,000 for these outcomes. Table 1 provides summary statistics for the sample. 6% of students are first-generation immigrants, while a further 9% are second-generation immigrants. The vast majority of students' parents have completed high school, with 36% of mothers and 23% of fathers having achieved a university degree. At their highest level of education, only 7% of mothers studied in a STEM track, but 45% of fathers in our sample have done so. 78% of students graduated high school and 28% had enrolled in university on or before the spring semester of the year in which they turned 21.

4 Empirical Strategy

We apply a fuzzy difference-in-discontinuity design to identify the causal effect of more informative grading on outcomes. In this section, we introduce the constituent parts of our model, building from a simple regression discontinuity design to a sharp difference-in-discontinuity design before reaching our final specification.

Conceptually, we begin with a regression discontinuity design to exploit the date of birth-based assignment of children to school cohorts. This research design would compare the outcomes of students born at the beginning of 1997 to those born at the end of 1996 as these students were exposed to different grading systems at compulsory school. If we were to assume full compliance with school cohort assignment and restrict our sample to students born within the 12-month window around the 1996/97 cohort assignment cut-off of January 1st, 1997, the basic regression discontinuity design can be estimated using the following linear model:

$$Y_i = \alpha + \beta \cdot \mathbb{1}[x_i \geq c] + \gamma_0 \cdot x_i + \gamma_1 \cdot \mathbb{1}[x_i \geq c] \cdot x_i + \epsilon_i \tag{16}$$

where Y_i is the chosen outcome of individual i, x is month of birth, re-centered around the assignment cut-off c and $\mathbb{1}[\cdot]$ is an indicator function. The slope of the relationship between birth month and the outcome is allowed to vary on either side of the assignment cut-off and β represents the effect of being assigned to the first school cohort to receive grades under the new scale.

Table 1: Summary Statistics

	Mean	Std. Dev.	Obs.
Pre-determined characteristics			
Female	0.49	(0.50)	547508
Number of siblings	2.10	(1.50)	547508
Birth order	2.11	(1.23)	547508
Twin	0.03	(0.16)	547508
Other multiple birth	0.03	(0.17)	547508
Has adoptive parent	0.01	(0.10)	547508
Has two adoptive parents	0.01	(0.09)	547508
First generation immigrant	0.06	(0.24)	547508
Second generation immigrant	0.09	(0.28)	547508
Mother has HS degree	0.90	(0.30)	547508
Mother has university degree	0.36	(0.48)	547508
Father has HS degree	0.85	(0.36)	547508
Father has university degree	0.23	(0.42)	547508
Mother studied STEM	0.07	(0.26)	543955
Father studied STEM	0.45	(0.50)	533420
Average Household income	517082.27	(329086.20)	547508
7th-9th grade class size	23.11	(5.17)	506951
High School Outcomes			
Graduate from High School	0.78	(0.41)	545760
Academic track	0.46	(0.50)	545760
STEM trac	0.17	(0.38)	545760
Econ. / Soc. Sci. / Hum. track	0.23	(0.42)	545760
Art track	0.06	(0.23)	545760
University Outcomes			
Enrolled in spring age 21	0.28	(0.45)	541302
Number of credits enrolled in	13.29	(31.13)	541302
Number of credits passed	9.16	(21.37)	541302
Enrolled in a top 5 university	0.13	(0.33)	541302
STEM track	0.08	(0.28)	508015
Econ. / Soc. Sci. / Hum. track	0.05	(0.22)	508015
Art track	0.01	(0.08)	508015
Other track	0.09	(0.28)	508015

Notes: This table presents summary statistics of background characteristics and the main outcome variables of our analysis. Variations in sample size are due to missing values in the data. Household income is calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000. High school outcomes are observed only for those finishing high school. University enrolment outcomes are observed at the end of the spring semester of the year an individual turns 21, while university credit enrolments are observed at the end of the autumn semester of the year an individual turns 20. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

If we were to estimate equation 16, the estimated effect of more informative grading would however be confounded by the fact that children born early in the calendar year benefit from the positive effects of being relatively older than their peers (Black et al., 2011; Fredriksson and Öckert, 2014).[7] This brings us to the sharp difference-in-discontinuity design. To deal with this issue, we extend the model to include a series of comparison cut-offs. The design estimates the discontinuity in outcomes around the 1996/97 assignment cut-off and differences out the discontinuity in outcomes found at our comparison cut-offs. The remaining discontinuity represents the effect of being exposed to the new grading scale, net of any school-starting age effects. This can be identified by estimating the following equation:

$$\begin{aligned} Y_i = {} & \beta_0 \cdot \mathbb{1}[x_i \geq c] + \gamma_0 \cdot x_i + \gamma_1 \cdot \mathbb{1}[x_i \geq c] \cdot x_i \\ & + \mathbb{1}[A_i = 1997] \cdot \{\beta_1 \cdot \mathbb{1}[x_i \geq c] + \gamma_2 \cdot x_i + \gamma_3 \cdot \mathbb{1}[x_i \geq c] \cdot x_i\} + \lambda_{A_i} + \epsilon_i \end{aligned} \tag{17}$$

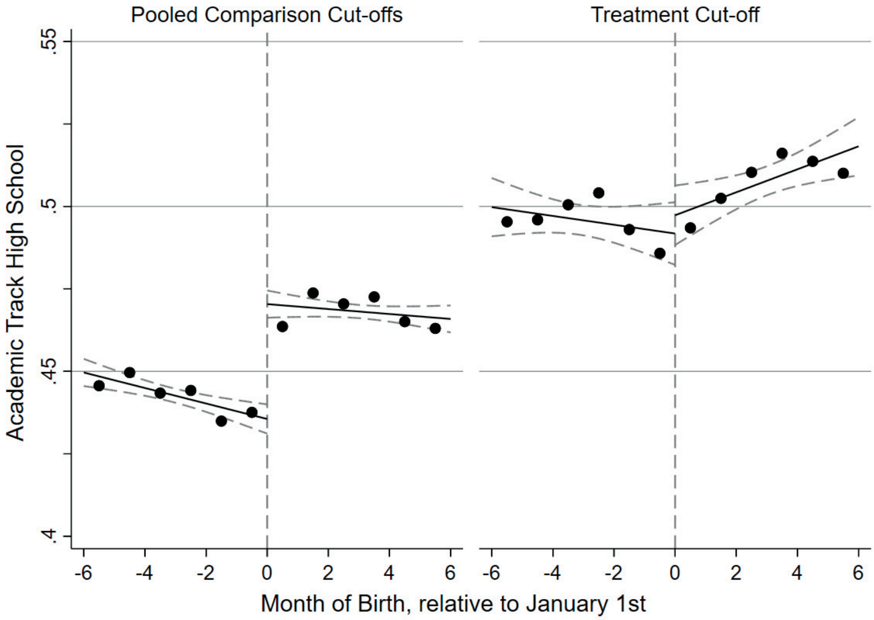
where A represents the 12 month period surrounding a school cohort assignment cut-off in which a student was born, λ_A comprises a vector of assignment cut-off fixed effects, β_0 represents the school starting age effect and β_1 is the coefficient of interest.

We include a number of comparison school cohort assignment cut-offs in order to capture any typical year-to-year variation in the school-starting age effect and to improve statistical power. As a comparison group, we include the 12 month windows around the 1991/92-1993/94 and 1995/96 assignment cut-offs. The 1994/95 cut-off is excluded. This is because, in 2011, a reform was passed that changed several important aspects of the high school system. Among other features of high school, the reform changed the eligibility requirements, introduced the A-F grading scale and augmented the differences between the academic and vocational tracks. Before the reform, all students who graduated from high school were eligible to enrol in higher education. Since the reform, this applied only to students in the academic tracks, but students in the vocational tracks have the opportunity to undertake extra courses to fulfil the qualification requirements and thus progress to higher education. After graduating from high

[7] So-called 'school-starting age effects' refer to the phenomenon that in countries where children are assigned to school cohorts according to their year of birth, children born early in the calendar year typically perform better than those born later in the calendar year. This is generally attributed to two main mechanisms. These are absolute maturity, i.e. that learning in the school environment is more or less effective at certain ages, and relative maturity, i.e. that being relatively older than one's peers gives children an advantage in early school that persists as children get older (Fredriksson and Öckert, 2014).

school, students have the opportunity to take a preparatory year at university in order to become eligible for specific university tracks. The first cohort of students affected by this reform entered high school in the 2011/12 academic year, which corresponds to the 1995 birth cohort. This reform therefore discontinuously affected the outcomes of individuals born just after the 1994/95 cut-off. Including students born in this period could confound the result if the high school reform led to discontinuous changes in factors affecting high school and university outcomes across the 1994/95 assignment cut-off. After excluding this from our comparison cut-offs, for each of the remaining school cohort assignment cut-offs in our sample, individuals born on either side of January 1st were subject to the same high school system and any discontinuities in outcomes are due to school starting age effects. While it is possible that the high school reform may have affected school starting age effects, as a robustness check we show that our results are unaffected if excluding the 1991/92-1993/94 cut-offs from our analysis. We also show that our results are not sensitive to the inclusion or exclusion of any specific birth cohorts.

The difference-in-discontinuity research design is presented graphically in figure 1, using as an example outcome whether a student graduated from an academic high school track. The figure shows the observed share of individuals born in each month in the window surrounding January 1st in each period and a linear fit of the relationship between the outcome and birth month, estimated separately for each side of the January 1st assignment cut-off. The left panel shows the discontinuity in the outcome estimated for the pooled comparison cut-offs. Over this period, students born just after January 1 exhibit a higher likelihood of graduating from an academic high school track than those born just prior to January 1st. This is the school starting age effect. The corresponding discontinuity at the treatment cut-off (i.e. 1996/97) is shown in the right panel. The difference between these two estimated discontinuities represents the treatment effect. As can be seen, there is a large reduction in the discontinuity in the outcome at the treatment cut-off, relative to the comparison cut-offs, indicating that the grading reform had a negative effect on the likelihood of graduating from an academic high school track which is approximately equal in magnitude to the school starting age effect. After January 1st, 1997, the slope of the relationship between graduating from an academic track and month of birth does appear to change although this is likely due to the choice of functional form. Appendix figure A1 shows the same relationship using a quadratic functional form, in which the slopes of the relationship after January 1st are very similar around our pooled comparison cut-offs and our treatment cut-off. We use a linear functional form to ensure a more parsimonious econometric model but show in a robustness check that the results for all of our outcomes are robust to using a quadratic form.



Notes: This figure plots graduation from an academic high school track (as an example outcome) by month of birth (dots), linear fits through 6 month bandwidths on either side of the January 1st school cohort assignment cut-off (solid lines) and robust 95% confidence intervals (dotted lines). The left panel shows the school starting age effect present at the 1991/91, 1992/93, 1993/94 and 1995/96 cut-offs. The right panel shows the 1996/97 cut-off, corresponding to the introduction of more informative grading. The treatment cut-off discontinuity (right) minus the pooled comparison cut-offs discontinuity (left) provides the difference-in-discontinuity estimate of the effect of more informative grading.

Figure 1: Difference-in-Discontinuity Visualisation

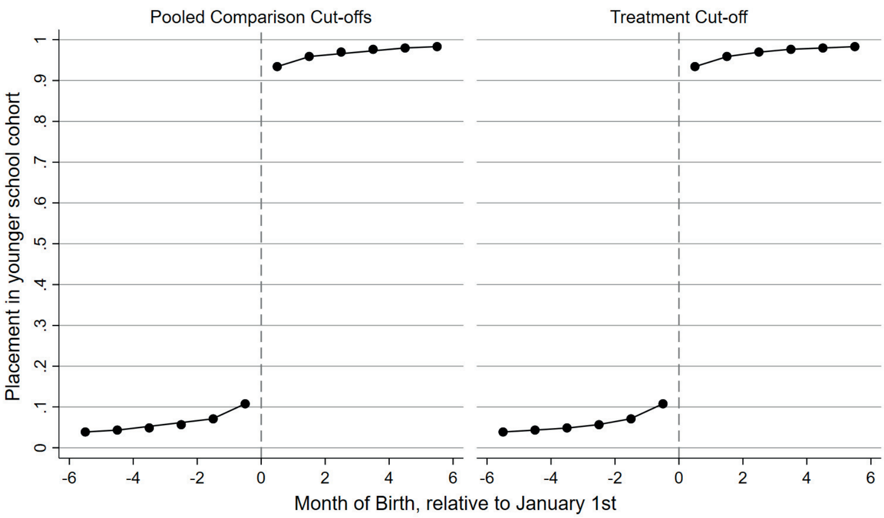
While assignment to school cohorts in Sweden is based on year of birth, this is not a completely binding rule. Failure to account for this would lead to our results being biased toward zero due to attenuation bias in our treatment variables. We therefore exploit a fuzzy difference-in-discontinuity design, instrumenting assignment to the school cohort born after the nearest assignment cut-off with month of birth. Equation 17 can thus be considered the reduced form relationship between educational outcomes and month of birth. Identification requires the estimation of the following system of equations:

$$
\begin{aligned}
S_i = {} & \delta_0 \cdot \mathbb{1}[x_i \geq c] + \rho_0 \cdot x_i + \rho_1 \cdot \mathbb{1}[x_i \geq c] \cdot x_i \\
& + \mathbb{1}[A_i = 1997] \cdot \{\delta_1 \cdot \mathbb{1}[x_i \geq c] + \rho_2 \cdot x_i + \rho_3 \cdot \mathbb{1}[x_i \geq c] \cdot x_i\} + \phi_{A_i} + \nu_i
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
Y_i = {} & \beta_0 \cdot \hat{S}_i + \gamma_0 \cdot x_i + \gamma_1 \cdot \hat{S}_i \cdot x_i \\
& + \mathbb{1}[A_i = 1997] \cdot \{\beta_1 \cdot \hat{S}_i + \gamma_2 \cdot x_i + \gamma_3 \cdot \hat{S}_i \cdot x_i\} + \lambda_{A_i} + \epsilon_i
\end{aligned} \tag{19}
$$

where equations 18 and 19 represent the first and second stages, respectively, and S is an indicator for whether a child is placed in the school cohort in which students born after the January 1st assignment cut-off nearest to their birth month are assigned, or in a later school cohort. Heteroscedasticity robust standard errors are used and to increase precision we include all students born in the 12 month window around each assignment cut-off. Because our running variable is discrete, as a robustness check, we reduce the bandwidth around the cut-offs, as recommended by Kolesár and Rothe (2018). The main specification is estimated using rectangular kernel weights and, as a robustness check, we show that alternative kernel weights do not affect our results.

Our identification strategy requires the assumption that school starting age effects do not vary from year to year in the absence of treatment. As we discuss in section 5.3, to test this assumption we perform a series of placebo tests and show that this is the case. In addition, fuzzy regression discontinuity designs require local randomisation, monotonicity, excludability and a strong first stage (Lee and Lemieux, 2010). We argue that our difference-in-discontinuity design ensures local randomisation. In this setting, monotonicity implies that being born just after the school cohort assignment cut-off of January 1st, 1997 does not cause any individuals to sort into a different cohort. As discussed in section 2.1, it is rare that students change cohort and unlikely that students could change cohort as a consequence of the introduction of more informative grading. Excludability requires that being born just after January 1st affects student outcomes only through exposure to more informative grading and/or the school starting age effect. As discussed in section 2.2, we believe that we are able to identify the effect of more informative grading, as distinct form other aspects of the Lgr 11 reform. We also examine whether our estimates could be affected by other simultaneous shocks. Assuming that these assumptions hold and assuming that the model is correctly specified, $\hat{\beta_1}$ can be interpreted as the LATE of exposure to more informative grading for students born close to the 1996/97 assignment cut-off. In a typical regression discontinuity framework, it is assumed that the treatment variable varies discontinuously across the assignment cut-off while all other variables related to the outcome are continuous across that cut-off. In the difference-in-discontinuity design, we do not require this continuity across the assignment cut-off but rather that any discontinuities across the cut-off are not different at the treatment cut-off, relative to the comparison cut-offs.



Notes: This figure plots on the y-axis S, which represents assignment to the school cohort of those born after the January 1st admissions cut-off nearest to an individual's birth month, against month of birth. Unrestricted means are plotted in bins and a local polynomial is fitted separately on either side of the January 1st cut-off.

Figure 2: Assignment to School Cohort

Supporting the assumption of a strong first stage, it is a very low proportion of students who are placed in a different school cohort than to their assigned cohort. Figure 2 shows the relationship between cohort assignment and month of birth, which shows large compliance with assignment. Although compliance does fall as month of birth gets closer to the assignment cut-off date, sorting non-compliance does not differ around the pooled control and treatment cut-offs.[8] One threat to identification is that students, or their parents, are able to manipulate school cohort assignment in order to receive grades under their preferred system, and thereby violating the assumption of local randomisation. As discussed in section 2.1, the vast majority of changes in school cohort occur before starting school or during the early years of schooling. To be certain that manipulation of school cohort assignment does not occur, we test that month of birth does not predict school cohort assignment or any pre-determined

[8]The difference-in-discontinuity estimate for the likelihood of being placed in the cohort with those born on or after January 1st is -0.0026 with a standard error of 0.0037.

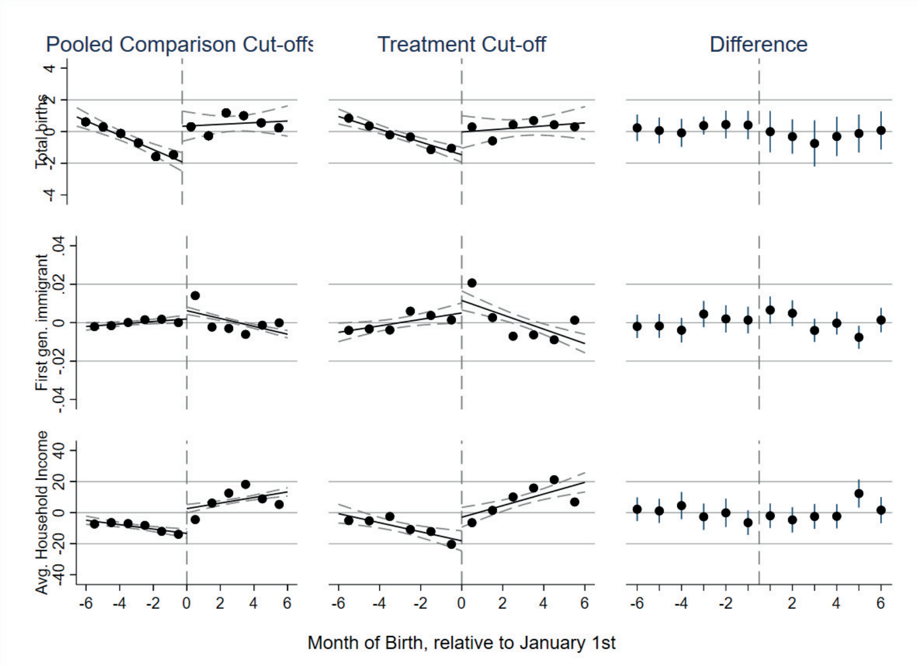
characteristics. Were this to be the case, it may be a signal that certain groups of students or parents may be manipulating school cohort assignment. Sample results of these tests are presented in figure 3. For the examples shown, while in some cases we find evidence of a discontinuous jump across the January 1st cut-off at both the treatment and pooled comparison cut-offs, what is important in our setting is that any discontinuous jump does not change when comparing the comparison cut-offs to the treatment cut-off. There is no noticeable evidence of any statistically significant difference in the discontinuities observed for these pre-determined characteristics. Full results are presented in appendix table A1, in which we find that the model predicts changes to two out of 18 characteristics included. These are the likelihoods of being one of three or more simultaneous births and of having two adoptive parents, both of which occur very rarely and are unlikely to bias our results.

5 Main Results

5.1 Effects of More Informative Grading on Educational Outcomes

Table 2 reports the estimated effects of exposure to more informative grading on high school outcomes. The first row shows the effect of being born just after January 1st, relative to being born just before January 1st, for those in the comparison group. This represents the school-starting age effect. The second row shows the corresponding effect for those in the treatment group. This coefficient therefore represents the sum of the school-starting age effect and the effect of more informative grading. The third row shows the differences between the two coefficients, which reflects the effect of more informative grading only.

Column 1 of table 2 shows how more informative grading affects the likelihood of graduating from high school and column 2 reports how the likelihood of graduating from an academic track is affected. The likelihood of graduating from any high school is reduced by 2.6 percentage points, while the likelihood of graduating from an academic track falls by 3.5 percentage points. Since just over 45% of the students graduate from an academic track, this corresponds to a reduction of almost 8%. Columns 3 through 5 present the effects on the likelihood of graduating from different specialisations within the academic track. The likelihood of graduating from a STEM track falls by 2 percentage points as a result of more informative grading. This corresponds to a 12% decrease in the likelihood of graduating from a STEM track. While the likelihood of graduating



Notes: This figure plots in the left and middle panels three pre-determined characteristics of the sample: total monthly births, first generation immigrant status and average yearly household income (measured in 000s of SEK), each of which have been de-meaned, by month of birth (dots), linear fits through 6 month bandwidths of either side of each year's January 1st cut-off (solid lines) and robust 95% confidence intervals (dotted lines). The right panel shows the estimated difference between the treatment and pooled comparison cut-offs and robust 95% confidence intervals. Average household income is calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000. The difference-in-discontinuity design do not require continuity across the assignment cut-off but rather that any discontinuities across the cut-off are not different at the treatment cut-off, relative to the comparison cut-offs.

Figure 3: Covariate Balance

from the social science, economics or humanities tracks (column 4) appears to be unaffected by the change in the grading scale, the likelihood of graduating from the art track (column 5) is reduced by 1.6 percentage points.

Overall, table 2 shows that the introduction of a more informative grading scale had a negative impact on the share of students who graduated from high school. To put these effects in context, a 2.6 percentage point reduction in the likelihood of graduating from high school is approximately equal to the effect found

Table 2: Effects on High School Outcomes

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
Pooled Comparison Cut-offs	0.011***	0.042***	0.022***	0.026***	–0.007***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)
Treatment Cut-off	–0.016**	0.006	0.002	0.026***	–0.023***
	(0.007)	(0.008)	(0.006)	(0.007)	(0.004)
Difference	–0.026***	–0.035***	–0.020***	0.001	–0.016***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.004)
Pre-reform mean	.793	.456	.168	.231	.0537
N: Pooled Comparison Cut-offs	449,487	449,487	449,487	449,487	449,487
N: Treatment Cut-off	96,273	96,273	96,273	96,273	96,273
N: Total	545,760	545,760	545,760	545,760	545,760

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. High school outcomes are observed only for those finishing high school.

by Elsner and Isphording (2017) of a decrease in one's high school ability ranking of 0.5 within-cohort standard deviations. In the Swedish context, this effect is slightly smaller than that found by Hall (2012), who identified that making upper secondary vocational school more comprehensive led to an increase in the likelihood of dropout by 3.8%. It is also worth pointing out that a change in track composition is not necessarily a bad thing if it leads to improved matching of students to tracks. As we find that changes in track composition are accompanied by lower graduation rates, this appears not to be the case.

Table 3 shows the effect of more informative grading on university enrolment, again including the effect of being born just after January 1st rather than just before January 1st around our pooled comparison cut-offs and around the treatment cut-off, followed by the difference between the two. More informative grading does not seem to have had any effect on whether a student enrolled in university during or before the spring of the year they turn 21 (column 1). Nor does it seem to have affected the likelihood of enrolling in one of the top five Swedish universities (column 2), where the top five are defined as the five universities receiving the most applications during the period of interest.[9] At the intensive margin, which we define as the number of university credits enrolled in and the number of credits passed (columns 3 and 4, respectively), we find that the number of credits passed falls by 0.65 credits, which corresponds to a 7% reduction. In the second row of table 3, field of study is divided into four groups corresponding to the academic high school tracks discussed above. The

[9]These universities are, in alphabetical order: Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University.

likelihood of choosing a STEM track is decreasing as a result of more informative grading and the magnitude is relatively large. Enrolment in a STEM track in university falls by 1.4 percentage points or more than 16%. We do not find any significant changes in enrolment in economics, social science or humanities tracks, in art tracks or in other tracks.

Table 3: Effects on University Outcomes

	(1) Enrolled Spring Age 21	(2) In top 5 University	(3) Number of credits enr. in	(4) Number of credits passed
Pooled Comparison Cut-offs	0.011***	0.002	1.237***	1.029***
	(0.003)	(0.002)	(0.233)	(0.159)
Treatment Cut-off	0.003	0.000	0.597	0.376
	(0.007)	(0.005)	(0.483)	(0.335)
Difference	–0.008	–0.001	–0.640	–0.654*
	(0.008)	(0.006)	(0.536)	(0.370)
Pre-reform mean	.278	.127	13.4	9.22
N: Pooled Comparison Cut-offs	445,537	445,537	445,537	445,537
N: Treatment Cut-off	95,765	95,765	95,765	95,765
N: Total	541,302	541,302	541,302	541,302
	(5) Stem track	(6) Econ./Soc.Sci./ Hum. Track	(7) Art Track	(8) Other track
Pooled Comparison Cut-offs	0.008***	0.006***	–0.001	0.005**
	(0.002)	(0.002)	(0.001)	(0.002)
Treatment Cut-off	–0.005	0.007*	–0.001	0.006
	(0.005)	(0.004)	(0.001)	(0.005)
Difference	–0.014***	0.001	–0.000	0.000
	(0.005)	(0.004)	(0.001)	(0.005)
Pre-reform mean	.0838	.0527	.0059	.0875
N: Pooled Comparison Cut-offs	417,046	417,046	417,046	417,046
N: Treatment Cut-off	90,969	90,969	90,969	90,969
N: Total	508,015	508,015	508,015	508,015

Notes: Robust standard errors in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

Given that we find no effects on university enrolment, it is possible that the negative effect on high school graduation is driven by students who would not have enrolled in university anyway. However, as we identify a negative effect on points passed, there does appear to be negative effects for those still enrolling in university. We can also see that the reduction in the likelihood of graduating from a STEM track in high school is persistent, with lower STEM enrolment at university. As a robustness check, we re-estimate our models reflecting track choice using alternative definitions of field of study.[10] These results are presented in appendix table A2 and are almost identical to the main results, supporting the robustness of our findings.

It is well established that there exist gender differences in participation in different educational tracks. Prior to the introduction of more informative grading, girls were much more likely to graduate from an academic high school track and from an economics, social science or humanities track, while boys were more likely to graduate from STEM tracks in high school and to enrol in STEM tracks at university. As our results identify effects of more informative grades on track choice, it is possible that there may exist gender differences in these effects.

Table 4 reports estimates corresponding to those in the third row of tables 2 and 3, now interacting the model with an indicator for being female. Note that from this point forward, results table include only the effect of more informative grading, net of the school-starting age effect. The first row presents the effect on boys, the second row presents the difference in outcomes for girls, relative to boys and the third row presents the effects of more informative grading on girls, relative to boys. If there were significant gender differences, this would be reflected in the third row, but as can be seen in the first panel, none of these differences are statistically significant for high school outcomes.

Looking at university enrolment (columns 6 through 9 of table 4), boys appear to experience more negative consequences of receiving more informative grades than girls, but none of these differences are statistically significant. Looking at track choices in university (columns 10 through 13), no gender differences can be detected in the likelihood of enrolling in economics, social science and humanities tracks, nor the group comprising other tracks. Girls become slightly less likely to enrol in an art track relative to boys. The most striking difference is found in the likelihood of enrolling in a STEM track. The negative effect on enrolling in a STEM track seems to be mainly driven by boys. As a consequence of more informative grading, the likelihood of enrolling in a STEM track in university is decreasing by 2.2 percentage points for boys while the corresponding decrease for girls is 0.5 percentage points. This is, however, only significant at the ten percent level.

[10]These alternative definitions correspond to those used by the Swedish government to assign university funding. Details of the definitions can be found in appendix table D2.

Table 4: Gender Differences

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
Panel A: High School Outcomes					
Treatment	–0.025**	–0.038***	–0.028***	0.001	–0.011**
	(0.011)	(0.012)	(0.010)	(0.010)	(0.005)
Female	0.069***	0.141***	–0.052***	0.137***	0.054***
	(0.005)	(0.006)	(0.004)	(0.005)	(0.003)
Treatment*Female	–0.003	0.005	0.016	–0.003	–0.010
	(0.015)	(0.017)	(0.013)	(0.015)	(0.008)
Pre-reform: boys	.766	.401	.194	.172	.0332
Pre-reform: girls	.82	.513	.14	.293	.075
N	545,760	545,760	545,760	545,760	545,760

	(6) Enrolled Spring Age 21	(7) In top 5 University	(8) Number of credits enr. in	(9) Number of credits passed
Panel B: University Outcomes				
Treatment	–0.019*	–0.010	–0.804	–0.796
	(0.010)	(0.007)	(0.694)	(0.485)
Female	0.082***	0.054***	4.217***	2.967***
	(0.005)	(0.004)	(0.360)	(0.245)
Treatment*Female	0.020	0.017	0.270	0.244
	(0.015)	(0.011)	(1.070)	(0.739)
Pre-reform: boys	.239	.104	11.3	7.81
Pre-reform: girls	.319	.151	15.5	10.7
N	541,302	541,302	541,302	541,302

	(10) Stem track	(11) Econ./Soc.Sci./ Hum. Track	(12) Art Track	(13) Other track
Treatment	–0.022***	0.006	0.002	–0.000
	(0.008)	(0.005)	(0.002)	(0.005)
Female	–0.057***	0.025***	–0.000	0.090***
	(0.003)	(0.003)	(0.001)	(0.003)
Treatment*Female	0.017*	–0.011	–0.005*	0.001
	(0.010)	(0.008)	(0.003)	(0.010)
Pre-reform: boys	.114	.0386	.00583	.0422
Pre-reform: girls	.0517	.0679	.00598	.136
N	508,015	508,015	508,015	508,015

Notes: Robust standard errors in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. High school outcomes are observed only for those finishing high school. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

That we do not detect any gender differences in high school outcomes and only to a small degree in university outcomes is in line with the findings of Azmat and Iriberri (2010) and Azmat et al. (2019), who, in looking at the effects of relative grading in university and high school, respectively, do not find any gender differences in outcomes. We also look at heterogeneous effects based on immigrant status. As can be seen in appendix table A3 we do not find evidence of economically or statistically significant variation in treatment effects across these groups.

A concern with regard to our identification strategy is the external validity of our results. As in any regression discontinuity framework, identification relies on estimating the treatment effect on those individuals with values of the running variable around some threshold value. In our case, identification relies on those born early in the school year. It is therefore a concern that the effect of treatment may differ for those born early in the year relative to those born later in the year, given that it is well-established that those born early tend to have better educational and labour market outcomes. However, the relationship between family background and outcomes is stronger in magnitude than that of month of birth. For example, appendix figure A1 compares the unconditional relationship between a subset of our outcomes and month of birth with that of household income, divided into 12 quantiles. As can be seen, the income gradient is much steeper than the month of birth gradient for the likelihood of graduating from high school, graduating from a STEM high school track or to enrol in a STEM track at university. If students who are expected to perform better are affected by more informative grades differently to the rest of the population, we should expect to find heterogeneous treatment effects among groups of different backgrounds. We do not find evidence of heterogeneous treatment effects by gender or immigrant status nor, as we discuss in section 6.3, by parental income and parental education. This leads us to believe that the effects we estimate are not driven by those born earlier in the year being affected differently by more informative grading.

5.2 Estimating Long-Term Effects on Income

Ideally, we would like to examine how more informative grading affects students' long-term labour market outcomes. As the treatment group was born in 1997 and our income data is observed only as far as 2016, the year in which the treated cohort turned 19, we are unable to directly examine earnings later in life. In the absence of data, we instead identify the effect of more informative grading on future earnings using a surrogate index, as developed by Athey et al.

(2019). The surrogate index represents the predicted value of future earnings, estimated as a function of the intermediate outcomes examined in tables 2 and 3. More precisely, we estimate the effect of more informative grading on the log of average yearly earned income between the ages of 28 and 30, adjusted according to Statistics Sweden's Consumer Price Index to the year 2000. We use earnings from ages 28-30 as an outcome because, as the average age of graduation from university in Sweden is 28 (OECD, 2020), this window best reflects the starting salary of university graduates.

The surrogate index is constructed using older cohorts for whom we have access to data on the intermediate outcomes and earned income in adulthood. As surrogates, we use the population of Sweden born in 1985 and 1986. We regress average earned income at ages 28-30 on the high school and university outcomes that are presented in tables 2 and 3 and use the results of this regression to fit predicted earnings in our analysis sample. These predicted values serve as the surrogate index and are used to calculate the effect of more informative grading on future earnings.

While it is common to use a short-term proxy to estimate treatment effects on longer-term outcomes, it requires the strong assumption that the short-term proxy can fully predict the relationship between treatment and the longer-term outcome. Athey et al. (2019) show that by combining several short-term proxies into a surrogate index, weaker assumptions are required. Under the assumptions of unconfoundedness, surrogacy and comparability, the average treatment effect on the long run outcome is equal to the average treatment effect on the surrogate index. Unconfoundedness is the requirement that treatment is randomly assigned. We argue that our difference-in-discontinuity approach ensures this.

Surrogacy requires that the conditional distribution of future earnings are independent of the treatment given the surrogate index. That is that the intermediate outcomes we use together span all of the causal path of the effect of more informative grading on earnings at ages 28-30. This means that some intermediate outcomes may be excluded without threatening surrogacy if that excluded variable is fully captured, i.e. it acts as a mechanism for or is determined by a combinations of some or all of the intermediate variables we use. We argue that this is a reasonable assumption to make in this setting since any causal effect of more informative grading should be reflected in treated students' future educational choices and performance.

Comparability requires that the relationship between the intermediate outcomes and future earnings is the same in the main sample and the sample used to construct the surrogate index. In our setting, this implies that the relationship between the high school and university outcomes used and average earnings at age 28-30 is comparable between those born between 1985 and 1986 and those born between 1991 and 1997. While we cannot fully test this assumption given that we cannot observe incomes for those in our main sample, we argue that it is a reasonable assumption to make by comparing the relationship between our intermediate outcomes and earnings for those born between 1985 and 1986 to those born between 1980 and 1981. As shown in appendix figure A3, the coefficients from a regression of log earnings on our intermediate outcomes are remarkably similar using both samples. As comparability holds from 1980 and 1981 to 1985 and 1986, we argue that it is also likely to hold from 1985 and 1986 to 1991 and onward. While identification relies on these three assumptions, if any of them were to not hold in our setting, the results of this analysis may still be seen as suggestive evidence of the effect of more informative grading on earnings.

Table 5 shows that the introduction of more informative grading causes a significant decrease in yearly income of approximately 1.8%. To put the size of this effect into context, we compare our finding to that of Fredriksson et al. (2013), who exploit a maximal class size rule in Sweden in order to estimate the effect of smaller classes on later earnings. They find that a class size reduction of seven students (equivalent to the class size reduction in project STAR) increased adult wages by 4.4%. The introduction of a more informative grading system thus causes a negative effect on earnings of a magnitude a little under half of the class size effect.

Table 5: Long-Term Effect on Income

	(1) Log Avg. Income Age 28-30
Treatment	–0.018**
	(0.007)
Pre-reform mean	11.5
N	508,015

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Income is estimated using a surrogate index, using as surrogates all individuals in Sweden born in 1985 and 1986. Income is measured as average total yearly income and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000.

5.3 Robustness of the Main Results

Our identification strategy relies on identifying discontinuous variation in outcomes across the 1996/97 school cohort assignment cut-off of January 1st, 1997, relative to our comparison school cohort assignment cut-offs. A primary concern with our identification strategy is thus that our results may be confounded by idiosyncratic school cohort-specific shocks affecting outcomes. If cohort-specific shocks are indeed a common occurrence then we would also expect to find significant changes in outcomes across individual cut-offs in our comparison group, relative to the remaining comparison cut-offs. We therefore conduct a series of placebo tests within our comparison group, using each assignment cut-off in turn as a placebo treatment cut-off. In particular, if the introduction of new syllabi discussed in section 2.2 had any impact on student outcomes, we would expect significant effects when using the 1995/96 assignment cut-off as a placebo. The results of these tests for our high school and university outcomes are presented in tables 6 and 7, respectively. Each of the placebo estimates are all close to zero and only 3 out of 52 estimates are significant at the 10% level, one of which is significant at the 5% level, less than what could be expected to occur by chance.

Table 6: Placebo Tests - High School Outcomes

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
1992	0.005	–0.012	–0.002	–0.009	0.001
	(0.007)	(0.008)	(0.006)	(0.007)	(0.004)
1993	–0.001	0.000	–0.002	–0.002	0.000
	(0.007)	(0.008)	(0.006)	(0.007)	(0.004)
1994	–0.016**	0.003	–0.005	0.007	0.000
	(0.007)	(0.008)	(0.006)	(0.007)	(0.004)
1996	0.013*	0.010	0.010	0.004	–0.002
	(0.008)	(0.009)	(0.007)	(0.008)	(0.004)
N	449,487	449,487	449,487	449,487	449,487

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. The left-most column details the cut-off year used as a placebo treatment cut-off. Estimates are taken from separate difference-in-discontinuity regressions for each outcome and placebo cut-off. High school outcomes are observed only for those finishing high school.

Idiosyncratic cohort-specific shocks may also affect the discontinuity in outcomes during the control period in a manner which leads to an apparent significant effect in the treatment period. We show in appendix tables A4 and A5 that our results are not driven by such shocks by excluding each comparison cut-off in turn and re-estimating our models. In addition, it is possible that as students born before 1995 entered a different high school system to those born after 1995, the inclusion of assignment cut-offs before 1995 in our comparison group may affect the results. Appendix table A6 shows that our results are essentially unchanged when using only the 1995/96 assignment cut-off as a comparison. Furthermore, we use a randomisation inference design to include all possible placebo assignment cut-offs in our period of interest. We use as placebo cut-offs

Table 7: Placebo Tests - University Outcomes

	(1) Enrolled Spring Age 21	(2) In top 5 University	(3) Number of credits enr. in	(4) Number of credits passed
1992	–0.003	–0.002	–0.401	–0.024
	(0.007)	(0.006)	(0.528)	(0.361)
1993	–0.007	–0.006	–0.399	–0.550
	(0.008)	(0.006)	(0.532)	(0.363)
1994	0.000	0.009	0.569	0.343
	(0.008)	(0.006)	(0.544)	(0.368)
1996	0.011	–0.001	0.287	0.262
	(0.008)	(0.006)	(0.551)	(0.378)
N	445,537	445,537	445,537	445,537
	(5) Stem track	(6) Econ./Soc.Sci./ Hum. Track	(7) Art Track	(8) Other track
1992	0.001	0.003	0.001	–0.007
	(0.005)	(0.004)	(0.001)	(0.005)
1993	–0.003	–0.002	–0.001	–0.002
	(0.005)	(0.004)	(0.001)	(0.005)
1994	–0.007	0.002	0.000	0.006
	(0.005)	(0.004)	(0.001)	(0.005)
1996	0.009*	–0.004	–0.001	0.002
	(0.005)	(0.004)	(0.001)	(0.005)
N	417,046	417,046	417,046	417,046

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. The left-most column details the cut-off year used as a placebo treatment cut-off. Estimates are taken from separate difference-in-discontinuity regressions for each outcome and placebo cut-off. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Umeå University, Lund University, Stockholm University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

all birth months included in our analysis sample, excluding the three months book-ending each period from which we draw our control group in order to allow enough data on each side of the cut-off for identification. The periods used for randomisation inference are thus October 1991-March 1994 and October 1995-March 1997. Table 8 presents the p-values estimated in tables 2, 3 and 5 alongside the corresponding p-values estimated using randomisation inference. For university credits enrolled in and university credits passed, the results in fact become statistically significant at the 1% level using randomisation inference. As a result of these robustness checks, we believe that it is very unlikely that our results could be driven by any idiosyncratic shocks coinciding with the treatment cut-off.

Table 8: Randomisation Inference

	Original p-value	R.I. p-value
Graduate High School	0.00	0.02
Academic Track High School	0.00	0.02
STEM Track High School	0.00	0.04
Econ./ Soc. Sci./ Hum. Track High School	0.93	0.47
Art Track High School	0.00	0.05
Enrolled in uni spring age 21	0.28	0.12
Enrolled in a top 5 university	0.83	0.44
University credits enrolled in	0.23	0.01
University credits passed	0.08	0.00
STEM track university	0.01	0.13
Econ./ Soc. Sci./ Hum. Track university	0.81	0.44
Art Track	0.94	0.49
Other Track	0.93	0.47
Log Avg. Income aged 28-30 (Surrogate)	0.02	0.08

Notes: The original p-values presented in this table are those of the corresponding estimates in tables 2, 3 and 6. The second column of p-values presented were calculated using randomisation inference over all months included in the analysis sample, excluding the three months book-ending each period in order to allow for identification. The periods used for randomisation inference are therefore October 1991 - March 1994 and October 1995 - March 1997. High school outcomes are observed only for those finishing high school. University enrolment outcomes are observed at the end of the spring semester of the year an individual turns 21. University credit enrolment outcomes are observed at the end of the autumn semester of the year an individual turns 20. Income is estimated using a surrogate index, using as surrogates all individuals in Sweden born in 1985 and 1986. Income is measured as average total yearly income and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000.

The next set of robustness checks are presented in appendix tables A7 and A8, respectively. Our data only allows the use of a discrete running variable, namely month of birth. In such settings, Kolesár and Rothe (2018) recommend, in cases with large enough sample size, to reduce the bandwidth to a small enough level to reduce bias in the estimates and employ conventional heteroscedasticity-robust standard errors. In our main analysis we use a 12 month period around each January 1st assignment cut-off in order to reduce variance. As a robustness check, we reduce our bandwidth to 6 months. As can be seen in appendix tables A7 and A8 our results are robust to this reduction in bandwidth. It can be argued whether a linear model is the best fit for the data. As an example, appendix figure A2 shows graphically how a quadratic fit would look, using the likelihood of completing an academic high school track as an outcome. This appears to fit the data quite well. In appendix tables A7 and A8, we re-estimate our models with quadratic rather than linear functional forms and show that none of our estimates are driven by mis-specification bias. Gelman and Imbens (2019) show that higher order polynomials can cause bias in regression-discontinuity designs due to overfitting, a concern which is especially acute in this case as our running variables takes only six discrete values on either side of the assignment threshold. As such, we employ only linear and quadratic functional forms.

As discussed in section 4, we do not find evidence of any differences in the variation in background characteristics between our pooled comparison cut-offs and our treatment cut-off that might impact our results. To be certain that the results are not driven by changes to the composition of students on either side of the thresholds, we re-estimate our models including a series of individual characteristics as control variables. As can be seen in appendix tables A7 and A8, the introduction of these controls do not change our results. While it is possible to pass through compulsory school in Sweden in a different cohort to that assigned, these changes generally occur in the first few years of schooling and thus before anyone could have been aware of a change to the grading scale. It is still theoretically possible, however, that a student may have been able to change cohort to receive grades on their preferred grading scale. To ensure that this is not affecting our results, we also run a reduced form model, where we find smaller effect sizes but similar levels of significance. Finally, it is commonplace in the regression discontinuity literature to weight observations in a manner that gives more importance to observations closest to the cut-off. While our main specification weights all observations equally, we show that our results are robust to implementing a triangular kernel. As can be seen in the last two rows of appendix tables A7 and A8, our results are robust to each of these checks.

Under this battery of robustness tests, we lose some significance in cases where we drop observations, which is to be expected. However, given that the majority of our estimates remain similar in magnitude and significance to our original estimates, we conclude that our estimates are largely robust and unlikely to have been caused by chance or any other simultaneous shock to outcomes.

6 Mechanisms

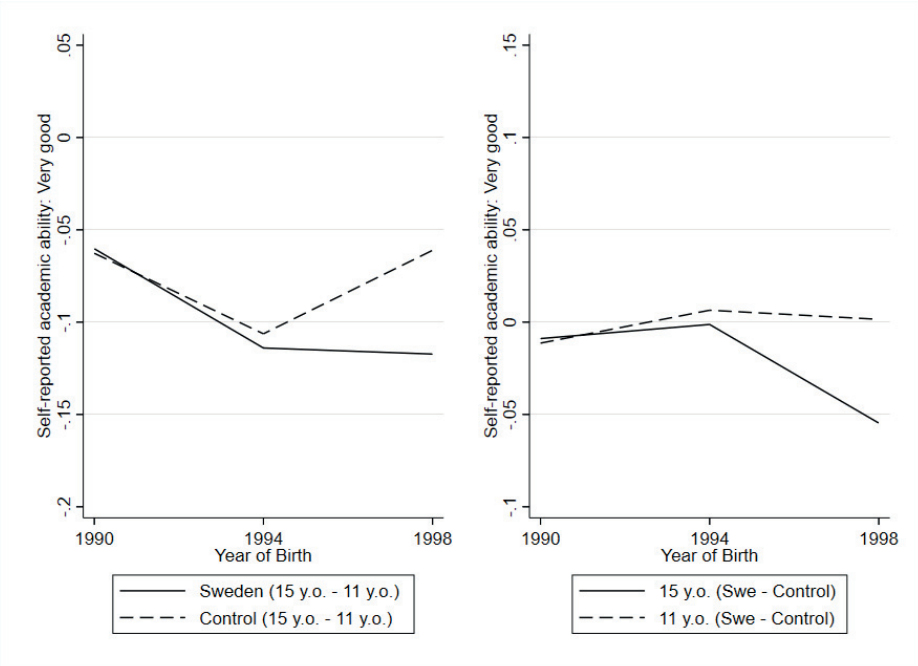
Our main results show long-term negative effects of more informative grading. In this section, we explore mechanisms related to academic achievement. These are students' perceived ability and attitudes to school, followed by mental health. We then discuss a number of other mechanisms which we believe are not driving our results.

6.1 Perceived Ability and Attitudes to School

To examine how the introduction of more informative grading affects students' perceived ability and attitudes to school, we exploit data from the Health Behaviour in School-Aged Children survey (HBSC). The HBSC is an international study organised by the World Health Organisation, comprising a self-reported questionnaire-based survey of schoolchildren every four years with nationally representative samples of children aged 11, 13 and 15. The study has been conducted since 1982 and includes over 50 countries and regions. We use data from surveys conducted in 2002, 2006, 2010 and 2014. The Swedish 1998 birth cohort were treated by the grading scale reform and this cohort was sampled at age 11 in the 2010 wave (before the grading reform) and again at age 15 in the 2014 wave (after the grading reform) of the HBSC. We exploit a pseudo-panel strategy to identify the effects of the reform. In addition to the 1998 birth cohort, the 1990 and 1994 birth cohorts, who finished compulsory school before the grading reform (and thus were never treated), were also sampled at ages 11 and 15.

To examine perceived ability and attitudes to school, we estimate the effects of more informative grading on the answers to a question regarding students' opinions of their own academic ability, a question asking students how much they like school and a question on how pressured students feel by their schoolwork. Appendix C gives an overview of the exact questions used in the HBSC. We exploit a triple-differences estimation strategy, examining changes in responses to HBSC questions within the 1998 birth cohort between ages 11 and 15, relative to the same changes in previous cohorts and comparing this change within Sweden to the corresponding changes in Sweden's neighbouring Scandinavian countries of Denmark and Norway. Comparing outcomes between 11 and 15 year olds from the same birth cohort removes any cohort specific variation and comparing outcomes across birth cohorts removes any age-specific variation, while comparing across countries removes any time-specific variation. The key assumption required to identify a causal effect of more informative grading is not necessarily that the Danish and Norwegian compulsory school systems represent a suitable counterfactual for the Swedish compulsory school system in the absence of any reforms, but rather that the difference in outcomes between 15 and 11 year olds in both groups (or the difference between groups at both ages) would follow the same path in the absence of treatment.

A visual interpretation of our estimation strategy is provided in figure 4, showing from two perspectives the progression of differences in the likelihood of a child reporting their school performance to be very good. In the left panel the solid line shows the differences in that likelihood between Swedish 15 and 11 year olds. The dashed line shows the same for Denmark and Norway. The slope of the two lines are on a similar trend before diverging at the 1998 cohort, with a relative decrease for Swedish 15 year olds, which comprise the treatment group. The right panel shows a similar pattern from an age perspective. The solid line shows the responses of Swedish 15 year olds relative to Danish and Norwegian 15 year olds, while the dashed line shows the same for 11 year olds. Again, we see a similar trend between the 1990 and 1994 cohorts followed by a divergence at the 1998 cohort.



Notes: This figure visualises the triple differences model using the other Scandinavian countries of Denmark and Norway as a control group. This graph uses as an outcome the proportion of respondents to the Health Behaviour of School Children survey who answered the question “In your opinion, what does your class teacher(s) think about your school performance compared to your classmates?” with “very good”. The left panel shows the development of the differences in the outcome between 15 and 11 year old children in Sweden and the control countries. The right panel shows the development of the differences in the outcome between Sweden and the control countries for 15 and 11 year old children. Both panels show a relative decrease in the proportion of Swedish 15 year old children believing their performance to be very good.

Figure 4: Triple Differences Visualisation

Table 9 provides the estimated effects of receiving more informative grades on the likelihood of choosing specific answers to the above questions. Appendix table A9 provides a more detailed set of results, including both pre-reform birth cohorts in the regression table in an event study setting in order to examine pre-trends. There is a large reduction in the likelihood of students believing that they are good or very good in school. By comparing the estimate (row 1, column 2) with the pre-reform mean for Swedish 15-year old's we can see that the effect size corresponds to a drop in the likelihood of reporting that they are good or very good in school by 32%. Reductions in the likelihood of reporting to be good or very good in school may point toward negative effects only for better students, which is slightly at odds with our findings that more informative grading mainly acts at the margin of high school graduation. We do find effects on some university outcomes, however, which would affect mainly higher ability students. More importantly, the proportion of students who, prior to the introduction of more informative grading, believed themselves to be good or very good in school comprised the majority of students at 58.1% of the sample of 15 year olds born in 1990 and 1994.

With less informative grades, students appear to be more confident in their academic ability, which receives a negative shock due to the introduction of more informative grades. This could simply reflect students learning that they are in fact performing worse than they previously would have thought under less informative grading. This is consistent with the findings of the experimental literature examining mechanisms through which feedback affects performance. Feedback about positive performance increases self-confidence (Möbius et al., 2014), which in turn leads to greater investment in learning and improved performance (Fischer and Sliwka, 2018). Our results are, however, inconsistent with Zimmermann (2020), who finds that the effects of negative feedback dissipate over time. Alternatively, this could be due to just missing out on reaching a specific grade threshold which previously would not have existed, leading to discouragement. This discouragement from failing to reach a grading threshold has been shown by Campos-Mercade and Wengström (2015) to have significant longer term effects on achievement, albeit at university level, where effects were only found for girls, while we find effects for both boys and girls in our sample.[11]

We find no evidence of any impact on liking school. We do, however, find evidence of an increase in the likelihood of students reporting some or more pressure from schoolwork. It is worth noting that, as shown in appendix table A9, there is evidence of differing pre-trends in the likelihood of reporting a

[11] In regressions not reported by the authors, we found the effects in table 9 to be similar for boys and girls.

Table 9: Attitudes to School

	(1)	(2)	(3)
	Very good	Good/ Very good	Average/ Good/ Very good
Academic Ability	–0.085***	–0.190***	–0.020
	(0.025)	(0.030)	(0.012)
Pre-reform mean	.145	.581	.934
N	25,241	25,241	25,241
	A lot	A little/ A lot	Not much/ A little/ A lot
Like School	0.034	0.006	–0.014
	(0.033)	(0.027)	(0.014)
Pre-reform mean	.125	.682	.908
N	25,241	25,241	25,241
	A lot	A little/ A lot	Some/ A little/ A lot
School Pressure	0.020	0.116***	0.118***
	(0.023)	(0.032)	(0.028)
Pre-reform mean	.2	.482	.924
N	25,241	25,241	25,241

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. The pre-reform mean refers to the mean of Swedish 15-year old's born in 1990 and 1994. Estimates are taken from separate triple-differences regressions for each outcome using the other Scandinavian countries of Denmark and Norway as a control group. Identification exploits differences in the outcomes of 15 year old and 11 year old children from the 1998 birth cohort in Sweden relative to the same changes in earlier birth cohorts in Sweden. This difference is further compared to the same differences in the control countries. All models include controls for gender and whether a child considers their family to be well-off. While most of models so not show statistically significant evidence of parallel trends, some do. This is relevant to the third row of column (2). The models presented in this table are re-produced including estimates of the treatment effect in the control periods in appendix table A9.

little or a lot of pressure at school so this result should be taken with a grain of salt. Nevertheless, these results provide evidence that the introduction of more informative grades at compulsory school had negative impacts on students' perceived ability and attitudes toward school.

6.2 Mental Health

We examine mental health impacts using both the HBSC survey data and data from administrative registers. Using the HBSC data, we are able to identify the effects of more informative grading on how often children experience mental health issues using the triple differences identification strategy described in section 6.1. We use as outcomes the answers to questions regarding how often children experience feelings of irritability or bad temper, nervousness and difficulties in getting to sleep. 7 gives an overview of the exact questions used to measure mental health in the HBSC survey.

Table 10 provides the main results.[12] While we do not find evidence of any strong patterns with regard to irritability or nervousness, we identify a large and strongly significant effect on the likelihood of reporting difficulties in getting to sleep. This is identified at all levels of instance, from monthly or more often up to experiencing sleep difficulties on a daily basis. As difficulties in sleeping are often associated with feelings of anxiety and stress, and taken together with the suggestive evidence on increased school pressure found in table 9, this suggests that more informative grading increases students stress levels. One would however expect that increased stress levels among students would also increase feelings of nervousness and irritability, outcomes for which we do not detect any significant changes. We speculate that it may be easier for individuals to estimate the number of times they have experienced difficulties in getting to sleep compared to feelings of nervousness and irritability and that this self-reported outcome is therefore less noisy and more reliable as a measure of stress levels. Indeed, both feeling irritable and nervous predict the likelihood of reporting difficulty sleeping with a high level of significance for the children in our sample, as shown in appendix table A11.

Using Swedish register data covering the entire population, we examine mental health diagnoses in specialised inpatient and/or outpatient care. We cannot observe mental health diagnoses in primary care since no such national register exists in Sweden. We observe diagnoses up to the end of 2012, which means up to midway through 9th grade for our treatment group. We therefore observe mental health outcomes at this point for all individuals in our sample. Using the fuzzy difference-in-discontinuity identification strategy described in section 4, we examine at the extensive margin the likelihood of receiving any mental health diagnosis and at the intensive margin the number of mental health diagnoses. The results of this analysis are presented in table 11. The first row uses as

[12]Appendix table A10 details the same but including pre-trends in the outcomes in an event-study setting.

Table 10: Self-reported Mental Health

	(1) Monthly / more often	(2) Weekly / more often	(3) More than weekly / more often	(4) Daily / more often
Irritable or Bad Temper	0.010	0.033	0.046*	0.013
	(0.028)	(0.034)	(0.026)	(0.014)
Pre-reform mean	.88	.606	.321	.0729
N	25,241	25,241	25,241	25,241
Nervous	–0.025	–0.003	0.019	0.000
	(0.033)	(0.031)	(0.023)	(0.011)
Pre-reform mean	.734	.41	.173	.0389
N	25,241	25,241	25,241	25,241
Difficulty Sleeping	0.094***	0.112***	0.105***	0.057***
	(0.036)	(0.032)	(0.028)	(0.019)
Pre-reform mean	.642	.42	.245	.0931
N	25,241	25,241	25,241	25,241

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. The pre-reform mean refers to the mean of Swedish 15-year old's born in 1990 and 1994. Estimates are taken from separate triple-differences regressions for each outcome using the other Scandinavian countries of Denmark and Norway as a control group. Identification exploits differences in the outcomes of 15 year old and 11 year old children from the 1998 birth cohort in Sweden relative to the same changes in earlier birth cohorts in Sweden. This difference is further compared to the same differences in the control countries. All models include controls for gender and whether a child considers their family to be well-off. The models presented in this table are re-produced including estimates of the treatment effect in the control periods in appendix table A10.

an outcome any mental health diagnoses, while the following rows divide these diagnoses into different sub-groups, namely self-harm, drug use, anxiety and education-related mental health issues. Details of the diagnoses categorised in each group are detailed in appendix table D3.

Mental health related diagnoses in either specialised inpatient or outpatient care are very rare in Sweden. As only 3.1% of our sample have any mental health diagnosis at the time of observation, with a sample mean of 0.09 diagnoses per person, these can be considered to be more extreme outcomes. While table 11 shows weak evidence of an increase in drug use and education-related mental health diagnoses, these effects are very small and only marginally significant and thus we conclude that more informative grading does not have an effect on

Table 11: Mental Health Diagnoses

	(1) Any diagnosis	(2) No. of diagnoses
Health	0.002	0.016
	(0.003)	(0.013)
	[.0311]	[.0823]
Self-harm	–0.000	–0.000
	(0.001)	(0.001)
	[.00254]	[.00334]
Drugs	0.001	0.003*
	(0.001)	(0.002)
	[.00512]	[.00638]
Anxiety	0.001	0.006
	(0.002)	(0.010)
	[.0135]	[.0428]
Education	0.002	0.015*
	(0.002)	(0.008)
	[.015]	[.0363]
N	547,508	547,508

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Pre-reform means in square brackets. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. As the mental health data we have received access to information only up to the December 2012, which is the year in which our treatment group is in 9th grade, outcomes here are measured up to that point for all individuals. Column (1) uses as an outcome an indicator for whether or not an individual received any mental health diagnosis. Column (2) uses as an outcome the number of mental health diagnoses received.

more severe mental health outcomes.

More informative grades lead to dis-improvements in mental health, which is associated with worse education outcomes at high school level and worse labour market outcomes (Cornaglia et al., 2015), along with worse education outcomes at the university level (Eisenberg et al., 2009). On the other hand, it is possible that more informative grading leads to worse education outcomes for reasons such as those discussed in section 6.1, which is in turn associated with poorer mental health (Chevalier and Feinstein, 2006). As such, the mental health effects we have identified may be both a cause and/or a symptom of the effects we find on education outcomes. The results in this section can therefore be considered more as suggestive evidence of mental health as a mechanism.

6.3 Other Mechanisms

We now turn to the possibility that the effects we find could be driven by factors other than perceived ability and mental health. Once such potential mechanism is the matching of students to different educational tracks and labour market options. For example, Bobba and Frisancho (2016) find that receiving feedback on mock high school admissions tests affects students' preference ordering over high school tracks. However, as discussed in section 5.1, we find that changes in the composition of high school track completions are accompanied by a lower likelihood of graduating from high school. Similarly, in section 5.2 we find that more informative grading led to lower labour market earnings. In both cases, the opposite effects would be expected if more informative grading led to improved matching.

Another potential mechanism could be that parents can now engage in more appropriate educational investment in their children's human capital as a result of more informative grading. For example, with more informative grading, parents may be more able to identify where their children are falling behind and tailor their investments to those areas (Cobb-Clark et al., 2021; Dizon-Ross, 2019). If this were to be the case, we may expect the children of wealthier parents, who have more resources to invest in their children, or more educated parents, who may be better able to identify appropriate educational investments, to be affected differently by the introduction of more informative grading. To test this hypothesis, we test for heterogeneity in the treatment effect across above and below median levels of household wealth and across higher and lower parental education, defined by whether at least one of a child's parents holds a university degree. Appendix tables A12 and A13 show that this is not the case.

Finally, changes in teacher behaviour after the reform may play a role in affecting student outcomes. If teachers struggle to adapt to the new grading scale and teaching dis-improves, then the effects we find could be driven by teacher behaviour as opposed to more informative grading. We argue that as the knowledge requirements for each grade level were very similar under the old and new scales (see appendix tables B1 to B3), teachers should not have had much difficulty in adjusting to using letter grades. In addition, as mentioned in section 2.2, the results of a survey of a large number of teachers showed that teaching practices did not change after the introduction of the Lgr11 reform (Wahlström and Sundberg, 2015).

7 Conclusion

A large literature exists examining how feedback plays a role in inducing greater performance both in the workplace and in education. The education literature tends to focus on how relative grading feedback increases students' motivation, relative to absolute grading. Grading systems remain, however, predominantly based on absolute levels of achievement and little light has been shed on the information content of absolute grading systems. In this paper we contribute to the literature by identifying a natural experiment which allows us to examine how the information content of grades, in terms of the number of discrete points on the grading scale, affects the education and labour market outcomes of students. We do this using Swedish administrative data comprising the population of Sweden born between 1991 and 1997. We employ a fuzzy difference-in-discontinuity design, assuming that, conditional on date of birth, exposure to more informative grading is as good as random and discontinuities in outcomes across school cohort admissions cut-offs caused by school starting age effects are consistent over time. We provide a number of specification checks that support this assumption.

We find that exposure to more informative grading has negative consequences for students. We estimate that those exposed to more informative grading are 2.6 percentage points less likely to go on to graduate from high school and 3.5 and 2 percentage points less likely to complete an academic track or an academic STEM track in high school, respectively. We find that the effect on STEM participation persists to the university level, with more informative grading causing a 1.4 percentage point reduction in the likelihood of STEM enrolment. Using a surrogate index approach, we estimate a long-term negative effect on income between the ages of 28-30 of approximately 1.8%.

Introducing a new data set and empirical strategy, we examine two mechanisms through which more informative grading affects outcomes. Treated students are found to be less likely to report themselves as being good or very good academically, relative to their peers, suggesting that with more information, students are relatively less optimistic about their abilities. More informative grading thus seems to provide a negative shock to self-confidence. Moreover, we identify that the treated students are more likely to report feeling pressure from their school work and to report stress-related mental health issues.

Our results give interesting insights into how receiving more informative feedback in one's formative years can have large and persistent impacts on outcomes, with important implications for the framing of feedback in the education setting. The findings of our research indicate that more information acts as a negative shock to people's self-belief, increasing stress-levels and leading, in turn, to worse educational outcomes and lower income later in life. While our results do not speak to whether children should be graded at all, when it comes to the information content of grades, less is more.

References

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization*, 100:44–63.

Athey, S., Chetty, R., Imbens, G. W., and Kang, H. (2019). The surrogate index: Combining short-term proxies to estimate long term treatment effects more rapidly and precisely. *NBER Working Paper*.

Azmat, G., Bagues, M., Cabrales, A., and Iriberri, N. (2019). What you don't know… can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science*, 65(8):3714–3736.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7-8):435–452.

Azmat, G. and Iriberri, N. (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics & Management Strategy*, 25(1):77–110.

Bandiera, O., Larcinese, V., and Rasul, I. (2015). Blissful ignorance? a natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34:13–25.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, 93(2):455–467.

Blanes i Vidal, J. and Nossol, M. (2011). Tournaments without prizes: Evidence from personnel records. *Management science*, 57(10):1721–1736.

Bobba, M. and Frisancho, V. (2016). Learning about oneself: The effects of performance feedback on school choice.

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, 62(11):3085–3099.

Campos-Mercade, P. and Wengström, E. (2015). Threshold incentives and academic performance.

Chevalier, A. and Feinstein, L. (2006). Sheepskin or prozac: The causal effect of education on mental health.

Chevalier, A., Gibbons, S., Thorpe, A., Snell, M., and Hoskins, S. (2009). Students' academic self-perception. *Economics of Education Review*, 28(6):716–727.

Cobb-Clark, D. A., Ho, T., and Salamanca, N. (2021). Parental responses to children's achievement test results.

Cornaglia, F., Crivellaro, E., and McNally, S. (2015). Mental health and education decisions. *Labour Economics*, 33:1–12.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review*, 109(8):2728–65.

Dobrescu, I., Faravelli, M., Megalokonomou, R., and Motta, A. (2021). Relative performance feedback in education: Evidence from a randomised controlled trial. *The Economic Journal*.

Eisenberg, D., Golberstein, E., and Hunt, J. B. (2009). Mental health and academic success in college. *The BE Journal of Economic Analysis & Policy*, 9(1).

Elsner, B. and Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3):787–828.

Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688.

Fischer, M. and Sliwka, D. (2018). Confidence in knowledge or confidence in the ability to learn: An experiment on the causal effects of beliefs on motivation. *Games and Economic Behavior*, 111:122–142.

Fischer, M. and Wagner, V. (2018). Effects of timing and reference frame of feedback: Evidence from a field experiment. *IZA Discussion Paper*.

Fredriksson, P. and Öckert, B. (2014). Life-cycle effects of age at school start. *The Economic Journal*, 124(579):977–1004.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1):249–285.

Gelman, A. and Imbens, G. (2019). Why higher-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

Gill, D., Kissová, Z., Lee, J., and Prowse, V. (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, 65(2):494–507.

Goulas, S. and Megalokonomou, R. (2015). Knowing who you are: The effect of feedback information on short and long term outcomes.

Hall, C. (2012). The effects of reducing tracking in upper secondary school: Evidence from a large-scale pilot scheme. *Journal of Human Resources*, 47(1):237–269.

Hannan, R. L., McPhee, G. P., Newman, A. H., and Tafkov, I. D. (2013). The effect of relative performance information on performance and effort allocation in a multi-task environment. *The Accounting Review*, 88(2):553–575.

Jalava, N., Joensen, J. S., and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior & Organization*, 115:161–196.

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

Kuhnen, C. M. and Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.

Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing self-confidence. *NBER Working Paper*.

OECD (2020). Education at a glance 2020: OECD indicators. Technical report, OECD Publishing, Paris.

SFS (2010). Lag (2010:801) om införande av skollagen (2010:800). Stockholm: Justitiedepartementet.

Sjögren, A. (2010). Graded children: Evidence of longrun consequences of school grades from a nationwide reform. *IFAU Working Paper*.

Skolverket (2016). Utvärdering av den nya betygsskalan samt kunskapskravens utformning. Technical report, Skolverket.

SOU (2007). Sou 2007:28: Tydliga mål och kunskapskrav i grundskolan: Förslag till nytt mål- och uppföljningssystem.

Stinebrickner, R. and Stinebrickner, T. (2014). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3):601–644.

Tafkov, I. D. (2013). Private and public relative performance information under different compensation contracts. *The Accounting Review*, 88(1):327–350.

Thaler, R., Sunstein, C., and Balz, J. (2013). Choice architecture. In Shafir, E., editor, *The Behavioral Foundations of Public Policy*, chapter 25, pages 428–439. North-Holland.

Tran, A. and Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96(9-10):645–650.

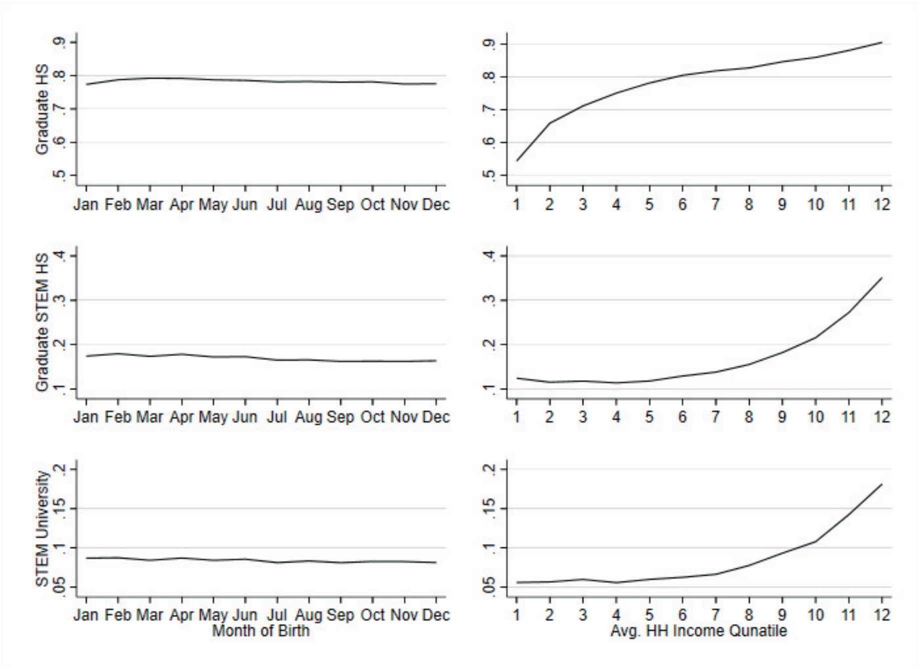
Utbildningsdepartementet (2009). Uppdrag att utarbeta nya kursplaner och kunskapskrav för grundskolan och motsvarande skolformer m.m. Technical report, Sveriges Regering.

Utbildningsdepartmentet (2008). Proposition 2008/09:66. Technical report, Sveriges Regering.

Wahlström, N. and Sundberg, D. (2015). Theory-based evaluation of the curriculum Lgr 11. *IFAU Working Paper*, page 50.

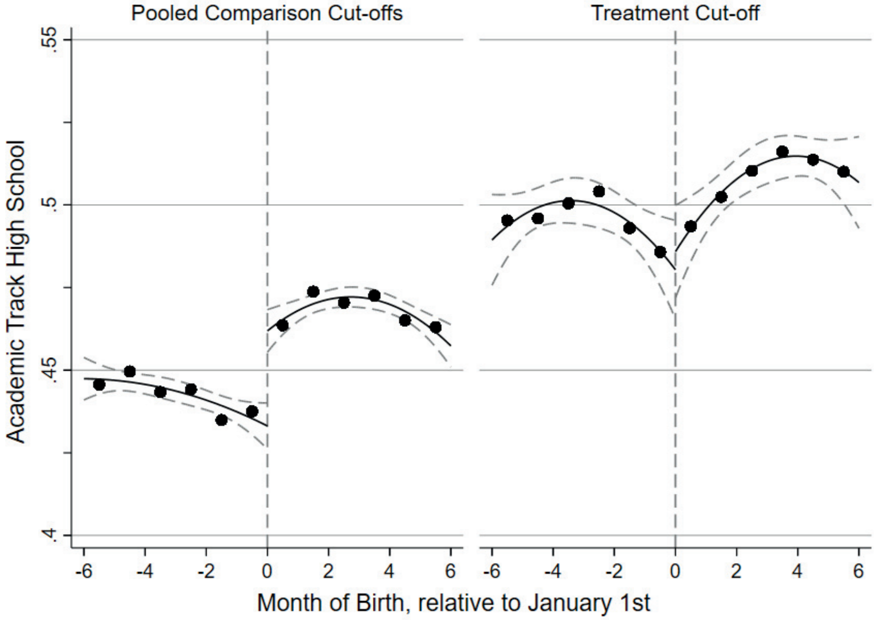
Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, 110(2):337–61.

Appendix A: Further Tables and Figures



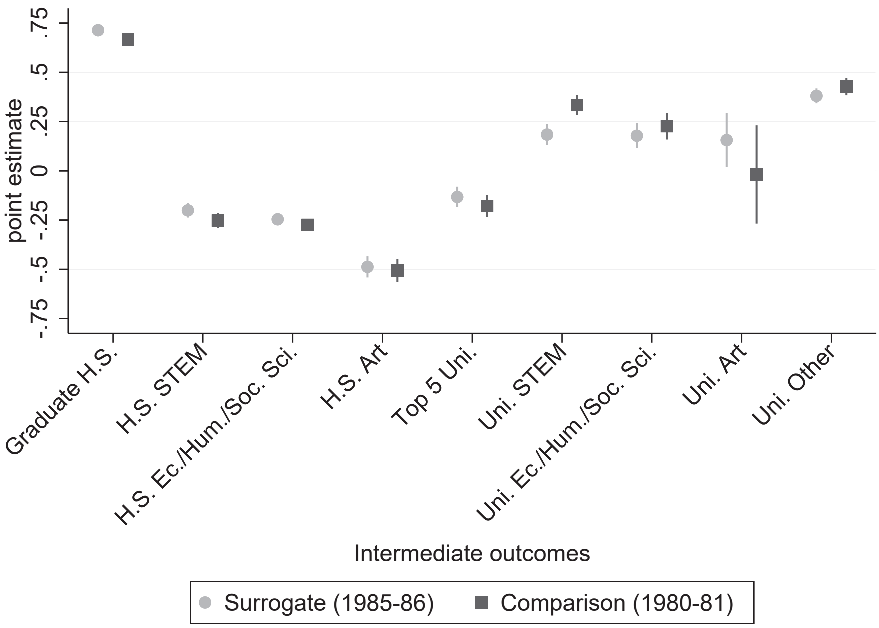
Notes: This figure plots the unconditional relationship between a sample of our outcomes and month of birth in the left panel and in the right panel shows the same for quantiles of average household income. Average household income is calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000.

Figure A1: Comparing Birth Month and Income Quantile Gradients



Notes: This figure plots graduation from an academic high school track (as an example outcome) by month of birth (dots), quadratic fits through 6 month bandwidths of either side of each year's January 1st cut-off (solid lines) and robust 95% confidence intervals (dotted lines).

Figure A2: Quadratic Functional Form



Notes: This figure plots the coefficients from regressions of log average earnings between the ages of 28 and 30 and the covariates included in our surrogate index and robust 95% confidence intervals.

Figure A3: Comparability of the Surrogate Index

Table A1: Covariate Balance

Panel A: Density of the running variable		
Total monthly births	–814.499	(511.891)
N	60	
Panel B: Balance Test		
School cohort assignment	–0.003	(0.004)
Sex	0.005	(0.009)
Number of siblings	0.040	(0.027)
Birth order	0.005	(0.022)
Twins	0.004	(0.003)
Triplets or more	0.001***	(0.000)
Immigrant (1st generation)	0.003	(0.005)
Immigrant (2nd generation)	0.006	(0.005)
Adoptive parents	0.003	(0.002)
Two adoptive parents	0.004**	(0.002)
Mother has HS degree	–0.009	(0.005)
Mother has Uni. Degree	–0.010	(0.008)
Father has HS degree	–0.002	(0.006)
Father has Uni. degree	–0.005	(0.007)
Mother, STEM	–0.007	(0.005)
Father, STEM	0.003	(0.009)
Household income	–1238.154	(5793.242)
Class size	0.082	(0.095)
N	547,508	

Notes: Robust standard errors in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. School cohort assignment refers to whether a student is assigned to the younger of two school cohorts surrounding the closest January 1st to a child's date of birth. Average household income is calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. The reported number of observations relates to the maximum number of observations in the sample. Due to missing information, fewer observations are included for three variables, namely 'Mother, STEM' (3,543 missing), 'Father, STEM' (14,088 missing) and 'Class size' (40,557 missing).

Table A2: University Outcomes - Alternative Definition of Tracks

	(1) Stem track	(2) Econ./Soc.Sci./ Hum. Track	(3) Art Track	(4) Other track
Treatment	–0.012**	–0.004	–0.000	0.005
	(0.005)	(0.005)	(0.001)	(0.004)
Pre-reform mean	.103	.116	.00396	.0674
N	541,302	541,302	541,302	541,302

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. University outcomes are observed at the end of the spring semester of the year an individual turns 21. The division into tracks is based on the governments definition of tracks that form the basis of their funding to universities. We assign tracks to students according to the field of study in which a student has taken the most credits.

Table A3: Heterogeneity by Immigrant Status

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
Panel A: High School Outcomes					
Treatment	–0.021***	–0.034***	–0.022***	0.006	–0.018***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.005)
Non-native	–0.157***	–0.053***	0.023***	–0.026***	–0.054***
	(0.008)	(0.009)	(0.007)	(0.007)	(0.003)
Treatment*Non-native	–0.016	–0.004	0.013	–0.032	0.017*
	(0.026)	(0.027)	(0.021)	(0.022)	(0.009)
Pre-reform: Native	.814	.456	.163	.23	.0594
Pre-reform: Non-native	.669	.455	.192	.236	.0201
N	545,760	545,760	545,760	545,760	545,760

	(6) Enrolled Spring Age 21	(7) In top 5 University	(8) Number of credits enr. in	(9) Number of credits passed
Panel B: University Outcomes				
Treatment	–0.016*	–0.000	–0.953*	–0.921**
	(0.008)	(0.006)	(0.543)	(0.382)
Non-native	0.083***	0.018***	4.744***	2.560***
	(0.008)	(0.006)	(0.642)	(0.418)
Treatment*Non-native	0.051**	–0.005	1.577	1.573
	(0.025)	(0.018)	(1.888)	(1.244)
Pre-reform: Native	.268	.124	12.6	8.8
Pre-reform: Non-native	.336	.146	18.1	11.7
N	541,302	541,302	541,302	541,302

	(10) Stem track	(11) Econ./Soc.Sci./ Hum. Track	(12) Art Track	(13) Other track
Treatment	–0.016***	–0.002	0.000	–0.001
	(0.005)	(0.004)	(0.001)	(0.005)
Non-native	0.017***	0.024***	–0.002	0.058***
	(0.005)	(0.005)	(0.001)	(0.006)
Treatment*Non-native	0.014	0.019	–0.003	0.010
	(0.016)	(0.014)	(0.003)	(0.018)
Pre-reform: Native	.0817	.049	.00624	.0807
Pre-reform: Non-native	.0965	.0743	.0039	.128
N	508,015	508,015	508,015	508,015

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. Non-native is defined as being a first or second generation immigrant. High school outcomes are observed only for those finishing high school. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

Table A4: Excluding Comparison Years - High School Outcomes

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
1992	–0.025***	–0.038***	–0.021***	–0.002	–0.016***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.004)
N	425,219	425,219	425,219	425,219	425,219
1993	–0.027***	–0.035***	–0.021***	0.000	–0.016***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.004)
N	430,075	430,075	430,075	430,075	430,075
1994	–0.030***	–0.034***	–0.021***	0.002	–0.016***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.004)
N	433,078	433,078	433,078	433,078	433,078
1996	–0.023***	–0.033***	–0.018***	0.002	–0.016***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.004)
N	445,181	445,181	445,181	445,181	445,181

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. The left-most column details the control cut-off excluded from the analysis. Estimates are taken from separate difference-in-discontinuity regressions for each outcome and control year excluded. High school outcomes are observed only for those finishing high school.

Table A5: Excluding Control Years - University Outcomes

	(1) Enrolled Spring Age 21	(2) In top 5 University	(3) Number of credits enr. in	(4) Number of credits passed
1992	–0.009	–0.002	–0.747	–0.659*
	(0.008)	(0.006)	(0.554)	(0.382)
N	422,083	422,083	422,083	422,083
1993	–0.010	–0.003	–0.743	–0.795**
	(0.008)	(0.006)	(0.553)	(0.382)
N	426,734	426,734	426,734	426,734
1994	–0.008	0.001	–0.498	–0.568
	(0.008)	(0.006)	(0.552)	(0.382)
N	429,575	429,575	429,575	429,575
1996	–0.006	–0.001	–0.577	–0.596
	(0.008)	(0.006)	(0.551)	(0.380)
N	441,279	441,279	441,279	441,279
	(5) Stem track	(6) Econ./Soc.Sci./ Hum. Track	(7) Art Track	(8) Other track
1992	–0.013**	0.002	0.000	–0.001
	(0.005)	(0.004)	(0.001)	(0.005)
N	397,309	397,309	397,309	397,309
1993	–0.014***	0.000	–0.000	0.000
	(0.005)	(0.004)	(0.001)	(0.005)
N	401,226	401,226	401,226	401,226
1994	–0.015***	0.002	–0.000	0.002
	(0.005)	(0.004)	(0.001)	(0.005)
N	403,252	403,252	403,252	403,252
1996	–0.012**	0.000	–0.000	0.001
	(0.005)	(0.004)	(0.001)	(0.005)
N	413,227	413,227	413,227	413,227

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. The left-most column details the control cut-off excluded from the analysis. Estimates are taken from separate difference-in-discontinuity regressions for each outcome and control year excluded. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97.* We assign tracks to students according to the field of study in which a student has taken the most credits.

Table A6: Using only 1995/96 as a Comparison Cut-off

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
Panel A: High School Outcomes					
Treatment	–0.036***	–0.043***	–0.028***	–0.003	–0.014***
	(0.010)	(0.011)	(0.008)	(0.010)	(0.005)
N					
prabove					
N	196,852	196,852	196,852	196,852	196,852

	(6) Enrolled Spring Age 21	(7) In top 5 University	(8) Number of credits enr. in	(9) Number of credits passed
Panel B: University Outcomes				
Treatment	–0.017*	–0.000	–0.864	–0.857*
	(0.010)	(0.007)	(0.683)	(0.471)
N				
prabove				
N	195,788	195,788	195,788	195,788

	(10) Stem track	(11) Econ./Soc.Sci./ Hum. Track	(12) Art Track	(13) Other track
Treatment	–0.020***	0.004	0.000	–0.001
	(0.006)	(0.005)	(0.002)	(0.006)
N				
prabove				
N	185,757	185,757	185,757	185,757

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. High school outcomes are observed only for those finishing high school. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

Table A7: Further Robustness Checks - High School Outcomes

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
6 Month Bandwidth	–0.026**	–0.026*	–0.024**	0.014	–0.016**
	(0.012)	(0.013)	(0.010)	(0.012)	(0.006)
N	260,780	260,780	260,780	260,780	260,780
Quadratic functional form	–0.028**	–0.030**	–0.023**	0.013	–0.020***
	(0.012)	(0.014)	(0.011)	(0.012)	(0.007)
N	545,760	545,760	545,760	545,760	545,760
Adding covariates	–0.022***	–0.032***	–0.018***	0.001	–0.015***
	(0.007)	(0.008)	(0.006)	(0.007)	(0.004)
N	529,355	529,355	529,355	529,355	529,355
Reduced form	–0.022***	–0.030***	–0.017***	0.001	–0.013***
	(0.006)	(0.007)	(0.006)	(0.006)	(0.004)
N	545,760	545,760	545,760	545,760	545,760
Triangular Kernel	–0.027***	–0.032***	–0.021***	0.007	–0.018***
	(0.009)	(0.010)	(0.008)	(0.009)	(0.005)
N	450,725	450,725	450,725	450,725	450,725

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate difference-in-discontinuity regressions for each outcome and specified robustness test. High school outcomes are observed only for those finishing high school. Covariates included are sex, number of siblings, birth order, whether a child is a twin or part of a higher order multiple birth, whether a child has one or two adoptive parents, whether a child is a first or second generation immigrant, mother and father's highest level of education, whether a child's mother or father studied a STEM subject in their highest level of education and average household income. Average household income is calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000.

Table A8: Further Robustness Checks - University Outcomes

	(1) Enrolled Spring Age 21	(2) In top 5 University	(3) Number of credits enr. in	(4) Number of credits passed
6 Month Bandwidth	–0.012	0.000	–1.548*	–1.186**
	(0.012)	(0.009)	(0.823)	(0.568)
N	258,653	258,653	258,653	258,653
Quadratic functional form	–0.014	0.000	–1.708*	–1.383**
	(0.013)	(0.009)	(0.873)	(0.602)
N	541,302	541,302	541,302	541,302
Adding covariates	–0.008	–0.001	–0.577	–0.606*
	(0.007)	(0.006)	(0.528)	(0.366)
N	525,236	525,236	525,236	525,236
Reduced form	–0.007	–0.001	–0.541	–0.552*
	(0.007)	(0.005)	(0.451)	(0.312)
N	541,302	541,302	541,302	541,302
Triangular Kernel	–0.011	–0.001	–1.153*	–1.003**
	(0.009)	(0.007)	(0.638)	(0.439)
N	447,050	447,050	447,050	447,050

	(5) Stem track	(6) Econ./Soc.Sci./ Hum. Track	(7) Art Track	(8) Other track
6 Month Bandwidth	–0.016**	0.007	–0.001	–0.006
	(0.008)	(0.006)	(0.002)	(0.008)
N	242,763	242,763	242,763	242,763
Quadratic functional form	–0.017**	0.013**	–0.001	–0.015*
	(0.008)	(0.007)	(0.002)	(0.008)
N	508,015	508,015	508,015	508,015
Adding covariates	–0.013***	0.002	0.000	0.001
	(0.005)	(0.004)	(0.001)	(0.005)
N	492,535	492,535	492,535	492,535
Reduced form	–0.011***	0.001	–0.000	0.000
	(0.004)	(0.003)	(0.001)	(0.004)
N	508,015	508,015	508,015	508,015
Triangular Kernel	–0.015**	0.007	–0.000	–0.007
	(0.006)	(0.005)	(0.001)	(0.006)
N	419,510	419,510	419,510	419,510

Notes: Robust standard errors in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Estimates are taken from separate difference-in-discontinuity regressions for each outcome and specified robustness test. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits. Covariates included are sex, number of siblings, birth order, whether a child is a twin or part of a higher order multiple birth, whether a child has one or two adoptive parents, whether a child is a first or second generation immigrant, mother and father's highest level of education, whether a child's mother or father studied a STEM subject in their highest level of education and average household income. Average household income is calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000.

Table A9: Attitudes to School, including Pre-Trends

	(1)	(2)	(3)
	Very good	Good/ Very good	Average/ Good/ Very good
Academic Ability			
Born 1990	0.013	0.021	–0.002
	(0.028)	(0.034)	(0.015)
Born 1994	0.000	0.000	0.000
	(.)	(.)	(.)
Born 1998	–0.078***	–0.180***	–0.023
	(0.029)	(0.033)	(0.015)
Pre-reform mean	.145	.581	.934
N	25,241	25,241	25,241
	A lot	A little/ A lot	Not much/ A little/ A lot
Like School			
Born 1990	0.029	–0.038	–0.026
	(0.037)	(0.031)	(0.017)
Born 1994	0.000	0.000	0.000
	(.)	(.)	(.)
Born 1998	0.050	–0.008	–0.024
	(0.038)	(0.030)	(0.015)
Pre-reform mean	.125	.682	.908
N	25,241	25,241	25,241
	A lot	A little/ A lot	Some/ A little/ A lot
School Pressure			
Born 1990	–0.008	0.092**	0.013
	(0.025)	(0.040)	(0.030)
Born 1994	0.000	0.000	0.000
	(.)	(.)	(.)
Born 1998	0.013	0.155***	0.121***
	(0.026)	(0.037)	(0.032)
Pre-reform mean	.2	.482	.924
N	25,241	25,241	25,241

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate triple-differences regressions for each outcome using the other Scandinavian countries of Denmark and Norway as a control group. Identification exploits differences in the outcomes of 15 year old and 11 year old children from the 1998 birth cohort in Sweden relative to the same changes in earlier birth cohorts in Sweden. This difference is further compared to the same differences in the control countries. All models include controls for gender and whether a child considers their family to be well-off.

Table A10: Self-reported Mental Health, including Pre-Trends

	(1) Monthly / more often	(2) Weekly / more often	(3) More than weekly / more often	(4) Daily / more often
Irritable or Bad Temper				
Born 1990	–0.040	–0.025	–0.027	0.002
	(0.032)	(0.038)	(0.030)	(0.017)
Born 1994	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Born 1998	–0.010	0.020	0.031	0.013
	(0.033)	(0.037)	(0.028)	(0.016)
Pre-reform mean	.88	.606	.321	.0729
N	25,241	25,241	25,241	25,241
Nervous				
Born 1990	–0.029	–0.028	0.020	–0.017
	(0.036)	(0.034)	(0.025)	(0.012)
Born 1994	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Born 1998	–0.041	–0.017	0.027	–0.009
	(0.038)	(0.035)	(0.026)	(0.012)
Pre-reform mean	.734	.41	.173	.0389
N	25,241	25,241	25,241	25,241
Difficulty Sleeping				
Born 1990	–0.010	–0.011	–0.026	–0.011
	(0.037)	(0.035)	(0.031)	(0.022)
Born 1994	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Born 1998	0.092**	0.107***	0.091***	0.051**
	(0.040)	(0.035)	(0.030)	(0.021)
Pre-reform mean	.642	.42	.245	.0931
N	25,241	25,241	25,241	25,241

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate triple-differences regressions for each outcome using the other Scandinavian countries of Denmark and Norway as a control group. Identification exploits differences in the outcomes of 15 year old and 11 year old children from the 1998 birth cohort in Sweden relative to the same changes in earlier birth cohorts in Sweden. This difference is further compared to the same differences in the control countries. All models include controls for gender and whether a child considers their family to be well-off.

Table A11: Relationship between Irritability and Nervousness and Sleeping Difficulties

	Difficulty Sleeping			
	(1)	(2)	(3)	(4)
	Monthly / more often	Weekly / more often	More than weekly / more often	Daily / more often
Irritability or bad temper				
About every month	0.142***	0.071***	0.020***	–0.003
	(0.008)	(0.007)	(0.006)	(0.004)
About every week	0.231***	0.190***	0.108***	0.024***
	(0.009)	(0.009)	(0.007)	(0.005)
More than once a week	0.279***	0.284***	0.225***	0.095***
	(0.010)	(0.010)	(0.009)	(0.007)
About every day	0.324***	0.359***	0.327***	0.232***
	(0.013)	(0.014)	(0.014)	(0.013)
Feeling nervous				
About every month	0.117***	0.061***	0.002	–0.015***
	(0.007)	(0.007)	(0.006)	(0.004)
About every week	0.200***	0.160***	0.078***	0.012**
	(0.009)	(0.009)	(0.008)	(0.006)
More than once a week	0.237***	0.257***	0.207***	0.088***
	(0.010)	(0.011)	(0.011)	(0.009)
About every day	0.260***	0.321***	0.323***	0.264***
	(0.014)	(0.016)	(0.017)	(0.017)
N	26,100	26,100	26,100	26,100

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates in each column are taken from separate regressions for each outcome using "Rarely or never" as reference category for both irritability and bad temper and feeling nervous.

Table A12: Heterogeneity by Household Income

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
Panel A: High School Outcomes					
Treatment	–0.034***	–0.036***	–0.015*	–0.009	–0.012**
	(0.013)	(0.013)	(0.009)	(0.010)	(0.006)
Above Median	0.183***	0.230***	0.099***	0.113***	0.019***
	(0.005)	(0.006)	(0.004)	(0.005)	(0.003)
Treatment*Above Median	0.012	0.005	–0.007	0.018	–0.006
	(0.015)	(0.017)	(0.013)	(0.015)	(0.008)
Pre-reform: below median	.724	.347	.119	.177	.048
Pre-reform: above median	.865	.57	.218	.288	.0597
N	545,760	545,760	545,760	545,760	545,760

	(6) Enrolled Spring Age 21	(7) In top 5 University	(8) Number of credits enr. in	(9) Number of credits passed
Panel B: University Outcomes				
Treatment	–0.000	–0.003	0.386	0.086
	(0.011)	(0.008)	(0.775)	(0.518)
Above Median	0.118***	0.074***	5.523***	4.157***
	(0.005)	(0.004)	(0.360)	(0.245)
Treatment*Above Median	–0.009	0.006	–1.553	–1.147
	(0.015)	(0.011)	(1.070)	(0.734)
Pre-reform: below median	.218	.0898	10.7	7.22
Pre-reform: above median	.34	.166	16.2	11.3
N	541,302	541,302	541,302	541,302

	(10) Stem track	(11) Econ./Soc.Sci./ Hum. Track	(12) Art Track	(13) Other track
Treatment	0.000	0.004	–0.001	–0.003
	(0.006)	(0.006)	(0.002)	(0.008)
Above Median	0.050***	0.024***	0.002**	0.019***
	(0.003)	(0.003)	(0.001)	(0.003)
Treatment*Above Median	–0.023**	–0.004	0.001	0.008
	(0.010)	(0.008)	(0.002)	(0.010)
Pre-reform: below median	.0586	.0393	.00498	.0791
Pre-reform: above median	.111	.0672	.0069	.0967
N	508,015	508,015	508,015	508,015

Notes: Robust standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. The sample is split at the median of average household income, calculated as the average yearly income of a child's parents in the three years prior to that child's entry to 9th grade and is adjusted according to Statistics Sweden's Consumer Price Index to the year 2000. High school outcomes are observed only for those finishing high school. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

Table A13: Heterogeneity by Parental Education

	(1) Graduate High School	(2) Academic track	(3) Stem Track	(4) Econ./Soc.Sci./ Hum. Track	(5) Art Track
Panel A: High School Outcomes					
Treatment	–0.029***	–0.034***	–0.011	–0.009	–0.015***
	(0.011)	(0.011)	(0.007)	(0.009)	(0.005)
Parent has Degree	0.148***	0.292***	0.150***	0.109***	0.029***
	(0.005)	(0.006)	(0.004)	(0.005)	(0.003)
Treatment*Parent has Degree	0.013	0.008	–0.016	0.027*	–0.001
	(0.015)	(0.017)	(0.014)	(0.015)	(0.009)
Pre-reform: No degree	.743	.333	.106	.184	.0411
Pre-reform: Parent has degree	.859	.618	.249	.293	.0703
N	544,968	544,968	544,968	544,968	544,968

	(6) Enrolled Spring Age 21	(7) In top 5 University	(8) Number of credits enr. in	(9) Number of credits passed
Panel B: University Outcomes				
Treatment	–0.011	–0.007	–1.073*	–0.913**
	(0.009)	(0.006)	(0.630)	(0.430)
Parent has Degree	0.197***	0.123***	9.965***	6.877***
	(0.005)	(0.004)	(0.372)	(0.253)
Treatment*Parent has Degree	0.013	0.017	1.325	0.828
	(0.016)	(0.012)	(1.100)	(0.761)
Pre-reform: No degree	.196	.0751	9.56	6.54
Pre-reform: Parent has degree	.386	.196	18.5	12.8
N	540,533	540,533	540,533	540,533

	(10) Stem track	(11) Econ./Soc.Sci./ Hum. Track	(12) Art Track	(13) Other track
Treatment	–0.006	–0.001	–0.002	–0.005
	(0.005)	(0.005)	(0.001)	(0.006)
Parent has Degree	0.079***	0.036***	0.004***	0.048***
	(0.004)	(0.003)	(0.001)	(0.004)
Treatment*Parent has Degree	–0.015	0.006	0.004*	0.015
	(0.011)	(0.008)	(0.003)	(0.011)
Pre-reform: No degree	.0533	.038	.00429	.069
Pre-reform: Parent has degree	.127	.0734	.00818	.114
N	507,256	507,256	507,256	507,256

Notes: Robust standard errors in parentheses. * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Estimates are taken from separate difference-in-discontinuity regressions for each outcome. The sample is split according to whether one or both of a child's parents holds a university degree. High school outcomes are observed only for those finishing high school. University outcomes are observed at the end of the spring semester of the year an individual turns 21, while credit enrolment is observed at the end of the autumn semester of the year an individual turns 20. Variations in sample size are due to missing values in the data. The universities considered as top-5 are chosen according to total applications to those universities. These are, in alphabetical order, Gothenburg University, Lund University, Stockholm University, Umeå University and Uppsala University. Track choices are defined according to *SUN2000*, which is the Swedish Education Nomenclature defined by Statistics Sweden. The *SUN2000* classification is constructed to be comparable with the *ISCED97*. We assign tracks to students according to the field of study in which a student has taken the most credits.

Appendix B: Further Information on Lgr11

Table B1: Comparison of the knowledge requirements in the old and new syllabus in Swedish - reading

Old syllabus	New syllabus
"The student is able to read age-appropriate fiction from Sweden, the Nordic countries and other countries as well as non-fiction and newspaper text about general subjects and is able to retell and reflect on the content in a coherent way."	"The student is able to read fiction and non-fiction ... Moreover, the student can ... conduct simple and to some extent substantiated reasoning about clearly prominent messages in different works"
["Eleven skall kunna läsa till åldern avpassad skönlitteratur från Sverige, Norden och från andra länder samt saklitteratur och tidningstext om allmänna ämnen, kunna återge innehållet sammanhängande samt kunna reflektera över det."]	["Eleven kan läsa skönlitteratur och sakprosatexter ... Dessutom kan eleven ... föra enkla och till viss del underbyggda resonemang om tydligt framträdande budskap i olika verk."]
"The student is able to read, reflect on and put into context some works of fiction and authorship with significance for people's way of life and thinking"	"The student is also able to conduct simple reasoning about the work and it's connection to the author. The student then draws ... conclusions about how the work has been influenced by the historical and cultural context in which it has been created"
["Eleven skall kunna läsa, reflektera över och sätta in i ett sammanhang några skönlitterära verk och författarskap med betydelse för människors sätt att leva och tänka."]	["Eleven kan också föra enkla resonemang om verket med kopplingar till dess upphovsman. Eleven drar då ... slutsatser om hur verket har påverkats av det historiska och kulturella sammanhang som det har tillkommit i."]

New syllabus - original text
"Eleven kan läsa skönlitteratur och sakprosatexter med flyt genom att, på ett i huvudsak fungerande sätt, välja och använda lässtrategier utifrån olika texters särdrag. Genom att göra enkla sammanfattningar av olika texters innehåll med viss koppling till tidsaspekter, orsakssamband och andra texter visar eleven grundläggande läsförståelse. Dessutom kan eleven utifrån egna erfarenheter, olika livsfrågor och omvärldsfrågor tolka och föra enkla och till viss del underbyggda resonemang om tydligt framträdande budskap i olika verk. Eleven kan också föra enkla resonemang om verket med kopplingar till dess upphovsman. Eleven drar då till viss del underbyggda slutsatser om hur verket har påverkats av det historiska och kulturella sammanhang som det har tillkommit i."

Notes: This table displays the knowledge requirements for reading in 9th grade Swedish in the old and new syllabi. The left panel displays the knowledge requirements from the old syllabus, which were listed as an individual bullet points and the right panel displays the equivalent requirements in the new syllabus, which are adapted from a longer form text. The complete paragraph, including the adapted knowledge requirement from the new syllabus, is listed in the bottom panel. All translations are made by the authors and the original text is displayed in square brackets below the translation.[a]

[a] The old and new knowledge requirements can be found in the old curriculum "1994 års läroplan för det obligatoriska skolväsendet" (Lpo94) and the new curriculum "Läroplan för grundskolan, förskoleklassen och fritidshemmet" (Lgr11), respectively. Both curriculum are accessible on the web page of the Swedish national agency for education.

Table B2: Comparison of the knowledge requirements in the old and new syllabus in Swedish - speaking

Old syllabus	New syllabus
"The students' knowledge about the language is sufficient for the student to make observations about one's own and other's use of language." ["Eleven skall ha kunskaper om språket som gör det möjligt att göra iakttagelser av eget och andras språkbruk."]	"The student is able to conduct ... reasoning about the history, origin and the distinctive features of the Swedish language as well as compare these with related languages and describe clearly salient similarities and differences." ["Eleven kan föra ... resonemang om svenska språkets historia, ursprung och särdrag samt jämföra med närliggande språk och beskriva tydligt framträdande likheter och skillnader."]
"The student is actively participating in conversations and discussions and understands other people's way of thinking as well as being able to make an oral presentation so that the content is clear and understandable." ["Eleven skall aktivt kunna delta i samtal och diskussioner och sätta sig in i andras tankar samt kunna redovisa ett arbete muntligt så att innehållet framgår och är begripligt."]	"The student is able to participate in conversations and discussions on various topics ... Moreover, the student is able to prepare and carry out simple oral presentations with a mainly functional structure and content..." ["Eleven kan samtala om och diskutera varierande ämnen . . . . Dessutom kan eleven förbereda och genomföra enkla muntliga redogörelser med i huvudsak fungerande struktur och innehåll. . . "]

New syllabus - original text
"Eleven kan samtala om och diskutera varierande ämnen genom att ställa frågor och framföra åsikter med enkla och till viss del underbyggda argument på ett sätt som till viss del för samtalen och diskussionerna framåt. Dessutom kan eleven förbereda och genomföra enkla muntliga redogörelser med i huvudsak fungerande struktur och innehåll och viss anpassning till syfte, mottagare och sammanhang. Eleven kan föra enkla och till viss del underbyggda resonemang om svenska språkets historia, ursprung och särdrag samt jämföra med närliggande språk och beskriva tydligt framträdande likheter och skillnader."

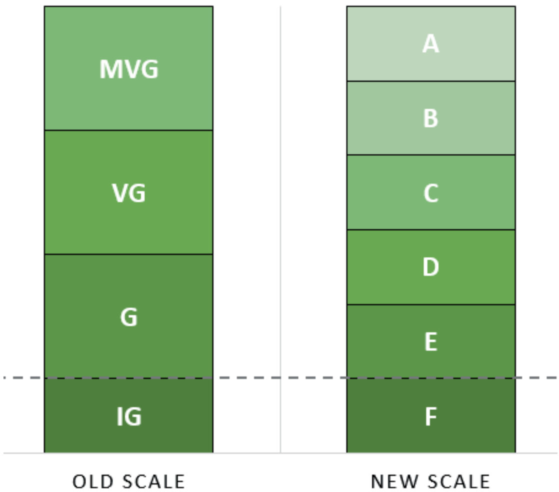
Notes: This table displays the knowledge requirements for speaking in 9th grade Swedish in the old and new syllabi. The left panel displays the knowledge requirements from the old syllabus, which were listed as an individual bullet points and the right panel displays the equivalent requirements in the new syllabus, which are adapted from a longer form text. The complete paragraph, including the adapted knowledge requirement from the new syllabus, is listed in the bottom panel. All translations are made by the authors and the original text is displayed in square brackets below the translation.[a]

Table B3: Comparison of the knowledge requirements in the old and new syllabus in Swedish - writing

Old syllabus	New syllabus
"The student is able to write different forms of text such that the content is clear as well as being able to apply linguistic norms, both when writing by hand and using a computer."	"The student is able to write different forms of text with ... a functional adaptation to the form, linguistic norms and structure."
["Eleven skall kunna skriva olika sorters texter så att innehållet framgår tydligt samt tillämpa skriftspråkets normer, både vid skrivande för hand och med dator"]	["Eleven kan skriva olika slags texter med ... fungerande anpassning till texttyp, språkliga normer och strukturer"]

New syllabus - original text
"Eleven kan skriva olika slags texter med viss språklig variation, enkel textbindning samt i huvudsak fungerande anpassning till texttyp, språkliga normer och strukturer. De berättande texter eleven skriver innehåller enkla gestaltande beskrivningar och berättargrepp samt dramaturgi med enkel uppbyggnad. Eleven kan söka, välja ut och sammanställa information från ett avgränsat urval av källor och för då enkla och till viss del underbyggda resonemang om informationens och källornas trovärdighet och relevans. Sammanställningarna innehåller enkla beskrivningar och förklaringar, enkelt ämnesrelaterat språk samt i huvudsak fungerande struktur, citat och källhänvisningar. Genom att kombinera olika texttyper, estetiska uttryck och medier så att de olika delarna samspelar på ett i huvudsak fungerande sätt kan eleven förstärka och levandegöra sina texters budskap. Dessutom kan eleven ge enkla omdömen om texters innehåll och uppbyggnad och utifrån respons bearbeta texter mot ökad tydlighet, kvalitet och uttrycksfullhet på ett i huvudsak fungerande sätt."

Notes: This table displays the knowledge requirements for writing in 9th grade Swedish in the old and new syllabi. The left panel displays the knowledge requirement from the old syllabus, which was listed as an individual bullet point and the right panel displays the equivalent requirement in the new syllabus, which is adapted from a longer form text. The complete paragraph, including the adapted knowledge requirement from the new syllabus, is listed in the bottom panel. All translations are made by the authors and the original text is displayed in square brackets below the translation.[a]



Notes: This figure is adapted from figure 1 of Skolverket (2016). The grades U, G, VG and MVG refer to *icke godkänt*, *godkänt*, *väl godkänt* and *mycket väl godkänt* which correspond to fail, pass, pass with distinction and pass with special distinction, respectively. The grades A-F refer to the letter-based scale, where F is a failing grade.

Figure B1: Comparison of Grading Scales

Table B4: Grading Scale by School Cohort

Cohort:	5th Grade	6th Grade	7th Grade	8th Grade	9th Grade
1991	.	.	.	G/VG/MVG	G/VG/MVG
1992	.	.	.	G/VG/MVG	G/VG/MVG
1993	.	.	.	G/VG/MVG	G/VG/MVG
1994	.	.	.	G/VG/MVG	G/VG/MVG
1995	.	.	.	G/VG/MVG	G/VG/MVG
1996	.	.	.	G/VG/MVG	G/VG/MVG
1997	.	.	.	Letter Grades	Letter Grades
1998	.	.	.	Letter Grades	Letter Grades
1999	.	.	Letter Grades	Letter Grades	Letter Grades
2000	.	Letter Grades	Letter Grades	Letter Grades	Letter Grades

Notes: This table lists the grading scale used to provide feedback to each school cohort from those born in 1991 to those born in 2000 and the grades at which grading feedback was administered.

Appendix C: Variables taken from the Health Behaviour in School-Aged Children Survey

Table 9 uses as outcomes the answers to the following 3 questions:

- "In your opinion, what does your class teacher(s) think about your school performance compared to your classmates?"
  - Answers from: Very good, Good, Average, Below average
- "How do you feel about school at present?"
  - Answers from: Like a lot, Like a little, Don't like much, Don't like at all
- "How pressured do you feel by the schoolwork you have to do?"
  - Answers from: A lot, A little, Some, Not at all

Table 10 uses as outcomes the answers to the following 3 questions:

- In the last 6 months: how often have you had the following. . . ?
  - Irritability or bad temper
  - Feeling nervous
  - Difficulties in getting to sleep
- Answers from: Rarely or never, About every month, About every week, More than once/week, About every day

Appendix D: Variable Classifications

Appendix Table D1: SUN2000

Track:	**Definition:**	**SUN2000 Code:**
Stem	Science	42, 44, 46 and 48
	Engineering	52, 54 and 58
Econ./Soc.Sci./Hum.	Humanities	22
	Social and behavioural science	31
	Business and administration	34
Art	Art	21
Other	General programs	01, 08 and 09
	Teacher training	14
	Journalism and information	32
	Law	38
	Agriculture	62 and 64
	Health and welfare	72 and 76
	Services	81, 84, 85 and 86
	Unspecified	99

Notes: The table lists the *SUN2000* codes used to create the track classifications at the university level. The first two digits of the code was used. *SUN2000* is the Swedish Education Nomenclature defined by Statistics Sweden. *SUN2000* classification is constructed to be comparable with the *ISCD97* classification.

Appendix Table D2: Utbildningsområde

Track:	Definition:	Utbildningsområde:
Stem	Science	NA
	Engineering	TE
Econ./Soc.Sci./Hum.	Humanities	HU
	Social science	SA
Art	Dance	DA
	Design	DE
	Art	KO
	Music	MU
	Opera	OP
	Theater	TA
Other	Pharmaceutical	FA
	Physical education	ID
	Law	JU
	Journalism and information	MM
	Medicine	ME
	Odontological	OD
	Theological	TL
	Teacher training	LU and VFU
	Health and welfare	VÅ
	Other	ÖV

Notes: The table lists the *Utbildningsområde* codes used to create the track classifications in Table A2. *Utbildningsområde* is a classification used by the Swedish government to divide university courses into tracks that form the basis of their funding to universities.

Appendix Table D3: ICD-10

Variable name:	Definition:	ICD-10 Code:
Health	Psychoactive substance use	F10-F19
	Mood disorders	F30-F39
	Neurotic, stress-related and somatoform disorders	F40-F49
	Behavioural syndromes	F50-F59
	Intellectual disability	F70-F79
	Disorders of psychological development	F80-F89
	Behavioural and emotional disorders	F90-F99
	Problems related to education and literacy	Z55
	Intentional self-harm	X60-X84
Self-harm	Intentional self-harm	X60-X84
Drugs	Psychoactive substance use	F10-F19
Anxiety	Mood disorders	F30-F39
	Neurotic, stress-related and somatoform disorders	F40-F49
	Behavioural syndromes	F50-F59
Education	Intellectual disability	F70-F79
	Disorders of psychological development	F80-F89
	Behavioural and emotional disorders	F90-F99
	Problems related to education and literacy	Z55

Notes: The table lists the *ICD-10* codes used to create the mental health groups used in Table 11. *ICD-10* is a classification of diseases and related health problems published by WHO.

Paper IV

The Effect of University Grade Inflation on Graduate Outcomes

Co-authored with Judith Delaney[13] and Therese Nilsson[14]
Accepted for Publication in the *Journal of Applied Econometrics*

Abstract

We exploit a series of reforms inducing grade inflation in English universities, using a staggered difference-in-differences strategy to identify the causal effect of grade inflation on education and labour market outcomes. Policies inducing grade inflation led to an increase in the proportion of students attaining first class honours and a decrease in the proportion obtaining the lowest final grades. We find that grade inflation reduces the likelihood of full-time employment but not the likelihood of being in any employment, while at the same time increasing the likelihood of students pursuing further studies six months after graduating. While we find no average effect on graduate salaries, we find that grade inflation led to a significant increase in the salary of graduates in the bottom decile of the earnings distribution and at the top five percentiles for male graduates. Our findings have particular relevance for policy as they highlight that grade inflation increases participation in further study and improves salary outcomes for the lowest-paid graduates.

Keywords: grade inflation; signalling; human capital
JEL Classifications: I23, I26, J24

[13]University of Bath, University College London, IZA
[14]Lund University, Research Institute of Industrial Economics

1 Introduction

An increasing share of university students are receiving better grades. The proportion of A grades awarded in US universities rose from 33% to 45% between 1988 and 2008 (Rojstaczer and Healy, 2012), while in the UK, the proportion of first class honours degrees awarded rose from 18% to 28% between 2014 and 2018 (HESA, 2018, 2019). Similar trends have been noted across other European countries like Germany (Müller-Benedict and Gaens, 2020) and Italy (Biancardi, 2017). Increasing proportions of good grades and degrees could be the result of better quality or better prepared students being admitted to university, improved teaching, technological advancements leading to improved learning or greater effort by students. An alternative explanation, which has received growing attention and concern among policy makers and the media, is that universities have been grading students more leniently over time, resulting in a trend of grade inflation. In other words, higher grades and good degrees are being awarded without a corresponding improvement in students' abilities or achievements. While the occurrence of grade inflation in universities has been widely documented, we know little about whether grade inflation actually matters for education and labour market outcomes.

Theoretically, universities may want to inflate grades to improve their graduates' job prospects (Popov and Bernhardt, 2013; Yang and Yip, 2003), increase graduates' income (Betts, 1998), induce greater student effort (Boleslavsky and Cotton, 2015; Dubey and Geanakoplos, 2010) or appeal to students who may be attracted by lower grading standards.[1] The expansion of higher education in recent decades may also have led to lower grading standards to cater to lower ability students (Zubrickas, 2015), while non-tenured lecturers may award more favourable grades in an effort to obtain improved teaching evaluations (Greenwald and Gillmore, 1997; Keng, 2018). Theory also provides insights regarding the implications of grade inflation. Grade inflation compresses the grade distribution around higher grades, reducing their signalling quality and, as employers use education as a signal of productivity (Spence, 1973), they may place more weight on the specific university a student attended (Belfield et al., 2018) or devise more costly methods to identify the most productive graduates. Greater uncertainty around productivity may also reduce allocative efficiency in the matching of graduate employees to employers, with employers offering lower entry wages, which increase over time as employers learn about workers' true productivity (Altonji and Pierret, 2001). On the other hand, more lenient

[1] Bar et al. (2012) show theoretically that when students are provided with grading information, they are more likely to select into more lenient courses.

grading may lead to reduced student effort, particularly among higher-ability students (Betts, 1998).

In this paper, we examine the effect of university grade inflation on students' education outcomes and graduates' labour market outcomes. Previous literature has identified how some of the specific mechanisms mentioned above determine outcomes, including signalling, student motivation, educational aspirations and employer hiring practices, among others. Yet it is unclear whether and to what extent the incidence of grade inflation itself affects outcomes. The main contribution of this work is therefore to identify the reduced-form causal effect of grade inflation on education and labour market outcomes.

We exploit a quasi-natural experiment to identify the causal effect of grade inflation. Typically, it is difficult to disentangle whether rising grades are due to better technology, better teaching methods, more diligent students, or grade inflation. We avoid this issue and separate the effects of grade inflation from other factors affecting grades. In England, universities typically award a final grade based on a 4 or 5 point scale. The algorithm used to calculate the final grade involves some weighted combination of individual course grades in the final and penultimate years, a weighted combination of all years, or some other ad hoc awarding scheme with the weighting being rather arbitrary and differing greatly across universities. Universities have full autonomy over how final grades are calculated and many universities have reformed their degree algorithm in recent years, making it easier for students to achieve a better final grade than students with the same profile of course grades who graduated prior to a reform. We specifically use the fact that different universities changed their grading algorithms in different years. In a staggered difference-in-differences framework using the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020), we identify the effect of policies inducing grade inflation on a range of outcomes, including employment status, salary and whether students enrol in further study. We also implement a changes-in-changes estimator to identify the distributional effects of grade inflation on earnings.

We collect data from university administrative records on changes to the algorithms used to assign degree class, matched with survey and register data on graduates from higher education institutions. Specifically, we use the UK's higher education register for all individuals graduating between the academic years 2004/05 and 2012/13, including information on year of entry, field of study, and students' final grade on their degree, in addition to background characteristics of students such as gender and socio-economic status. We complement this register data with information from the *Destinations of Leavers of Higher Education* (DLHE) survey, which is sent to all graduates in England six months after graduation. The DLHE includes information on respondents' primary economic activities and salaries at the time of the survey.

Our results show that policies inducing grade inflation led to an increase in the proportion of students graduating with first class honours and a decrease in the proportion graduating with the lowest degree classes. We find that grade inflation had a significant impact on short-term labour market outcomes and educational choice. Specifically, grade inflation reduces the likelihood of full-time employment but not the likelihood of being in any employment, while the treatment at the same time decreases the likelihood of unemployment. Grade inflation increases the likelihood of students pursuing further studies six months after graduating from their first degree. While there is no significant average effect on graduate salaries, this finding masks significant heterogeneity in effects across the earnings distribution. We find that grade inflation led to a significant increase in the salary of graduates in the bottom decile of the earnings distribution and at the top five percentiles for male graduates. The results also point to important heterogeneous effects, with grade inflation increasing the average salary of graduates in fields of study with higher levels of salary variation, who are more likely to be employed in the private sector. As discussed by Toft Hansen et al. (2021), this is in line with the hypothesis that the labour market structure is an important factor in determining the effect of having an improved signal in the form of better grades.

We examine possible threats to our identification strategy. As we rely on the assumption that grade inflation exogenously affects students, we show that our results are robust to restricting our sample only to students already enrolled in university at the time of grade inflation reforms. We also show that our results are robust to residual dynamic effects from unobserved reforms occurring before our period of analysis, to changes in the composition of universities in our sample, to alternative difference-in-differences estimators and to the potential issues of mis-reporting and censoring in our survey data.

This paper provides three important contributions to the literature. First, despite the substantial rise in university grades and the accompanying public debate around grade inflation and grading standards, the effects of grade inflation in higher education is an empirical question that is not very well examined. Denning et al. (2020) find that rising college graduation rates in the US are partly due to grade inflation, while Babcock (2010) uses survey data from a US university to examine how the perceived leniency of grading at the course-level relates to motivation and finds that higher expected grades are correlated with lower student effort. Ahn et al. (2019) focus on grading differences between STEM and non-STEM classes, suggesting that stricter standards reduce the demand for university courses. With regard to identifying the causal effect of grade inflation, Butcher et al. (2014) examine the effect of stricter grading standards in a US university, finding that students in treated departments gave less favourable teaching evaluations and became less likely to major in stricter graded subjects. We contribute to this literature by examining the role of grading leniency for educational outcomes among the population of university graduates in England and in doing so, provide evidence that grading leniency has a causal impact on further study and salaries in the short term.

Second, we contribute to the literature examining the effect of grade inflation at different stages of education. Several papers examine the effects of high school grade inflation. For example, Betts and Grogger (2003) find that high school grade inflation in the United States led to a decrease in the graduation rate of minorities, but find no effect on the graduation of other students or on college attendance. Hvidman and Sievertsen (2019) find that students in Danish high schools who were downgraded following a national reform performed better on subsequent high school exams and were more likely to attend university. Diamond and Persson (2016) examine the manipulation of high school math grades by teachers in Sweden, finding that students whose grades were inflated completed more years of schooling and those whose grades were inflated from a pass to pass with distinction had higher earnings at age 23. Nordin et al. (2019) examine grade inflation in high schools in Sweden and find that affected males are more likely to achieve higher levels of education and subsequently higher earnings, while there are negative effects on later outcomes for females. The effects of high school grade inflation on labour market outcomes tend to be indirect and operate through the impact of grade inflation on sorting into higher education. We contribute to this literature by shedding light on the direct relationship between university grade inflation and labour market outcomes.

Third, our work contributes to the larger literature on how grading feedback can be leveraged to impact education and labour market outcomes. Bandiera et al. (2015) show that feedback on past achievements can improve student outcomes at university by allowing students to tailor their efforts where needed. Many studies in this literature examine the signalling value of educational attainment rather than that of grades or degree classes. For example, Clark and Martorell (2014), Freier et al. (2015), Khoo and Ost (2018) and Feng and Graetz (2017) study the returns to reaching specific grading thresholds in high-school or in higher education, with significant and positive earnings effects following the latter. While one channel through which grade inflation affects outcomes is via improving the grades or degree class of graduates, grade inflation also works through other channels, including student motivation and employer expectations, as discussed above. Furthermore, Toft Hansen et al. (2021) examine the effect of changes to the signalling value of grades, exploiting a national reform in Denmark. This work is closely related to ours in that the reform studied generated GPA changes unrelated to achievements, although not being reflective of a policy to grade more leniently. Their results show that reform-induced increases in GPA caused large increases in earnings in the short run, but that this positive effect dissipated within three years of graduation, suggesting that employers rapidly learn about employees' true productivity. While our setting does not allow for an examination of long-run labour market outcomes, our paper complements the literature by identifying the reduced form causal effect of university grade inflation on labour market outcomes, giving insights from an alternative source of variation in graduates' signals.

The remainder of the paper proceeds as follows: in the next section we describe the institutional context of the English university grading system. In section 3, we describe the data and in section 4 we outline the empirical strategy. Section 5 provides our main estimates for the effects of grade inflation on graduate outcomes, results on heterogeneous effects, some tests for the robustness of our results and our analysis of distributional effects on earnings. Section 6 concludes.

2 Institutional Context

In England, along with the rest of the United Kingdom, universities typically award a final grade for undergraduate and masters programmes. This final grade, referred to as the degree classification or degree class, is generally assigned on a four or five point scale consisting of First Class Honours (1.1), Upper Second Class Honours (2.1), Lower Second Class Honours (2.2), Third Class Honours (3.1) and Pass, with many universities not offering any distinction between the latter two. Universities maintain autonomy over how degree class is calculated, with substantial variation existing across universities in the methods used to do so. In general, degree programmes in England are strictly defined with regard to the courses students are required to take and the semester and years, or stages, in which they are taken. The algorithm used to assign degree class often then begins with a weighted average of course grades across the different years of a student's studies, with weights assigned according to the credit value of the course and a weighting corresponding to the stage of a degree in which a course was taken. The number of courses or credits contributing to the calculation sometimes varies across stages and can include all or only a number of a student's best grades. For example, the highest graded 80 out of 120 credits may be used at a particular stage. When the weighted average is calculated, degree class is assigned based on whether the weighted average grade is at or above specific thresholds. The algorithm may also or otherwise require a minimum number of courses or credits to be at or above a certain grade. Some universities may also limit students to a maximum number of credits below certain grades. In addition, many universities operate a borderline criteria, where students whose average grade is below a given threshold but within a certain percentage of that threshold may receive the higher degree class based on certain criteria, ranging from the examiners discretion to specific performance requirements in a student's final year.

In recent years, many universities have reformed their degree algorithm, often making it easier for students with the same portfolio of grades to obtain a better final grade. These reforms often involve amendments such as allowing students to remove their lowest course grades in a stage from contributing to the calculation of their final grade or increasing the borderline range for different degree classes. Appendix table B1 provides two examples of grading algorithms before and after reforms took place. In the majority of cases, these reforms affect only the algorithm used to calculate final grades and not the minimum requirements to graduate or to progress between stages during an individual's studies. In addition, some, although not all, of these reforms affect incumbent students in addition to new entrants to a university.

While the algorithm used to assign degree class is public information, it is essentially unknown to students before enrolling in university. Information on the regulations surrounding degree class are typically included in the academic regulations or student handbook provided to students during their induction to university, a point at which students have already enrolled. When applying for a place in university, students most commonly find information on universities and the programmes they offer from the Universities and Colleges Admissions Service (UCAS), from prospectuses provided online by each university and sometimes distributed to secondary schools and from university open days. From searching the UCAS website, the authors did not find any information regarding grading algorithms. Furthermore, in all of the universities from whom we collected information on grading algorithms, none of the universities provided information on grading algorithms in sections devoted to prospective students, but rather under information for current students or in sections devoted to administrative and legal information. Indeed, according to a report on grading algorithms released by UK Standing Committee for Quality Assessment (2020, p. 21), 80% of universities examined do not provide information about the grading algorithm on what they refer to as student-facing web pages. As a result, we do not believe that students engage in sorting based on grading algorithms. As discussed in section 4, we also do not find evidence of increased or decreased enrolment coinciding with reforms which induce grade inflation.

3 Data

Our analysis is based on a comprehensive dataset combining administrative and survey data provided by the UK's Higher Education Statistical Agency (HESA), which is matched to a novel dataset containing information on the algorithms used to assign degree class in English universities.

The university register contains information on individuals who graduated from their first degree in English universities between the 2004/05 and 2012/13 academic years. The data tracks students from the year in which they enrolled in university, including which institution they attended, their field of study, according to 19-digit JACS codes, expected time to completion, date of graduation and degree class. In addition, HESA collects the *Destinations of Leavers of Higher Education Survey* (DLHE), which is administered six months after graduation and is sent to every student graduating from a UK higher education institution and includes information on graduates' main economic activity and yearly salary. Our data include all graduates who completed this survey, which amounts to approximately 77% of all graduates.[2]

To match students in the HESA data to grading algorithms and reforms thereto, we have created a novel dataset detailing the degree algorithms used over time in English universities. For all universities with 5,000 or more graduates in our sample, we have compiled data comprising the degree algorithms used during the period, when algorithms were reformed, whether this led to grade inflation or deflation and whether any or all incumbent students were affected by the reforms. These data have been compiled from a combination of publicly available information on university websites, contact with university registry staff and freedom of information requests sent directly to the universities in question. As not all universities maintain full records of the algorithms used in the past, we have compiled algorithm data for 86 out of 89 universities. We exclude one university which does not award degree class. We exclude universities who instituted a reform in our first year of observation as we do not possess a suitable pre-treatment comparison for these universities. We also exclude the post-reform period for universities who instituted reforms that did not clearly lead to grade inflation. Such reforms include, for example, reforms which served

[2]A response rate of 77% is based on figures released by HESA for the 2015/16 and 2016/17 DLHE surveys.

only to change the weighting allocated to each stage. For universities instituting multiple reforms, we exclude the time period after the second reform.

Table 1 details the number of universities instituting reforms that affected students according to year of enrolment and does so separately for three- and four-year programs. The number of universities switching to treatment reflects the number introducing a reform which led to grade inflation for students enrolling in that year, both in total and among top-40 universities only.[3] The total number of treated universities reflects the number of universities present in the sample who have introduced a reform leading to grade inflation for students entering during that year or before. The number of universities not yet treated includes all universities who have not instituted a reform inducing grade inflation for students enrolling up to that year. There is attrition from both treated and untreated universities due to missing information on grading algorithms, reforms that did not clearly introduce grade inflation and multiple treatments, as discussed above. There is an increasing trend, with a higher number of universities introducing a reform in later years. The majority of reforms are introduced by universities outside the top-40.[4]

Table 1: Reforms instituted over time

	3-year programmes				4-year programmes			
Entry year	Switch to treatment	Switchers from top-40	Total treated	Not yet treated	Switch to treatment	Switchers from top-40	Total treated	Not yet treated
2002/03	0	0	0	66	0	0	0	43
2003/04	1	1	1	66	2	1	2	43
2004/05	2	0	3	66	1	0	3	43
2005/06	2	1	4	60	2	1	5	42
2006/07	4	1	8	55	1	0	5	38
2007/08	1	0	7	56	2	0	6	32
2008/09	5	0	11	45	3	0	9	25
2009/10	3	0	14	38	2	0	10	23
2010/11	1	0	11	34	.	.	.	.

Notes: This table details the number of universities instituting reforms that affected students according to year of enrolment and does so separately for three- and four-year programs. The number of universities switching to treatment reflects the number introducing a reform which led to grade inflation for students enrolling in that year, both in total and among top-40 universities only. The total number of treated universities reflects the number of universities present in the sample who have introduced a reform leading to grade inflation for students entering during that year or before. The number of universities not yet treated includes all universities who have not instituted a reform inducing grade inflation for students enrolling up to that year. There is attrition from both treated and untreated universities. This is because of missing information on grading algorithms in certain years, because we exclude the post-reform period for universities who instituted reforms that did not clearly lead to grade inflation and because, for universities instituting multiple reforms, we exclude the time period after the second reform.

Our data includes all students who graduated from three- or four-year degrees

[3] We define top-40 universities according to The Times University Rankings for 2020.

[4] This introduces the issue that there may be selection into the sample. In section 4 we test for selection and in section 5.3, we show that the results are robust to excluding top-40 universities from both the treatment and control groups.

in English universities having entered between 2002/03 and 2010/11 who completed the DLHE survey six months after graduation and for whose university we possess information on degree class algorithms. This results in a final sample of 709,180 graduates from 77 universities. Due to many students not earning any salary and non-response, we possess salary information for 228,028 graduates.

As degree outcomes, we examine the effects of grade inflation on the degree class obtained by graduates and the likelihood of students finishing their degree on time. As outcomes six months after graduation, we use the likelihood of reporting to be in full-time employment, any employment or unemployment, the likelihood of reporting to be in full-time study or any study 6 months after graduation and log salary, inflated to 2019 British Pounds. Table 2 presents descriptive statistics for those included in our sample. The average entrant is aged 18.45 years old, 57% of entrants are female, and 58% are from a higher socio-economic status, defined by having a parent in a professional occupation.[5] Across the sample, 70% of students achieve either a first class or upper second class honours degree and 92% graduate from their degree on time. 10% of our sample are treated by grade inflation. An average log salary of 9.97 corresponds to 21,375 GBP. Salary information is censored at 10,000 and 100,000 GBP. In section 5.3, we show that grade inflation does not affect the likelihood of reporting a censored value.

4 Empirical Strategy

We use a difference-in-differences approach to identify the causal effect of grade inflation on each of our outcomes of interest. Our framework employs a staggered DiD set-up, exploiting variation in the algorithm used to assign degree class across universities and over time. Typically, studies using a staggered difference-in-differences design have estimated two-way fixed effects models, which in our case would involve estimating the following equation:

$$y_{igt} = \alpha + \beta D_{gt} + \gamma_g + \rho_t + X_i'\lambda + \varepsilon_{igt} \tag{20}$$

where y_{igt} denotes the outcome of interest for individual i who entered university in academic year t, where g represents the group in which students are treated.

[5] We define socio-economic status according to the UK's NRS Social Grades, which are assigned to students according to the social grade of their parents. We define higher socio-economic status as those from the A, B and C1 social grades, which corresponds to higher and intermediate managerial, administrative or professional occupations and supervisory or clerical and junior managerial occupations in administrative or professional occupations.

Table 2: Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.	N
Treatment and background variables					
Female	0.57	0.50	0.0	1.0	709,180
Age on entry	18.45	0.64	17.0	20.0	709,180
A, B and C1 socioeconomic grade	0.58	0.49	0.0	1.0	709,180
Course length	3.20	0.40	3.0	4.0	709,180
Treated	0.10	0.31	0.0	1.0	709,180
Degree outcomes					
First class honours	0.15	0.35	0.0	1.0	709,180
Upper second class honours	0.55	0.50	0.0	1.0	709,180
Lower second class honours	0.26	0.44	0.0	1.0	709,180
Third class honours/pass	0.03	0.18	0.0	1.0	709,180
Graduate on time	0.92	0.27	0.0	1.0	709,180
Outcomes six months after graduation					
Full-time employment	0.53	0.50	0.0	1.0	709,180
Any employment	0.71	0.46	0.0	1.0	709,180
Unemployed	0.08	0.27	0.0	1.0	709,180
Full-time study	0.14	0.35	0.0	1.0	709,180
Any study	0.22	0.41	0.0	1.0	709,180
Log salary	9.97	0.30	9.3	11.8	228,028

Notes: This table presents descriptive statistics for the treatment and background characteristics, degree outcomes and outcomes six months after graduation for our sample.

Since universities offer both 3 and 4 year degree programmes, the level of treatment is at the university-course length level. In some cases, universities use different algorithms in different academic departments and in these cases, we assign treatment at the university-department-course length level (Hereafter, we use university-department-course length and university-course length synonymously). γ_g and ρ_t are university-course length and entry year fixed effects, respectively. D is a binary variable indicating whether the cohort in each group is exposed to an inflationary grading reform and X_i is a vector of controls for pre-determined student characteristics including field of study, age at entry, gender and socio-economic status. ε_{igt} is the error term.

Estimating equation 20 in our case could lead to a biased estimate of the treatment effect, however.[6] An ever-growing literature on difference-in-differences estimation has pointed out that two-way fixed effects is inconsistent in the

[6]In section 5.3, we compare the estimate produced using two-way fixed effects to that of our chosen estimator.

presence of a staggered rollout of the treatment in combination with heterogeneous and/or dynamic treatment effects (Athey and Imbens, 2018; Goodman-Bacon, 2020; De Chaisemartin and d'Haultfoeuille, 2020; de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020; Baker et al., 2021; Borusyak et al., 2021). To identify a consistent estimate of the effect of grade inflation, we employ the difference-in-differences estimator proposed by de Chaisemartin and D'Haultfœuille (2020). Intuitively, this estimator first categorises the treatment group according to the time they are treated, hereafter referred to as 'timing groups'. For each timing group, the treatment effect at the first lead (i.e. the first period after treatment) is estimated by comparing the change in the outcome in the timing group from the last period before treatment to the first period after treatment to the corresponding change in all universities who are still untreated at that time. This is done for all timing groups to calculate the timing group-specific treatment effect at the first lead. A weighted average treatment effect is then calculated across all timing groups, weighted by the number of treated individuals within each group. This is repeated for each lead to identify every lead-specific treatment effect. The average treatment effect is then calculated as a weighted average of lead-specific treatment effects, weighted by the number of treated individuals at each lead.

More formally, as our setting provides a staggered roll-out of the treatment, where once treated, a university-course length group remains treated and as we do not discount the benefits or costs of grade inflation over time, the proposed estimator collapses to the following specification.

$DID_{+,t,l}$ represents the difference-in-differences estimate when comparing the evolution of the outcome of interest, Y, from the last period before treatment, $t-l-1$, to period t for groups, indexed by g, who were treated for the first time in period $t-l$, to all groups who are still untreated at time t. These untreated groups are assumed to represent the counterfactual evolution of Y that would occur in the treatment group in the absence of treatment. This is an unbiased estimator of the effect of the treatment l periods after treatment occurred and is calculated as follows:

$$
\begin{aligned}
DID_{+,t,l} = & \sum_{g:F_{g,1}=t-l} \frac{N_{g,t}}{N^1_{t,l}} (Y_{g,t} - Y_{g,t-l-1} - (X_{g,t} - X_{g,t-l-1})'\hat{\theta}_0) \\
& - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N^{n,t}_{t}} (Y_{g,t} - Y_{g,t-l-1} - (X_{g,t} - X_{g,t-l-1})'\hat{\theta}_0) \qquad (21)
\end{aligned}
$$

The evolution of y is conditional on a vector of pre-determined covariates, X, which include dummies for university-course length groups, gender, field of study, socio-economic status and age on entry to university. $\hat{\theta}_0$ is estimated by regressing $Y_{g,t} - Y_{g,t-l-1}$ on $X_{g,t} - X_{g,t-l-1}$ in the sample of groups untreated at period t. $N^1_{t,l} = \sum_{g:F_{g,1}=t-l} N_{g,t}$ denotes the number of observations in groups, treated for the first time in period $t - l$, while $N^{n,t}_t = \sum_{g:F_{g,1}>t} N_{g,t}$ denotes the number of observations in groups still untreated at period t. $DID_{+,t,l}$ is identical to β in equation 20 if including in the estimation sample only observations from periods $t - l - 1$ and t from groups treated first in period $t - l$ and those still untreated at t in the canonical 2x2 difference-in-differences design.

After calculating $DID_{+,t,l}$ for every l period after treatment for sets of groups treated at each value of t, the next step is to aggregate the estimates into a series of estimates of $DID_{+,l}$. $DID_{+,l}$ is a weighted average of $DID_{+,t,l}$, weighted by the number of observations at period t in groups treated for the first time at $t - l$, aggregated across all groups at every l and is calculated as follows:

$$DID_{+,l} = \sum_{t=l+2}^{NT} N^1_{t,l} DID_{+,t,l} \tag{22}$$

where NT denotes the last period where there is still a group that has been untreated since period 1. This aggregated estimate can be thought of as the treatment effect at each lead in an event study design.

Finally, each $DID_{+,l}$ is aggregated across all values of l to calculate δ^+, the average treatment effect:

$$\delta^+ = \sum_{l=0}^{L_{nt}} w_{+,l} DID_{+,l} \tag{23}$$

where L_{nt} denotes the number of time periods between the earliest t at which a group is first treated and NT, i.e. the largest l at which a $DID_{+,l}$ can be calculated. δ^+ is thus a weighted average of treatment effects at each l, weighted by $w_{+,l}$, which represents the proportion of observations used as weights to calculate each $DID_{+,l}$ and is calculated as follows:

$$w_{+,l} = \frac{\sum_{L_{nt}}^{l=0} N^1_{t,l}}{\sum_{L_{nt}}^{l=0} \sum_{NT}^{t=l+2} N^1_{t,l}} \tag{24}$$

Standard errors are calculated by cluster-bootstrap over the whole procedure, which we perform at the university(-department)-course length level. Clustering at this level is consistent with Abadie et al. (2017) when including cluster fixed effects in the presence of clustering in treatment assignment and heterogeneous treatment effects, which we discuss in section 5.2.[7]

A key reason behind choosing this estimator, as opposed to other proposed difference-in-differences estimators when there exists variation in treatment timing, is that the composition of universities in our sample changes over the time period we examine. This is because some universities implemented multiple reforms. As using the period between a first and second treatment may lead to confounded pre-trends in the presence of dynamic effects, we exclude individuals affected by a second reform. In addition, compositional changes occur due to algorithm information for specific universities not being available for all years in our sample, because some universities did not exist in every year due to mergers and formations and because we exclude reforms which did not clearly induce grade inflation.[8] As this estimator is underpinned by a series of 2x2 difference-in-differences estimates, the treatment effect identified is robust to changes in the composition of the control group over time.

This identification strategy relies on a number of key assumptions to identify a causal effect, namely a sharp and non-pathological design, common trends, exogeneity and no anticipation. A sharp design requires that everyone in the treatment group is in fact affected by the treatment. This holds in our setting as all students are subject to the university regulations. A non-pathological design requires that, for every group that switches from being untreated to treated at a given time t, there exists a group that is untreated in both $t-1$ and t, which holds for each t in our setting. Common trends, which in our setting is conditional on covariates, requires that, in the absence of treatment, outcomes for the treatment and control groups would follow a common trend. In section 5.1, we plot our estimated placebo treatment effects in event study figures, finding from visual inspection that the common trend assumption appears to hold.

Exogeneity requires that students and/or universities do not select into treatment. It is possible that students may select into more lenient universities as information on degree algorithms is publicly available. As discussed in section 2, however, we do not believe that students are aware of the grading algorithm when selecting into specific universities. Furthermore, we test whether our treatment predicts a number of outcomes that may indicate changes in student selection. These are the number of entrants to a university in any given year

[7] Appendix tables A1 and A2 show that the results are robust to clustering at the university level.

[8] Reforms inducing grade deflation typically constitute very small reforms and are therefore not comparable in magnitude to the inflationary reforms we use in our analysis.

(which is based on external data on university-level enrolment from HESA), the number of total graduates, males and females in our sample within a university-course length group and the likelihood of an individual reporting their salary, if working, in the DLHE survey. The results of this analysis are presented in table 3, which shows that grade inflation did not affect any of these measures at conventional levels of statistical significance.

Table 3: Selection into Universities, Survey Responses and Salary Reporting

	(1) New Entrants	(2) Total in Sample	(3) Total Males in Sample	(4) Total Females in Sample	(5) Reports Salary
Grade Inflation	83.340	11.061	–1.772	12.833	–0.061
	(98.87)	(35.84)	(16.82)	(19.99)	(.041)
# of observations	1,425		2,317		1,388,828
# of individuals	543		874		501,282
# of clusters	77		131		131
# of universities	77		77		77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

It is also possible that universities may select into treatment if they were to decide to impose a more lenient grading algorithm as a response to, for example, declining education outcomes. Finally, no anticipation is a common assumption in difference-in-difference designs and requires that the treatment group does not change it's behaviour before the treatment occurs in anticipation of the effects of that treatment. In section 5.1, we show using event study figures, that there does not appear to be non-parallel trends or any anticipation effects.

5 Effects of Grade Inflation

5.1 Main Results

Table 4 presents the main results for the degree class obtained by students and the likelihood of graduating on time. Columns 1-4 show the average effect of grade inflation on the likelihood of graduating with first, upper second, lower second or third class honours, respectively, while column 5 shows the effect on the likelihood of graduating with at least upper second class honours, which is a common requirement to enter masters programmes or to secure employment in many large companies. Column 6 then shows the effect on graduating on time.

Overall, grade inflation leads to a statistically significant increase in first class honours and decreases in lower second and third class honours, with an overall increase in the likelihood of receiving at least upper second class honours. This can be seen as a sort of first-stage result, showing that the reforms we believe are inducing grade inflation do in fact lead to improved final grades. We do not find any significant effect on graduating on time. This is to be expected, given that the reforms we identify typically do not alter the requirements to move from one year to the next or to graduate. This does point, however, to no effect of lower grading standards on students motivations, at least not among those students at the margin of graduating on time.

Table 4: Effects of Grade Inflation on Degree Outcomes

	(1) First Class Honours	(2) Upper Second Class Honours	(3) Lower Second Class Honours	(4) Third Class Honours/Pass	(5) 1st/Upper 2nd Class Honours	(6) Finish on Time
Grade Inflation	0.017	0.013	–0.023	–0.007	0.030	–0.008
	(.006)	(.01)	(.011)	(.003)	(.013)	(.006)
# of observations	1,977,777					
# of individuals	709,180					
# of clusters	131					
# of universities	77					

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Table 5 presents the main results for outcomes measured 6 months after graduating. Columns 1 and 2 show the effect of grade inflation on the likelihood of being in full-time employment and any employment, respectively, and column 3 shows the effect on unemployment. Columns 4 and 5 show the effect on the likelihood of being in full-time or part-time study, with column 6 showing log salary. Grade inflation leads to a 2 percentage point decrease in the likelihood of being in full-time employment but we do not find a significant impact on the likelihood of being in any employment, i.e. being in either full- or part-time employment. This implies that grade inflation leads to a shift from graduates being in full-time to part-time employment. We find that grade inflation leads to a 1.2 percentage point decrease in the likelihood of unemployment six months after graduation, which is accompanied by 1.4 and 2.4 percentage point increases in the likelihood of being in full-time and part-time study, respectively. This could be explained by students who would otherwise be in unemployment choosing instead to study, as individuals who are studying are not considered to be part of the labour force (unless simultaneously working) and thus not considered to be unemployed. Combined with the findings on full-time and part-time employment, this may point toward some graduates who would otherwise be in full-time employment engaging in further studies and working part-time to support such endeavours. Many master programs in England require that students have achieved at least an upper second class honours degree during their bachelors studies. It is therefore likely that as grade inflation increases the likelihood of graduating with such honours, it opens the door to further studies for more graduates. We estimate that grade inflation led to a 0.7% increase in salary but this is very imprecisely estimated and we cannot rule out negative average effects.

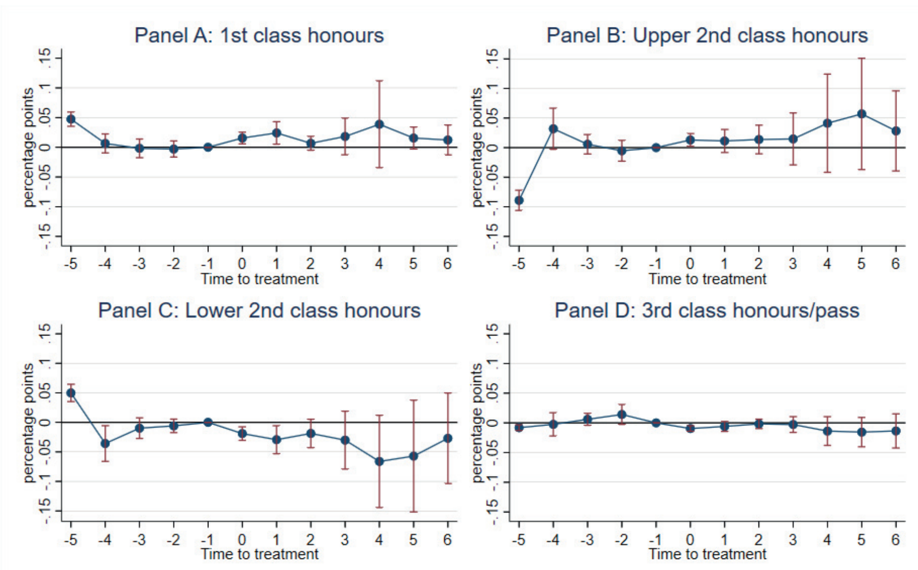
Table 5: Effects of Grade Inflation on Outcomes 6 Months After Graduation

	(1) Full-Time Employment	(2) Any Employment	(3) Unemployed	(4) Full-Time Study	(5) Any Study	(6) Log Salary
Grade Inflation	–0.020	–0.007	–0.012	0.014	0.024	0.007
	(.006)	(.01)	(.011)	(.003)	(.013)	(.006)
# of observations			1,977,777			622,207
# of individuals			709,180			228,028
# of clusters			131			131
# of universities			77			77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

It is unclear exactly what mechanisms dominate in driving these average effects. Theory suggests that grade inflation will improve the signal of students at the margin of two grade levels, improving their labour market prospects (Betts, 1998; Popov and Bernhardt, 2013; Yang and Yip, 2003). Assuming employers are not immediately aware of the reforms, this should not impact on students whose signal is unchanged as a result. If graduates use their improved signal to improve their prospects of engaging in further study, the results we find could be explained by signalling. If this were the only mechanism underlying the results, we could identify the effect of achieving a better degree class on the likelihood of further study, using grade inflation as an instrument for degree class. However, students just below the margin of two grade levels may be more motivated by the possibility of working harder to get a higher degree class (Boleslavsky and Cotton, 2015; Dubey and Geanakoplos, 2010). This motivational effect may also drive some of the effects identified, which would invalidate the exclusion restriction required to use grade inflation as an instrument. Motivation may also fall for high ability students, who no longer have to work so hard to achieve the highest degree class (Betts, 1998), which we discuss in section 5.4.

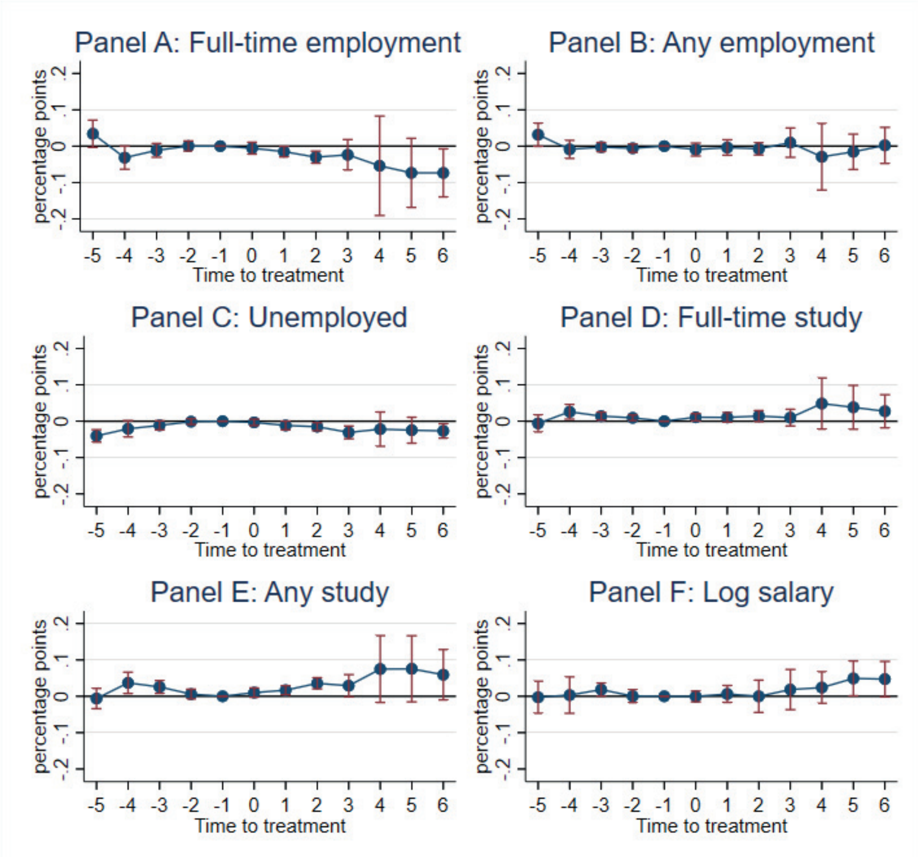
Figure 1 presents event study plots for the likelihood of graduating with each possible degree class and figure 2 presents event study plots for each of our outcomes measured six months after graduation. For each of the degree class outcomes, the effects seem to be relatively constant over time. Looking at panels A, D and E of figure 2, we see that the effects of grade inflation on the likelihood of being employed full-time and of being in full-time or any study appear to be increasing over time. This could signal that as grades become more compressed, students become more likely to engage in further study to improve their labour market signal, moving from full-time to part-time employment to support their studies. Students treated a number of years after the reform also enrolled in university after the reform and thus did not experience the stricter grading scheme. If these students had a higher likelihood of a top grade from the beginning, this could heighten their ambitions and therefore their motivations, making them more likely to engage in further study, and thus reducing employment.



Notes: These figures present event study figures for each degree class outcome and 95% confidence intervals. Confidence intervals are based on standard errors calculated using 100 bootstrap samples, clustered at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 1: Event Study Figures: Degree Outcomes

On the other hand, these results could point toward employer learning, although



Notes: These figures present event study figures for each outcome measured six months after graduation and 95% confidence intervals. Confidence intervals are based on standard errors calculated using 100 bootstrap samples, clustered at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 2: Event Study Figures: Outcomes 6 Months After Graduation

not in the traditional sense of learning about the productivity of individual workers. If universities award top grades to more students, this will lower the average productivity of graduates of each degree class. Although, similar to prospective students, employers do not observe the grading algorithm in place in different universities, the proportion of each degree class awarded by every university in the UK is made available publicly, albeit with a lag of more than one year. If employers expect or observe lower average productivity of graduates from treated universities, they may shift to employing more graduates from other universities. Weaker employment prospects may then push some students to engage in further study, either to further improve their own signal or to avoid unemployment.

Event study plots also allow us to visually assess the assumptions of common trends and no anticipation. For common trends to hold, we should see lags that are small in magnitude and not statistically significantly different to zero. For many of our outcomes, we do see some deviations in trends four to five years before treatment. Moving closer to treatment, however, pre-trends in all of our outcomes for the treatment group appear to converge towards the trends in the control group and from two to three periods before treatment are very similar. We believe that this provides evidence that in the absence of treatment, outcomes for the treatment and control group would follow a common trend. We see that as there are no significant effects in the period before treatment relative to the next previous period, that there does not appear to be any anticipation effect.

5.2 Heterogeneous Effects

In this section, we examine how the effects of grade inflation vary by gender, socio-economic status and field of study according to the level of variation in graduate salaries.

Heterogeneity by socio-economic status may arise due to differences in the resources available to students. Students from professional backgrounds may have access to a wider network of employment opportunities. For these students, the signal provided by grades is therefore less important for securing one's desired employment after graduation, which could lead to weaker effects of grade inflation for these students, relative to those from non-professional backgrounds (Toft Hansen et al., 2021). Students from professional backgrounds may also be in more secure financial positions, allowing them to more easily adjust their study efforts, to engage in a longer search for a better job match or to more easily pursue further study. We split the sample in to professional (A, B and C1) and non-professional (C2, D and E) status based on the socio-economic status of an individual's parents and re-estimate our models. As shown in table 6, the effects of grade inflation are larger for students from a professional background with regard to the likelihood of graduating with first class honours and with regard to engaging in full-time study or any study, although none of these differences are statistically significant at conventional levels. For our other outcomes, the results are largely the same in both groups.

Table 6: Heterogeneity by Socio-Economic Grade

	Education Outcomes						Labour Market Outcomes					
	(1) 1st Class	(2) Upper 2nd	(3) Lower 2nd	(4) 3rd Class	(5) 1st/ Upper 2nd	(6) Finish on Time	(7) Full-Time Employment	(8) Any Employment	(9) Unemployed	(10) Full-Time Study	(11) Any Study	(12) Log Salary
Panel A: C2 and D socioeconomic grades only												
Grade Inflation	0.011	0.021	–0.027	–0.005	0.032	–0.008	–0.016	–0.005	–0.015	0.009	0.019	–0.002
	(.007)	(.012)	(.011)	(.004)	(.013)	(.006)	(.01)	(.009)	(.007)	(.006)	(.009)	(.01)
Sample mean	.134	.536	.289	.04	.671	.917	.52	.714	.084	.135	.206	9.956
# of observations						811,487						253,945
# of individuals						296,723						94,518
# of clusters						131						131
# of universities						77						77
Panel B: A, B and C1 socioeconomic grades only												
Grade Inflation	0.022	0.015	–0.026	–0.010	0.036	–0.004	–0.021	–0.005	–0.010	0.017	0.028	0.010
	(.007)	(.012)	(.014)	(.004)	(.015)	(.007)	(.011)	(.01)	(.004)	(.007)	(.01)	(.011)
Sample mean	.157	.566	.246	.031	.723	.924	.53	.701	.071	.152	.223	9.988
# of observations						1,166,290						368,262
# of individuals						412,457						133,510
# of clusters						131						131
# of universities						77						77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Table 7: Heterogeneity by Salary Variation

	Education Outcomes						Labour Market Outcomes					
	(1) 1st Class	(2) Upper 2nd	(3) Lower 2nd	(4) 3rd Class	(5) 1st/ Upper 2nd	(6) Finish on Time	(7) Full-Time Employment	(8) Any Employment	(9) Unemployed	(10) Full-Time Study	(11) Any Study	(12) Log Salary
Panel A: Low salary variation only												
Grade Inflation	0.021	0.013	–0.024	–0.010	0.034	–0.009	–0.021	–0.008	–0.007	0.014	0.021	–0.009
	(.006)	(.011)	(.011)	(.004)	(.013)	(.005)	(.006)	(.006)	(.004)	(.004)	(.006)	(.009)
Sample mean	.135	.571	.264	.03	.706	.934	.538	.728	.073	.131	.193	9.916
# of observations						924,142						280,576
# of individuals						336,818						104,508
# of clusters						130						130
# of universities						76						76
Panel B: High salary variation only												
Grade Inflation	0.014	0.010	–0.019	–0.006	0.024	–0.004	–0.017	–0.000	–0.016	0.012	0.027	0.025
	(.008)	(.014)	(.015)	(.005)	(.014)	(.01)	(.015)	(.013)	(.007)	(.008)	(.011)	(.011)
Sample mean	.158	.539	.264	.039	.697	.909	.516	.687	.08	.158	.237	10.024
# of observations						1,053,635						341,629
# of individuals						372,362						123,520
# of clusters						130						130
# of universities						77						77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Heterogeneities across field of study by variation in graduate salaries are likely due to the salary structures present in different lines of work. Graduates from degrees in teacher training and graduates from degrees in medicine and related fields are typically employed in the public sector, where starting salaries are likely fixed nationally or according to the location of employment. Grade inflation is unlikely to affect students from such fields with regard to their starting salaries but may do so at other margins. For graduates from fields with less rigid salary structures, grade inflation may be more likely to allow students attain greater starting salaries by improving their labour market signal. To test for this, we split our sample into above and below median levels of salary variation, measured in our sample using untreated graduates. The results of this analysis are presented in table 7.[9] While there is little heterogeneity in the effects of grade inflation on degree outcomes or on the likelihood of employment or further study, there exists a striking difference with regard to log salary. While there is no significant effect of grade inflation on salaries for those in fields of study with low levels of salary variation, there is a positive and significant effect on salaries for those in higher variation fields. Among graduates of such fields of study, grade inflation leads to a 2.5% increase in salary six months after graduation. This difference across groups is significant at the 1% level. This implies that where it is possible to accrue gains with regard to salary, grade inflation does in fact lead to greater income for graduates.

Gender differences in the effects of grade inflation may occur for various reasons. Ahn et al. (2019) find that female students have a greater demand for better grades, while Toft Hansen et al. (2021) and Tan (2020) find that the signalling power of grades in the labour market is stronger for men, which the former attribute to lower wage variation for female graduates. Table 8 shows the results separately for males and females. The effect of grade inflation on degree outcomes appears to be larger for females than males, although these differences are not precisely estimated. Looking at log salary, although the difference is not statistically significant, there is a statistically significant effect of grade inflation on log salary for males of 2.3%. These results are consistent with the the above findings with regard to variation in starting salary, as 65% of male students sort into fields of study classed as possessing high levels of salary variation, compared to 43% of female students.

[9] Low variation fields are as follows: Subjects allied to medicine, Biological sciences, Veterinary science, Architecture, building and planning, Mass communications and documentation, Languages, Creative arts and design, Education and Combined fields of study. High variation field are as follows: Medicine and dentistry, Agriculture and related subjects, Physical sciences, Mathematical sciences, Computer science, Engineering and technology, Social studies, Law, Business and administrative studies and Historical and philosophical studies.

Table 8: Heterogeneity by Gender

	Education Outcomes						Labour Market Outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	1st Class	Upper 2nd	Lower 2nd	3rd Class	1st/ Upper 2nd	Finish on Time	Full-Time Employment	Any Employment	Unemployed	Full-Time Study	Any Study	Log Salary
Panel A: Male students only												
Grade Inflation	0.015	0.003	−0.009	−0.009	0.018	−0.009	−0.016	0.001	−0.016	0.013	0.026	0.023
	(.008)	(.008)	(.009)	(.005)	(.011)	(.007)	(.011)	(.011)	(.007)	(.008)	(.011)	(.013)
Sample mean	.153	.512	.287	.048	.665	.901	.515	.682	.098	.147	.214	10.03
# of observations						863,330						257,066
# of individuals						306,246						93,713
# of clusters						131						131
# of universities						77						77
Panel B: Female students only												
Grade Inflation	0.019	0.019	−0.030	−0.008	0.038	−0.003	−0.021	−0.009	−0.010	0.014	0.023	−0.003
	(.006)	(.014)	(.016)	(.003)	(.017)	(.005)	(.008)	(.007)	(.004)	(.006)	(.007)	(.01)
Sample mean	.143	.586	.246	.025	.729	.936	.535	.726	.06	.143	.217	9.936
# of observations						1,114,447						365,015
# of individuals						402,934						134,315
# of clusters						131						131
# of universities						77						77

Notes: This table presents δ^+, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Table 9: Heterogeneity by University Prestige

	Education Outcomes						Labour Market Outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	1st Class	Upper 2nd	Lower 2nd	3rd Class	1st/ Upper 2nd	Finish on Time	Full-Time Employment	Any Employment	Unemployed	Full-Time Study	Any Study	Log Salary
Panel A: Russell group and top-40 universities only												
Grade Inflation	0.011	−0.001	−0.003	−0.007	0.010	0.002	−0.023	−0.004	−0.030	0.031	0.042	0.017
	(.016)	(.015)	(.015)	(.007)	(.017)	(.015)	(.024)	(.022)	(.013)	(.02)	(.025)	(.026)
Sample mean	.183	.604	.188	.025	.787	.93	.488	.65	.07	.198	.279	10.03
# of observations						422,170						115,980
# of individuals						307,707						92,618
# of clusters						47						47
# of universities						24						24
Panel B: Other universities												
Grade Inflation	0.019	0.021	−0.032	−0.008	0.040	−0.004	−0.018	−0.006	−0.009	0.008	0.016	0.003
	(.008)	(.009)	(.01)	(.005)	(.012)	(.006)	(.008)	(.009)	(.006)	(.005)	(.007)	(.009)
Sample mean	.12	.515	.322	.043	.635	.915	.555	.751	.082	.104	.168	9.936
# of observations						1,079,887						359,752
# of individuals						401,473						135,410
# of clusters						84						84
# of universities						53						53

Notes: This table presents δ^+, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Finally, heterogeneities may exist according to how prestigious a university may be. Students in more prestigious universities might not react as strongly to greater leniency in the degree class algorithm as the signal provided by a more prestigious university might still dominate an improved grade from a less prestigious university. We group together universities in the Russell Group and other universities ranked in the UK's top-40 as the more prestigious universities in our sample.[10] Table 9 shows the results of our analysis, splitting the sample by prestige. Looking at education outcomes, the effects of grade inflation on degree class are much larger in less prestigious universities, implying that the motivational power of more lenient grading is stronger for these students. On the other hand, the effects of grade on outcomes six months after graduation are larger in magnitude for graduates from more prestigious universities, particularly with regard to unemployment and the likelihood of engaging in further study.

5.3 Robustness Checks

A key assumption underpinning our analysis is that of exogeneity. One threat to exogeneity is the ability of students to sort into universities according to their preferred grading algorithm. While we discuss in section 2 that we do not believe this to be a threat in our context, to test that student sorting is not affecting our results, we re-estimate our empirical models, excluding students who enrolled in university after a reform which induced grade inflation. Our treatment group in this analysis therefore comprises only students already enrolled at the time of the reform, ensuring exogeneity. The results of this analysis are presented in table 10 and, as can be seen, the results are almost identical to those presented in tables 4 and 5.

For universities that enacted a reform inducing grade inflation between the 2004/05 and 2012/13 academic years, our treatment variable is equal to zero for those entering university who were not affected by these reforms and one for those who were affected. For universities not enacting any reforms to their grading algorithm, treatment is always equal to zero. Unfortunately, we do not observe reforms prior to this. This means that, as we found evidence of dynamic effects in section 5, some of our treatment and control group may be affected by the residual effects of prior, unobserved reforms. To test that this is not biasing our results, as our data include entrants to university from 2002 to 2010, we re-estimate our models, excluding entrants from, cumulatively, 2002, 2003, 2004 and 2005 in turn. As shown in figures 3 and 4, doing so does not affect our results.

[10]We define top-40 universities according to The Times University Rankings for 2020.

Table 10: Treatment Group Restricted to Incumbent Students

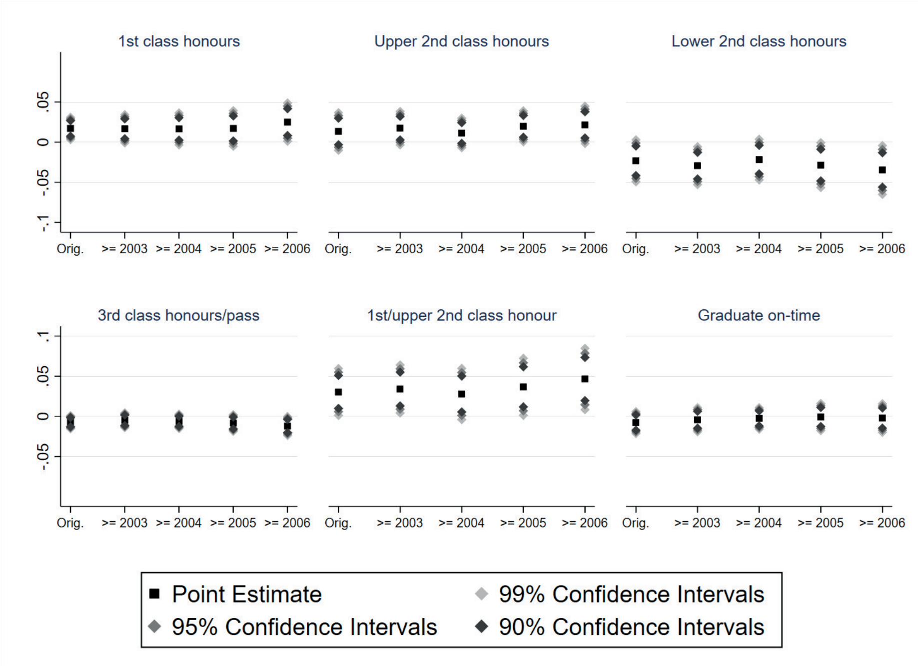
	(1) First Class Honours	(2) Upper Second Class Honours	(3) Lower Second Class Honours	(4) Third Class Honours/Pass	(5) 1st/Upper 2nd Class Honours	(6) Finish on Time
Grade Inflation	0.019	0.013	–0.025	–0.007	0.032	–0.002
	(.007)	(.007)	(.009)	(.004)	(.012)	(.007)
# of observations			1,491,018			
# of individuals			685,836			
# of clusters			131			
# of universities			77			
	Full-Time Employment	Any Employment	Unemployed	Full-Time Study	Any Study	Log Salary
Grade Inflation	–0.012	–0.007	–0.009	0.012	0.016	0.001
	(.007)	(.007)	(.009)	(.004)	(.012)	(.007)
# of observations			1,491,018			470,731
# of individuals			685,836			221,012
# of clusters			131			131
# of universities			77			77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

The composition of our sample with regard to the universities included varies over time. This is because algorithm information for specific universities was not available for all years in our sample, because some universities implemented multiple reforms and we exclude individuals affected by a second reform, because some universities did not exist in every year due to mergers and formations and because universities implemented reforms inducing grade deflation.[11] Although the difference-in-differences estimator proposed by de Chaisemartin and D'Haultfœuille (2020) is robust to compositional changes over time, our estimates may be biased due to other factors, such as if the treatment effect in universities who appear in our data for only a short period is different than to universities appearing in multiple years in our analysis. To show that this is not the case, we re-estimate our models, restricting our sample to universities who appear in at least 3, 4, 5 and 6 years of our data. As shown in figures 5 and 6, such restrictions do not affect our results.

As our analysis relies on data collected from a survey of university graduates, we must assume that the survey is completed truthfully. Although inaccurate responses should serve only to attenuate the estimated relationship between

[11] Reforms inducing grade deflation typically constitute very small reforms and are therefore not comparable in magnitude to the inflationary reforms we use in our analysis.

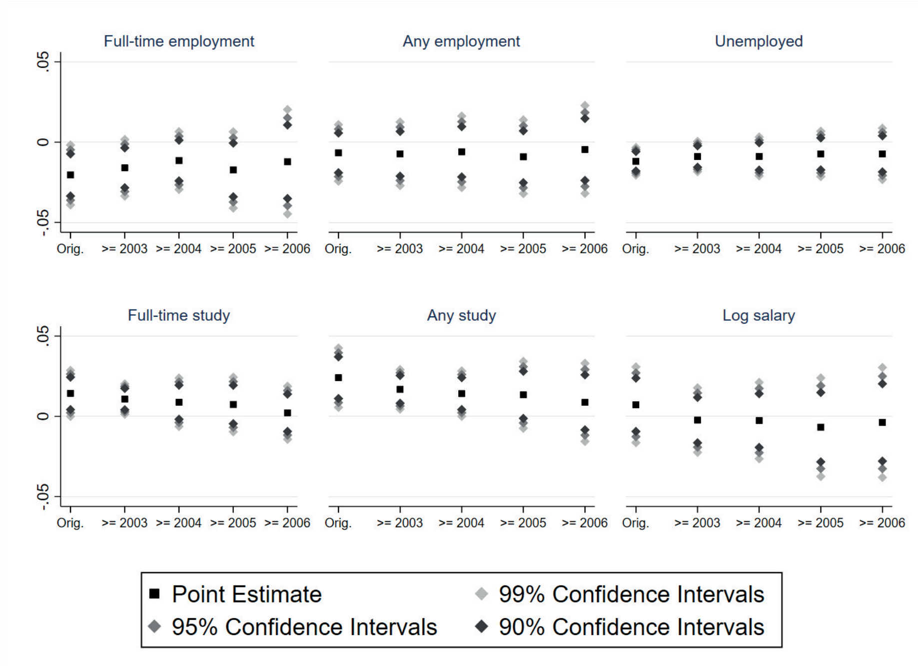


Notes: This table presents δ^+, which is the estimated effect of grade inflation on each outcome, excluding earlier years of enrolment from the sample. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects. These estimates are presented in table format in table A3.

Figure 3: Excluding Early Years from the Sample: Degree Outcomes

grade inflation and outcomes, we test that this is not affecting our results by excluding observations that are likely to be unreliable. A number of graduates report annual salaries of between 70,000 and 100,000 GBP, which are unrealistic salaries for most graduates to earn so soon after graduation. We exclude individuals reporting salaries more than three standard deviations above the mean and re-estimate our models. As shown in table 11, this does not affect our results.

A recent and expanding literature has emerged highlighting the pitfalls of using two-way fixed effects estimators to estimate difference-in-differences models

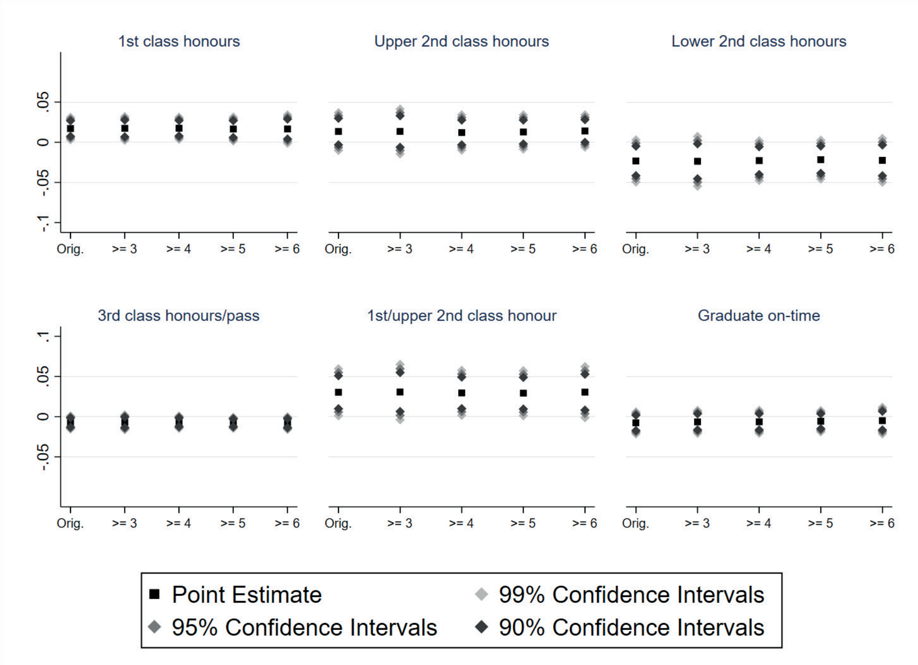


Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome, excluding earlier years of enrolment from the sample. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects. These estimates are presented in table format in table A3.

Figure 4: Excluding Early Years from the Sample: Outcomes 6 Months After Graduation

in the presence of variation in treatment timing and heterogeneous and/or dynamic treatment effects. As such, our main specification employs the estimator proposed by de Chaisemartin and D'Haultfœuille (2020). As various proposed solutions exist, to test whether our results are specific to our chosen estimator, we re-estimate the effect of grade inflation using five alternative estimators. Specifically, we use standard two-way fixed effects and the estimators proposed by Callaway and Sant'Anna (2020), Borusyak et al. (2021), Gardner (2021) and Cengiz et al. (2019).

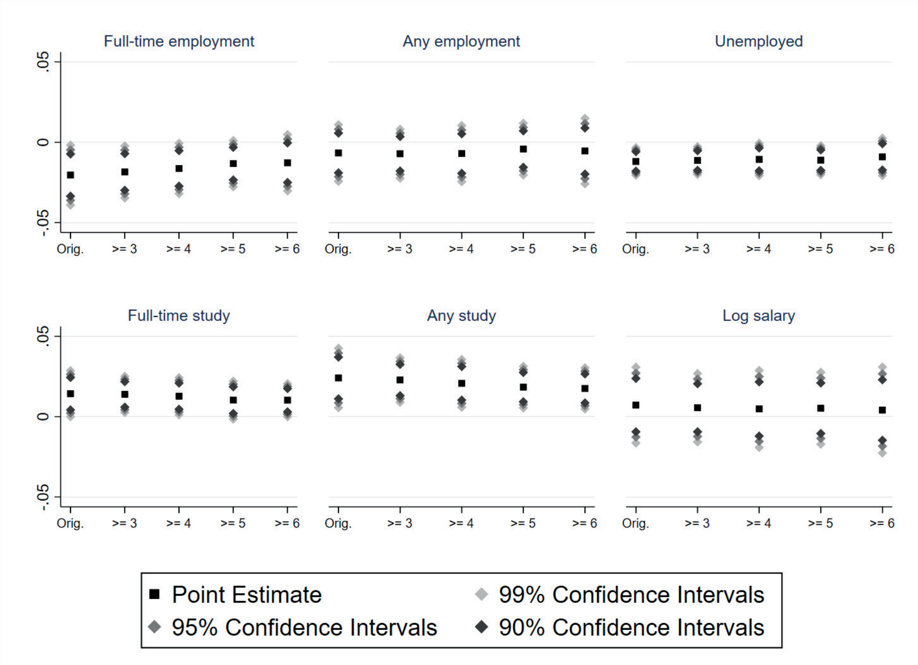
We graphically present the point estimate and confidence intervals from each of



Notes: This figure presents δ^{+}, which is the estimated effect of grade inflation on each outcome, separately for samples excluding universities present in the data for only a certain number of years. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects. These estimates are presented in table format in table A4.

Figure 5: Excluding Universities Appearing in a Small Number of Years: Degree Outcomes

these estimators, in comparison to our own specification, in figures 7 and 8. As can be seen, the point estimates produced by each estimator are broadly similar, albeit with some exceptions. Our main estimates are very similar, although more efficiently estimated, that those of Callaway and Sant'Anna (2020), which is to be expected as these propose very similar approaches, with the latter introducing a doubly-robust estimation procedure for each 2x2 difference-in-differences estimate. While some of the estimates with regard to outcomes six months after graduation are not significant when using two-way fixed effects, estimates when using this approach are likely to be biased. Finally, when looking at both the likelihood of engaging in any study and log salary, there is a deviation in



Notes: This figure presents δ^+, which is the estimated effect of grade inflation on each outcome, separately for samples excluding universities present in the data for only a certain number of years. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects. These estimates are presented in table format in table A4.

Figure 6: Excluding Universities Appearing in a Small Number of Years: Outcomes 6 Months After Graduation

estimates, with those of Borusyak et al. (2021), Gardner (2021) and Cengiz et al. (2019) all estimating negative coefficients for the effect of grade inflation. It is worth noting, however, that in comparison to de Chaisemartin and D'Haultfœuille (2020) and Callaway and Sant'Anna (2020), these approaches are not robust to changes in the composition of the control group over time, which could potentially explain these differences.[12]

[12] Borusyak et al. (2021) propose an extension to their standard estimator, whereby leads are estimated using only balanced controls.

Table 11: Excluding Unreliable Survey Responses

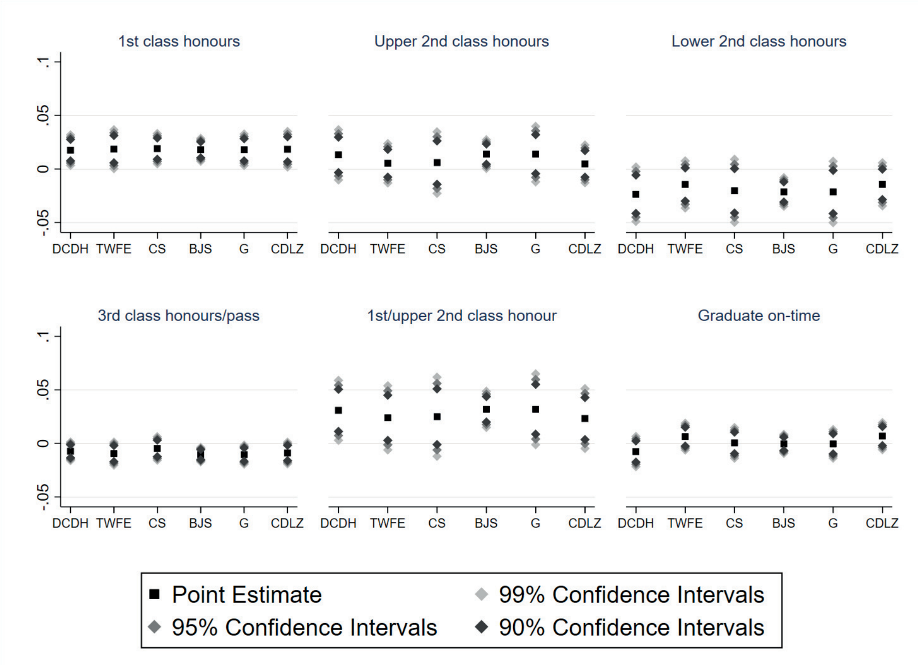
	(1) First Class Honours	(2) Upper Second Class Honours	(3) Lower Second Class Honours	(4) Third Class Honours/Pass	(5) 1st/Upper 2nd Class Honours	(6) Finish on Time
Grade Inflation	0.017	0.013	–0.023	–0.007	0.030	–0.008
	(.006)	(.01)	(.011)	(.003)	(.013)	(.006)
# of observations			1,974,733			
# of individuals			708,181			
# of clusters			131			
# of universities			77			
	Full-Time Employment	Any Employment	Unemployed	Full-Time Study	Any Study	Log Salary
Grade Inflation	–0.020	–0.007	–0.012	0.014	0.024	0.006
	(.006)	(.01)	(.011)	(.003)	(.013)	(.006)
# of observations			1,974,733			619,163
# of individuals			708,181			227,029
# of clusters			131			131
# of universities			77			77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

As salaries are censored at 10,000 and 100,000 GBP, it is possible that our results with regard to salary could be biased. As we found changes in the likelihood of moving from full-time to part-time work, if this resulted in a number of graduates moving from salaries above the lower bound to below, this could positively bias the estimated effect of grade inflation on salary. Likewise, moving from below the upper bound to above it would downward bias any estimated effects. To test that this is not the case, we re-estimate our model, using as outcomes the likelihood of reporting any censored salary, a salary censored at 10,000 and a salary censored at 100,000 GBP. As shown in table 12, grade inflation did not affect the likelihood of reporting a censored salary value.

5.4 Distributional Effects

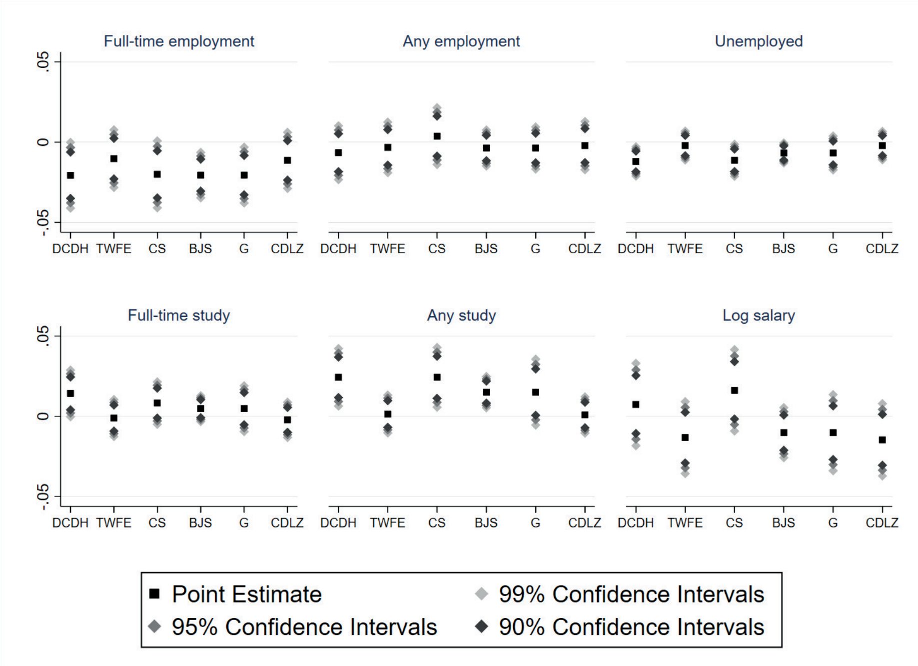
The previous section showed a positive effect of grade inflation on salary for students graduating from fields of study with higher levels of salary variation but no significant average effects for the full sample. Heterogeneous effects of salary may exist across the salary distribution, however. This is particularly likely with grade inflation as the reforms may compress the grade distribution and subsequently lead to very different effects across the earnings distribution.



Notes: These figures present estimates of the effect of grade inflation on degree outcomes and 90%, 95% and 99% confidence intervals. Each point estimate is taken from a separate estimation procedure. DCDH refers to the estimator proposed by de Chaisemartin and D'Haultfœuille (2020), TWFE refers to the standard two-way fixed effects model, CS refers to the estimator proposed by Callaway and Sant'Anna (2020), BJS refers to the estimator proposed by Borusyak et al. (2021), G refers to the estimator proposed by Gardner (2021) and CDLZ refers to the estimator proposed by Cengiz et al. (2019). Confidence intervals are calculated using procedures which are standard to each estimator and based on clustering at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 7: Alternative Difference-in-Differences Estimators: Degree Outcomes

In this section, we investigate the distributional effects of grade inflation on earnings at each percentile of the salary distribution of graduates. To do so, we use the changes-in-changes estimator first proposed by Athey and Imbens (2006). For comparison, a quantile difference-in-differences estimator would estimate a difference-in-differences model, comparing values of the outcome at specific percentiles of the outcome distribution between the treatment and control groups. The changes-in-changes model, on the other hand, estimates the quantile treatment effects as follows.



Notes: These figures present estimates of the effect of grade inflation on outcomes measured six months after graduation and 90%, 95% and 99% confidence intervals. Each point estimate is taken from a separate estimation procedure. DCDH refers to the estimator proposed by de Chaisemartin and D'Haultfœuille (2020), TWFE refers to the standard two-way fixed effects model, CS refers to the estimator proposed by Callaway and Sant'Anna (2020), BJS refers to the estimator proposed by Borusyak et al. (2021), G refers to the estimator proposed by Gardner (2021) and CDLZ refers to the estimator proposed by Cengiz et al. (2019). Confidence intervals are calculated using on procedures which are standard to each estimator and based on clustering at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 8: Alternative Difference-in-Differences Estimators: Outcomes 6 Months After Graduation

For a given quantile q, the econometrician identifies quantile q' in the control group, which corresponds to the value of the outcome at q in the pre-treatment distribution of the treatment group. The change in the value of the outcome at q' from before to after treatment in the control group represents the counterfactual change in the outcome at q in the treatment group. The value of this change is thus added to the value of the outcome at q in the treatment group to identify a counterfactual distribution of the outcome in the treatment group in the absence of treatment. By comparing the realised distribution of the outcome in the

Table 12: Selection into Reporting a Censored Salary Value

	(1) Censored	(2) Censored at 10,000	(3) Censored at 100,000
Grade Inflation	–0.001	–0.001	–0.000
	(.006)	(.006)	(0)
# of observations		622,207	
# of individuals		228,028	
# of clusters		131	
# of universities		77	

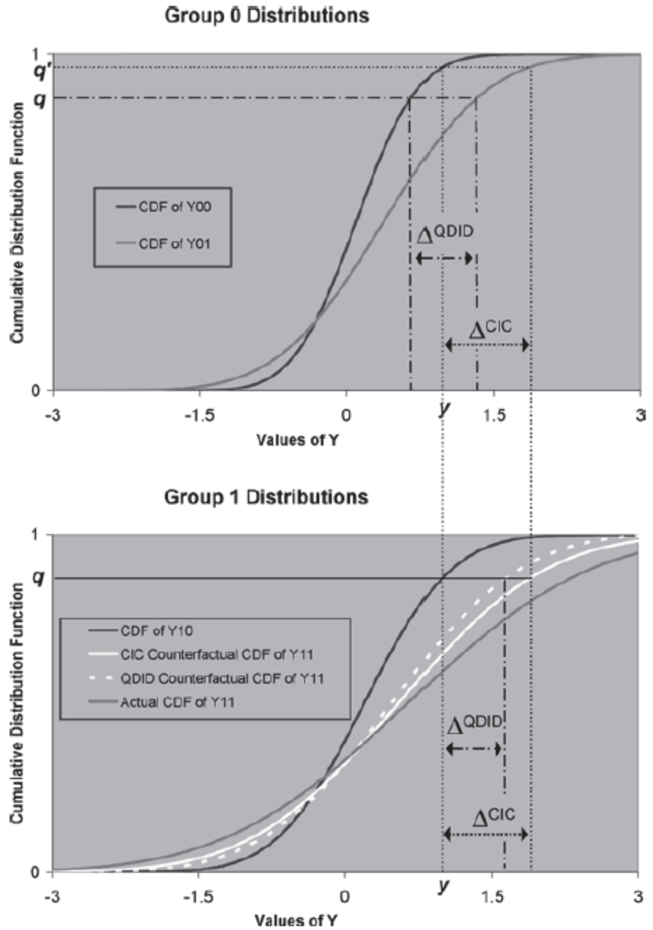
Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

treatment group to the counterfactual distribution, the quantile treatment effect is estimated. This process is outlined in figure 9, which is taken from Athey and Imbens (2006).

We use an adapted CiC estimator proposed by Melly and Santangelo (2015) to estimate the CiC model with covariates. As in our main analysis, we control for university-course length groups, gender, age on entry to university, field of study and parental socio-economic status.[13] Following de Chaisemartin and D'Haultfœuille (2020), we estimate a series of CiC models for each control and treatment year pair in our data. By aggregating these pairwise counterfactual and realised outcome distributions, weighted by the number of observations in each pair, and comparing the aggregated outcome distributions, we identify the average treatment effects at each quantile. To estimate the standard errors of the treatment effects, we bootstrap the whole procedure, using cluster bootstrapping at the university-course length level.

Figure 10 presents the quantile treatment effects of grade inflation on log salary six months after graduation. As can be seen, while we find no statistically significant average effects of grade inflation on earnings, this masks heterogeneity across the distribution, with statistically significant positive effects at the tails of the distribution. At the bottom decile of the salary distribution, grade inflation is found to increase earnings by between 2% and 3%. At the very top percentile, on the other hand, grade inflation leads to an almost 5% increase in salary. Figure 11 presents the corresponding results for male and female graduates in panels A and B, respectively. While similar patterns appear for both males

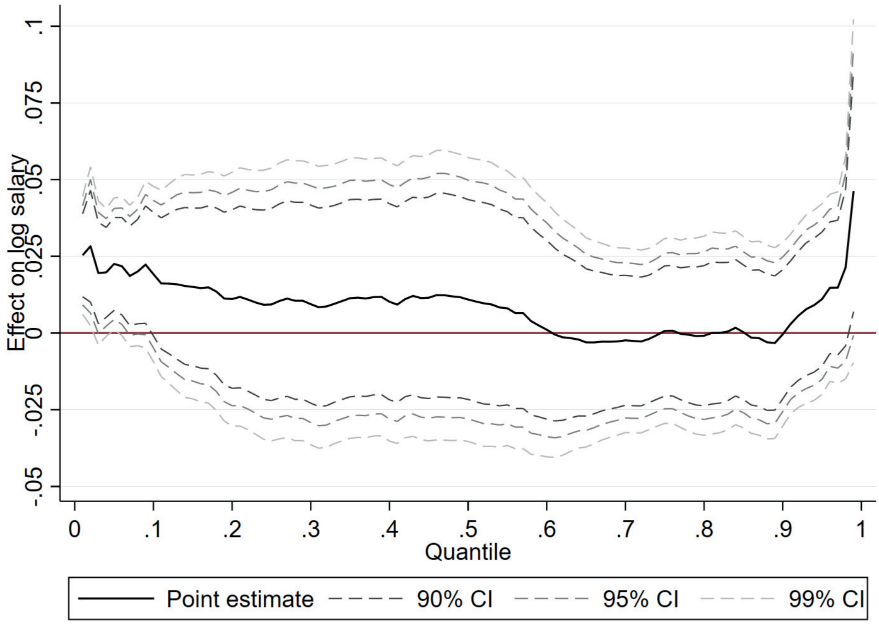
[13]To ensure convergence of the Melly and Santangelo (2015) model, we first residualise the outcome on the latter three controls before estimating quantile treatment effects.



Notes: This figures is taken from Athey and Imbens (2006) and explains how distributional effects are estimated using the changes-in-changes model. For a given quantile q, the econometrician identifies quantile q' in the control group, which corresponds to the value of the outcome at q in the treatment group. The change in the value of the outcome at q' from before to after treatment in the control group represents the counterfactual change in the outcome at q in the treatment group. The value of this change is thus added to the value of the outcome at q in the treatment group to identify a counterfactual distribution of the outcome in the treatment group in the absence of treatment. By comparing the realised distribution of the outcome in the treatment group to the counterfactual distribution, the quantile treatment effect is estimated.

Figure 9: The Changes-in-Changes Model

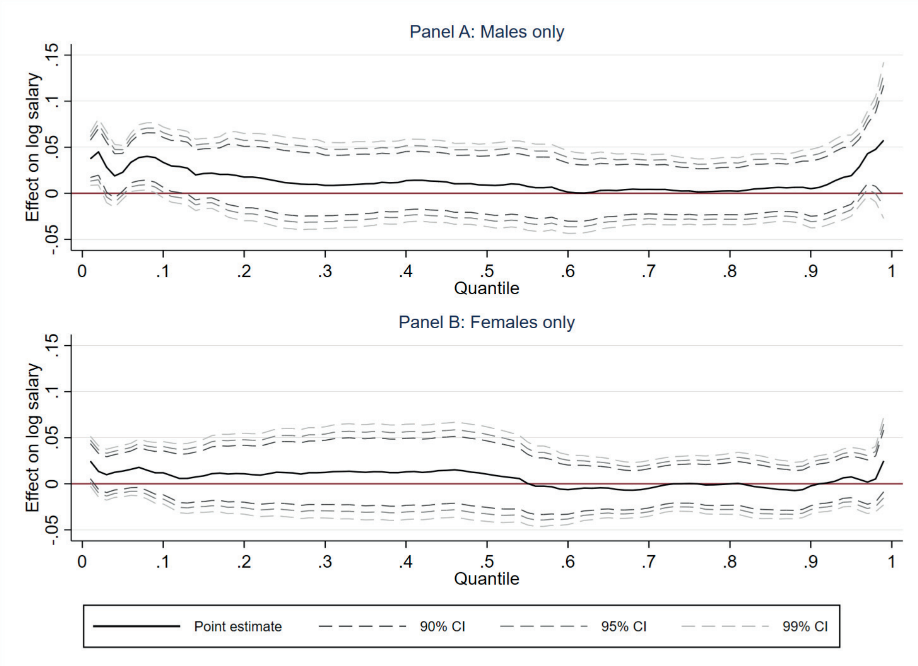
and females, the results presented in figure 10 are largely driven by males, particularly at the top of the distribution.



Notes: This figure plots the effects of grade inflation on log salary at each percentile of the salary distribution six months after graduation and 90%, 95% and 99% confidence intervals. Confidence intervals are based on standard errors calculated using 100 bootstrap samples, clustered at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 10: Distributional Effects of Grade Inflation on Log Salary

In section 5, we find that grade inflation has a positive effect on salary for graduates from fields of study with higher levels of variation in salary. As such, figure 12 presents the quantile treatment effects for graduates from low and high variation fields of study in panels A and B, respectively. It is clear from this figure that the positive effect on salary at the bottom decile of the salary distribution is primarily driven by those in high variation fields. While a similar pattern appears with regard to positive effects at the right tail of the distribution in high variation fields, this pattern is clearly more pronounced for those in low variation fields. This means that in fields of study with low levels of salary variation, grade inflation leads to an increase in variation by increasing the salaries of the top 2-3% of the distribution, while those at the bottom of the

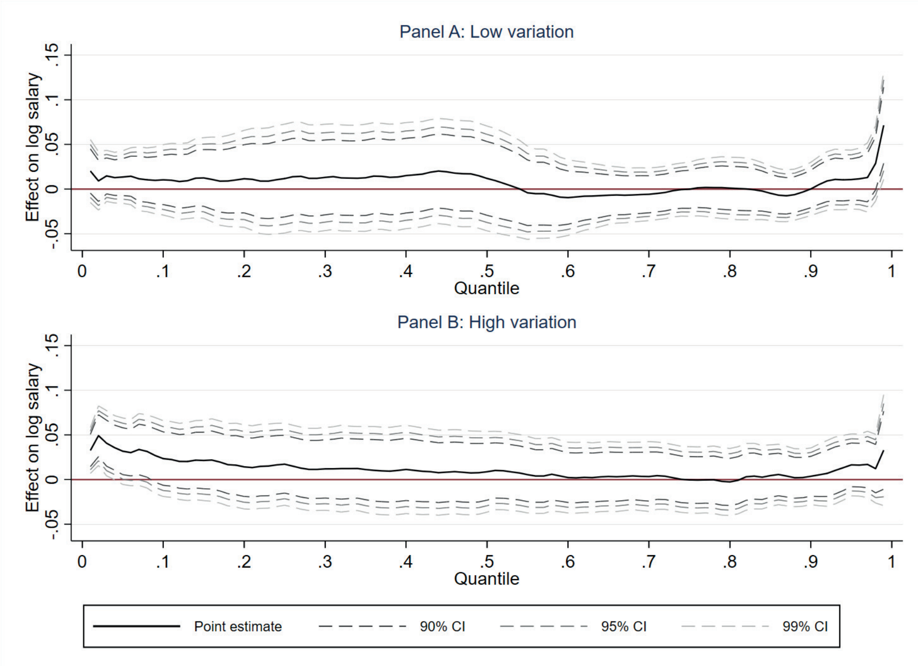


Notes: This figure plots the effects of grade inflation on log salary at each percentile of the salary distribution six months after graduation and 90%, 95% and 99% confidence intervals, separately for males and females. Confidence intervals are based on standard errors calculated using 100 bootstrap samples, clustered at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 11: Distributional Effects of Grade Inflation on Log Salary, by Gender

distribution are largely unaffected.

The positive effect of grade inflation on salary is mainly driven by graduates at the lower end of the salary distribution in fields of study with higher levels of salary variation. This is likely due to students at the lower end of the distribution receiving a better signal, either through the mechanical effect of grade inflation, from achieving higher grades due to the motivational effects of higher grades becoming more attainable, or both. This better signal is then likely to lead to a higher starting salary. Our priors may suggest that students at the bottom of the salary distribution are unlikely to be those students at the margin of receiving a higher degree class as a result of grade inflation. The distribution of starting salaries is very similar for graduates at all degree classes, however. In appendix figure A1, we show that is particularly true across upper second,



Notes: This figure plots the effects of grade inflation on log salary at each percentile of the salary distribution six months after graduation and 90%, 95% and 99% confidence intervals, separately for graduates from fields of study with lower and higher variation in salaries in the absence of treatment. Confidence intervals are based on standard errors calculated using 100 bootstrap samples, clustered at the university(-department)-course length level. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Figure 12: Distributional Effects of Grade Inflation on Log Salary, by Salary Variation

lower second and third class honours, with the main differences being that the distributions at lower classes possess shorter right-tails.

Theory suggests that grade inflation may reduce motivation for high ability students who no longer need to work as hard to achieve the top degree class (Betts, 1998). If this would lead to less human capital production for top students in treated universities and, in turn, lower productivity, this would be reflected in reduced earnings at the right tail of the distribution. We do not find any evidence that this is the case.

It is possible that the distributional effects identified are driven by shifts in the labour supply of graduates. In section 5.1, we find that grade inflation leads to a shift of graduates from full-time employment into part-time employment and further study. This would lead to graduates experiencing a decrease in earnings from moving into part-time employment, moving them down the salary distribution. This would raise the percentile rank of graduates with lower earnings and lead to negative estimated effects on earnings at lower points on the distribution. This would mean that the positive effects we identify are biased toward zero.

The model of employer learning proposed by Altonji and Pierret (2001) provides the implication that with weaker signals of productivity, employers are likely to offer similar wages to all graduates, increasing them over time as the true productivity of workers is revealed. If employers perceive that the signal provided by grades is becoming weaker, graduate wages would then begin to converge toward the median, potentially explaining the increase in wages we find at the bottom of the distribution. This convergence would likely occur across the board, however, with similar increases in wages at the bottom of the distribution in control universities, which would remove any estimated treatment effect of grade inflation. Finally, effects at the bottom and top percentiles of a distributional analysis can sometimes be driven by the presence of outliers in the data. However, as shown in section 5.3, we did not find any evidence that grade inflation affected the proportion of censored salary values, so this is unlikely to be a driver of these effects.

6 Conclusion

Grade inflation at all levels of education has become a growing concern in many countries. Given the key role grading plays in the educational experience of students and in the matching of graduates to employers, evidence of the effects of university grade inflation can have significant policy implications for the higher education sector. In this paper, we provide the first evidence on the effect of university grade inflation on graduation and labour market outcomes.

We employ a novel identification strategy, exploiting university-level reforms to grading algorithms which cause grade inflation in a difference-in-differences framework. We find that grade inflation did lead to an increase in the likelihood of graduating with first class honours and a decrease in the likelihood of graduating with lower second or third class honours. We find that grade inflation reduces the likelihood of full-time employment but not the likelihood of being in any employment, while at the same time decreasing the likelihood of unemployment. Grade inflation is also found to increase the likelihood of students pursuing further studies six months after graduating from their first degree. With regard to earnings, our results show that grade inflation led to increased salaries for students graduating from fields of study with higher levels of salary variation. Finally, we find that grade inflation led to a significant increase in the salary of graduates in the bottom decile of the earnings distribution and at the top five percentiles for male graduates and graduates from fields of study with low levels of variation in salary.

Our findings have various policy implications. First, we show that part of the noted increase in the proportion of students graduating with higher degree classes can be explained by university policies to grade more leniently. Second, our results show that more lenient grading increases both the access to and take-up of further education. Third, as we find effects of grade inflation on employment and salary outcomes, grade inflation may have distortionary effects on the labour market. On the other hand, as we find positive effects of grade inflation on salaries at the bottom of the salary distribution, grade inflation may work to reduce inequalities in earnings. Fourthly, while the grading system described in this paper is specific to the UK and certain other anglophone countries, the use of final grades is prevalent in other settings, such as the use of *cum laude* distinctions in the United States and various European countries. The findings may therefore be informative about the effects of grade inflation internationally. Moreover, the results of this analysis can be generalised to grading at the course-level in settings where students often include their university transcripts as part of graduate job applications and where students' transcripts include a GPA or equivalent aggregate grade.

The results of our analysis identify a reduced form estimate of the effect of grade inflation on education and labour market outcomes. This study is limited, however, by imprecision in our estimates of the dynamic effects of grade inflation across graduating cohorts and by the availability of data on graduate outcomes. Further research on grade inflation can shed light on the mechanisms involved, including but not limited to whether students who benefit from grade inflation become more or less motivated during their studies, whether the increased likelihood of further study is due to the mechanical effect of more students graduating with upper second class honours or due to students' higher perceived aptitudes for study and whether employers hiring practices adapt to grade inflation. In addition, further research can shed light on the longitudinal effects of grade inflation on graduate salaries.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering?

Ahn, T., Arcidiacono, P., Hopson, A., and Thomas, J. R. (2019). Equilibrium grade inflation with implications for female interest in stem majors.

Altonji, J. G. and Pierret, C. R. (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics*.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

Babcock, P. (2010). Real costs of nominal grade inflation? new evidence from student course evaluations. *Economic inquiry*, 48(4):983–996.

Baker, A., Larcker, D. F., and Wang, C. C. (2021). How much should we trust staggered difference-in-differences estimates? *Available at SSRN 3794018*.

Bandiera, O., Larcinese, V., and Rasul, I. (2015). Blissful ignorance? a natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34:13–25.

Bar, T., Kadiyali, V., and Zussman, A. (2012). Putting grades in context. *Journal of Labor Economics*, 30(2):445–478.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Van Der Erve, L., Sibieta, L., Vignoles, A., Walker, I., and Zhu, Y. (2018). The relative labour market returns to different degrees.

Betts, J. R. (1998). The impact of educational standards on the level and distribution of earnings. *The American Economic Review*, 88(1):266–275.

Betts, J. R. and Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22(4):343–352.

Biancardi, D. (2017). Empirical essays in education and labor economics.

Boleslavsky, R. and Cotton, C. (2015). Grading standards and education quality. *American Economic Journal: Microeconomics*, 7(2):248–79.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation.

Butcher, K. F., McEwan, P. J., and Weerapana, A. (2014). The effects of an anti-grade-inflation policy at Wellesley College. *Journal of Economic Perspectives*, 28(3):189–204.

Callaway, B. and Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Clark, D. and Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):282–318.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Difference-in-differences estimators of intertemporal treatment effects.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Denning, J. T., Mumford, K. J., Patterson, R. W., and Warnick, C. M. (2020). Why have college completion rates increased? an analysis of rising grades. Technical report.

Diamond, R. and Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Technical report, National Bureau of Economic Research.

Dubey, P. and Geanakoplos, J. (2010). Grading exams: 100, 99, 98,... or a, b, c? *Games and Economic Behavior*, 69(1):72–94.

Feng, A. and Graetz, G. (2017). A question of degree: the effects of degree class on labor market outcomes. *Economics of Education Review*, 61:140–161.

Freier, R., Schumann, M., and Siedler, T. (2015). The earnings returns to graduating with honors—evidence from law graduates. *Labour Economics*, 34:39–50.

Gardner, J. (2021). Two-stage differences-in-differences.

Goodman-Bacon, A. (2020). Difference-in-differences with variation in treatment timing. Technical report.

Greenwald, A. G. and Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American psychologist*, 52(11):1209.

HESA (2018). Higher education student statistics: UK, 2016/17 - qualifications achieved. Technical report, Higher Education Statistics Agency.

HESA (2019). Higher education student statistics: UK, 2017/18 - qualifications achieved. Technical report, Higher Education Statistics Agency.

Hvidman, U. and Sievertsen, H. H. (2019). High-stakes grades and student behavior. *Journal of Human Resources*, pages 0718–9620R2.

Keng, S.-H. (2018). Tenure system and its impact on grading leniency, teaching effectiveness and student effort. *Empirical Economics*, 55(3):1207–1227.

Khoo, P. and Ost, B. (2018). The effect of graduating with honors on earnings. *Labour Economics*, 55:149–162.

Melly, B. and Santangelo, G. (2015). The changes-in-changes model with covariates. Technical report.

Müller-Benedict, V. and Gaens, T. (2020). A new explanation for grade inflation: the long-term development of german university grades. *European Journal of Higher Education*, 10(2):181–201.

Nordin, M., Heckley, G., and Gerdtham, U. (2019). The impact of grade inflation on higher education enrolment and earnings. *Economics of Education Review*, 73:101936.

Popov, S. V. and Bernhardt, D. (2013). University competition, grading standards, and grade inflation. *Economic inquiry*, 51(3):1764–1778.

Rojstaczer, S. and Healy, C. (2012). Where a is ordinary: The evolution of american college and university grading, 1940-2009. *Teachers College Record*.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374.

Sun, L. and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Tan, B. (2020). Grades as noisy signals.

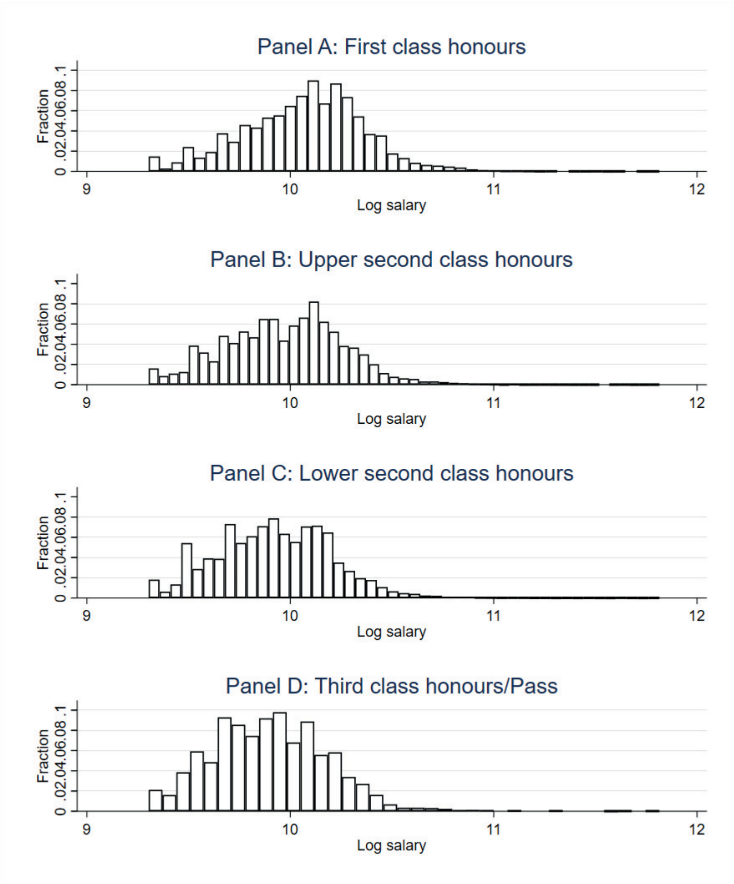
Toft Hansen, A., Hvidman, U., and Sievertsen, H. H. (2021). Grades and employer learning.

UK Standing Committee for Quality Assessment (2020). Degree algorithm practice in 2020. Research report, UK Standing Committee for Quality Assessment.

Yang, H. and Yip, C. S. (2003). An economic theory of grade inflation. *University of Pennsylvania*.

Zubrickas, R. (2015). Optimal grading. *International Economic Review*, 56(3):751–776.

Appendix A: Further Figures and Tables



Notes: This figure plots the distribution of log salary for graduates of each degree class

Figure A1: Salary Distribution, by Degree Class

Table A1: Effects of Grade Inflation on Degree Outcomes, Standard Errors Clustered by University

	(1) First Class Honours	(2) Upper Second Class Honours	(3) Lower Second Class Honours	(4) Third Class Honours/Pass	(5) 1st/Upper 2nd Class Honours	(6) Finish on Time
Grade Inflation	0.017	0.013	–0.023	–0.007	0.030	–0.008
	(.007)	(.013)	(.014)	(.003)	(.016)	(.006)
# of observations	1,977,777					
# of individuals	709,180					
# of clusters	131					
# of universities	77					

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Table A2: Effects of Grade Inflation on Outcomes 6 Months After Graduation, Standard Errors Clustered by University

	(1) Full-Time Employment	(2) Any Employment	(3) Unemployed	(4) Full-Time Study	(5) Any Study	(6) Log Salary
Grade Inflation	–0.020	–0.007	–0.012	0.014	0.024	0.007
	(.007)	(.013)	(.014)	(.003)	(.016)	(.006)
# of observations	1,977,777					622,207
# of individuals	709,180					228,028
# of clusters	131					131
# of universities	77					77

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Table A3: Excluding Early Years from the Sample

	Education Outcomes						Labour Market Outcomes					
	(1) First Class Honours	(2) Upper Second Class Honours	(3) Lower Second Class Honours	(4) Third Class Honours/Pass	(5) 1st/Upper 2nd Class Honours	(6) Finish on Time	(7) Full-Time Employment	(8) Any Employment	(9) Unemployed	(10) Full-Time Study	(11) Any Study	(12) Log Salary
Exc. 2002	0.017	0.017	–0.029	–0.005	0.034	–0.004	–0.016	–0.007	–0.009	0.011	0.017	–0.002
	(.008)	(.009)	(.01)	(.004)	(.013)	(.006)	(.008)	(.008)	(.004)	(.004)	(.005)	(.009)
# of observations						1,502,635						467,827
# of individuals						620,720						200,310
# of clusters						128						128
# of universities						76						76
Exc. 2003 and earlier	0.016	0.011	–0.022	–0.006	0.028	–0.003	–0.011	–0.006	–0.009	0.009	0.014	–0.003
	(.009)	(.008)	(.011)	(.004)	(.014)	(.006)	(.008)	(.01)	(.005)	(.006)	(.006)	(.01)
# of observations						1,117,300						342,829
# of individuals						620,720						200,310
# of clusters						122						122
# of universities						73						73
Exc. 2004 and earlier	0.017	0.020	–0.029	–0.008	0.037	–0.001	–0.017	–0.009	–0.007	0.007	0.013	–0.007
	(.01)	(.008)	(.012)	(.004)	(.015)	(.007)	(.01)	(.01)	(.006)	(.007)	(.009)	(.013)
# of observations						798,012						250,272
# of individuals						435,163						140,342
# of clusters						110						110
# of universities						65						65
Exc. 2005 and earlier	0.025	0.022	–0.035	–0.012	0.047	–0.002	–0.012	–0.005	–0.007	0.002	0.009	–0.004
	(.01)	(.01)	(.013)	(.005)	(.016)	(.008)	(.014)	(.012)	(.007)	(.007)	(.01)	(.015)
# of observations						517,895						175,229
# of individuals						347,525						115,598
# of clusters						99						99
# of universities						58						58

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Table A4: Excluding Universities Appearing in a Small Number of Years

	Education Outcomes						Labour Market Outcomes					
	(1) First Class Honours	(2) Upper Second Class Honours	(3) Lower Second Class Honours	(4) Third Class Honours/Pass	(5) 1st/Upper 2nd Class Honours	(6) Finish on Time	(7) Full-Time Employment	(8) Any Employment	(9) Unemployed	(10) Full-Time Study	(11) Any Study	(12) Log Salary
¿=3	0.017	0.013	−0.024	−0.007	0.031	−0.006	−0.018	−0.007	−0.011	0.014	0.023	0.006
	(.006)	(.012)	(.013)	(.004)	(.015)	(.006)	(.007)	(.007)	(.004)	(.005)	(.006)	(.009)
# of observations						1,901,176						597,715
# of individuals						699,155						224,487
# of clusters						121						121
# of universities						72						72
¿=4	0.017	0.011	−0.021	−0.007	0.028	−0.005	−0.015	−0.006	−0.010	0.011	0.018	0.004
	(.006)	(.009)	(.01)	(.003)	(.012)	(.006)	(.006)	(.007)	(.004)	(.004)	(.006)	(.01)
# of observations						1,618,350						507,391
# of individuals						699,155						224,487
# of clusters						112						112
# of universities						64						64
¿=5	0.016	0.013	−0.022	−0.008	0.029	−0.006	−0.013	−0.004	−0.011	0.010	0.018	0.005
	(.006)	(.009)	(.01)	(.003)	(.012)	(.006)	(.006)	(.007)	(.004)	(.005)	(.006)	(.01)
# of observations						1,583,990						497,388
# of individuals						648,425						208,861
# of clusters						103						103
# of universities						60						60
¿=6	0.016	0.014	−0.023	−0.008	0.031	−0.005	−0.013	−0.005	−0.009	0.010	0.018	0.004
	(.008)	(.009)	(.012)	(.003)	(.014)	(.007)	(.008)	(.009)	(.005)	(.004)	(.005)	(.011)
# of observations						1,299,077						410,838
# of individuals						635,515						204,471
# of clusters						95						95
# of universities						58						58

Notes: This table presents δ^{+}, which is the estimated effect of grade inflation on each outcome. As described in section 4, this is calculated as a weighted average of the treatment effect at each lead, relative to the last period before treatment. Standard errors, in parentheses, are based on 100 bootstrap samples, clustered at the university(-department)-course length level. Point estimates in each column of each panel are taken from separate estimations. Controls include gender, field of study, age on entry and socio-economic grade, in addition to cluster and entry year fixed effects.

Appendix B: Grading Reforms

Table B1: Example Reforms to Grading Algorithms

Pre-Reform Algorithm	Post Reform Algorithm	Students Affected
Example 1		
• Credit-weighted average is calculated for each year • Final weighted average calculates using weights of (0,0.4,0.6) for years 1,2 and 3 • Class assigned as follows: – 1.1: Average of $\geq$ 70% – 2.1: Average of 60% - 69.9% – 2.2: Average of 50% - 59.9% – 3.0: Average of 40% - 49.9% • Borderline zone: Students with an average within 1% of a higher class are graded up if their Y3 average is above 70/60/50%	• Credit-weighted average is calculated for each year • Final weighted average calculates using weights of (0,0.4,0.6) for years 1,2 and 3 • Class assigned as follows: – 1.1: Average of $\geq$ 70% – 2.1: Average of 60% - 69.9% – 2.2: Average of 50% - 59.9% – 3.0: Average of 40% - 49.9% • Borderline zone: Students with an average within 3% of a higher class are graded up if their Y3 average is above 70/60/50%	• All enrolled students
Example 2		
• Students take 8 courses each year • All year 3 grades and a student's best four year 2 grades are considered 'counting grades' – 1.1: 6 counting grades above 70%, max 2 below 60% – 2.1: 6 counting grades above 60%, max 2 below 50% – 2.2: 6 counting grades above 50% – 3.0: Pass all courses	• Students take 8 courses each year • A student's best six year 3 grades and a student's best six grades from year 2 and remaining year 3 courses are considered 'counting grades' – 1.1: 6 counting grades above 70%, max 2 below 60% – 2.1: 6 counting grades above 60%, max 2 below 50% – 2.2: 6 counting grades above 50% – 3.0: Pass all courses	• New entrants only

Notes: This table presents examples of university grading algorithms and reforms that may be enacted. These algorithms are not taken specifically from any one university but rather reflect typical grading algorithms that are in use.

Lund Economic Studies

1. Guy Arvidsson	Bidrag till teorin för verkningarna av räntevariationer, 1962
2. Björn Thalberg	A Trade Cycle Analysis. Extensions of the Goodwin Model, 1966
3. Bengt Höglund	Modell och observationer. En studie av empirisk anknytning och aggregation för en linjär produktionsmodell, 1968
4. Alf Carling	Industrins struktur och konkurrensförhållanden, 1968
5. Tony Hagström	Kreditmarknadens struktur och funktionssätt, 1968
6. Göran Skogh	Straffrätt och samhällsekonomi, 1973
7. Ulf Jakobsson och Göran Norman	Inkomstbeskattningen i den ekonomiska politiken. En kvantitativ analys av systemet för personlig inkomstbeskattning 1952-71, 1974
8. Eskil Wadensjö	Immigration och samhällsekonomi. Immigrationens ekonomiska orsaker och effekter, 1973
9. Rögnvaldur Hannesson	Economics of Fisheries. Some Problems of Efficiency, 1974
10. Charles Stuart	Search and the Organization of Marketplaces, 1975
11. S Enone Metuge	An Input-Output Study of the Structure and Resource Use in the Cameroon Economy, 1976
12. Bengt Jönsson	Cost-Benefit Analysis in Public Health and Medical Care, 1976
13. Agneta Kruse och Ann-Charlotte Ståhlberg	Effekter av ATP - en samhällsekonomisk studie, 1977
14. Krister Hjalte	Sjörestaureringens ekonomi, 1977
15. Lars-Gunnar Svensson	Social Justice and Fair Distributions, 1977
16. Curt Wells	Optimal Fiscal and Monetary Policy - Experiments with an Econometric Model of Sweden, 1978
17. Karl Lidgren	Dryckesförpackningar och miljöpolitik - En studie av styrmedel, 1978
18. Mats Lundahl	Peasants and Poverty. A Study of Haiti, London, 1979
19. Inga Persson-Tanimura	Studier kring arbetsmarknad och information, 1980
20. Bengt Turner	Hyressättning på bostadsmarknaden - Från hyresreglering till bruksvärdesprövning, Stockholm 1979
21. Ingemar Hansson	Market Adjustment and Investment Determination. A Theoretical Analysis of the Firm and the Industry, Stockholm 1981
22. Daniel Boda Ndlela	Dualism in the Rhodesian Colonial Economy, 1981
23. Tom Alberts	Agrarian Reform and Rural Poverty: A Case Study of Peru, 1981
24. Björn Lindgren	Costs of Illness in Sweden 1964-75, 1981

25. Göte Hansson	Social Clauses and International Trade. An Economic Analysis of Labour Standards in Trade Policy, 1981
26. Noman Kanafani	Oil and Development. A Case Study of Iraq, 1982
27. Jan Ekberg	Inkomsteffekter av invandring, 1983
28. Stefan Hedlund	Crisis in Soviet Agriculture?, 1983
29. Ann-Marie Pålsson	Hushållen och kreditpolitiken. En studie av kreditrestriktioners effekt på hushållens konsumtion, sparande och konsumtionsmönster, 1983
30. Lennart Petersson	Svensk utrikeshandel, 1871-1980. En studie i den intraindustriella handelns framväxt, 1984
31. Bengt Assarsson	Inflation and Relative Prices in an Open Economy, 1984
32. Claudio Vedovato	Politics, Foreign Trade and Economic Development in the Dominican Republic, 1985
33. Knut Ödegaard	Cash Crop versus Food Crop Production in Tanzania: An Assessment of the Major Post-Colonial Trends, 1985
34. Vassilios Vlachos	Temporära lönesubventioner. En studie av ett arbetsmarknadspolitiskt medel, 1985
35. Stig Tegle	Part-Time Employment. An Economic Analysis of Weekly Working Hours in Sweden 1963-1982, 1985
36. Peter Stenkula	Tre studier över resursanvändningen i högskolan, 1985
37. Carl Hampus Lyttkens	Swedish Work Environment Policy. An Economic Analysis, 1985
38. Per-Olof Bjuggren	A Transaction Cost Approach to Vertical Integration: The Case of Swedish Pulp and Paper Industry, 1985
39. Jan Petersson	Erik Lindahl och Stockholmsskolans dynamiska metod, 1987
40. Yves Bourdet	International Integration, Market Structure and Prices. A Case Study of the West-European Passenger Car Industry, 1987
41. Krister Andersson and Erik Norrman	Capital Taxation and Neutrality. A study of tax wedges with special reference to Sweden, 1987
42. Tohmas Karlsson	A Macroeconomic Disequilibrium Model. An Econometric Study of the Swedish Business Sector 1970-84, 1987
43. Rosemary Vargas-Lundius	Peasants in Distress. Poverty and Unemployment in the Dominican Republic, 1989
44. Lena Ekelund Axelson	Structural Changes in the Swedish Marketing of Vegetables, 1991
45. Elias Kazarian	Finance and Economic Development: Islamic Banking in Egypt, 1991
46. Anders Danielson	Public Sector Expansion and Economic Development. The Sources and Consequences of Development Finance in Jamaica 1962-84, 1991

47. Johan Torstensson	Factor Endowments, Product Differentiation, and International Trade, 1992
48. Tarmo Haavisto	Money and Economic Activity in Finland, 1866-1985, 1992
49. Ulf Grönkvist	Economic Methodology. Patterns of Reasoning and the Structure of Theories, 1992
50. Evelyne Hangali Maje	Monetization, Financial Development and the Demand for Money, 1992
51. Michael Bergman	Essays on Economic Fluctuations, 1992
52. Flora Mndeme Musonda	Development Strategy and Manufactured Exports in Tanzania, 1992
53. Håkan J. Holm	Complexity in Economic Theory. An Automata Theoretical Approach, 1993
54. Klas Fregert	Wage Contracts, Policy Regimes and Business Cycles. A Contractual History of Sweden 1908-90, 1994
55. Per Frennberg	Essays on Stock Price Behaviour in Sweden, 1994
56. Lisbeth Hellvin	Trade and Specialization in Asia, 1994
57. Sören Höjgård	Long-term Unemployment in a Full Employment Economy, 1994
58. Karolina Ekholm	Multinational Production and Trade in Technological Knowledge, 1995
59. Fredrik Andersson	Essays in the Economics of Asymmetric Information, 1995
60. Rikard Althin	Essays on the Measurement of Producer Performance, 1995
61. Lars Nordén	Empirical Studies of the Market Microstructure on the Swedish Stock Exchange, 1996
62. Kristian Bolin	An Economic Analysis of Marriage and Divorce, 1996
63. Fredrik Sjöholm	R&D, International Spillovers and Productivity Growth, 1997
64. Hossein Asgharian	Essays on Capital Structure, 1997
65. Hans Falck	Aid and Economic Performance - The Case of Tanzania, 1997
66. Bengt Liljas	The Demand for Health and the Contingent Valuation Method, 1997
67. Lars Pålsson Syll	Utility Theory and Structural Analysis, 1997
68. Richard Henricsson	Time Varying Parameters in Exchange Rate Models, 1997
69. Peter Hördahl	Financial Volatility and Time-Varying Risk Premia, 1997
70. Lars Nilsson	Essays on North-South Trade, 1997
71. Fredrik Berggren	Essays on the Demand for Alcohol in Sweden - Review and Applied Demand Studies, 1998
72. Henrik Braconier	Essays on R&D, Technology and Growth, 1998
73. Jerker Lundbäck	Essays on Trade, Growth and Exchange Rates, 1998
74. Dan Anderberg	Essays on Pensions and Information, 1998

75. P. Göran T. Hägg	An Institutional Analysis of Insurance Regulation – The Case of Sweden, 1998
76. Hans-Peter Bermin	Essays on Lookback and Barrier Options - A Malliavin Calculus Approach, 1998
77. Kristian Nilsson	Essays on Exchange Rates, Exports and Growth in Developing Countries, 1998
78. Peter Jochumzen	Essays on Econometric Theory, 1998
79. Lars Behrenz	Essays on the Employment Service and Employers' Recruitment Behaviour, 1998
80. Paul Nystedt	Economic Aspects of Ageing, 1998
81. Rasha M. Torstensson	Empirical Studies in Trade, Integration and Growth, 1999
82. Mattias Ganslandt	Games and Markets - Essays on Communication, Coordination and Multi-Market Competition, 1999
83. Carl-Johan Belfrage	Essays on Interest Groups and Trade Policy, 1999
84. Dan-Olof Rooth	Refugee Immigrants in Sweden - Educational Investments and Labour Market Integration, 1999
85. Karin Olofsdotter	Market Structure and Integration: Essays on Trade, Specialisation and Foreign Direct Investment, 1999
86. Katarina Steen Carlsson	Equality of Access in Health Care, 1999
87. Peter Martinsson	Stated preference methods and empirical analyses of equity in health, 2000
88. Klas Bergenheim	Essays on Pharmaceutical R&D, 2000
89. Hanna Norberg	Empirical Essays on Regional Specialization and Trade in Sweden, 2000
90. Åsa Hansson	Limits of Tax Policy, 2000
91. Hans Byström	Essays on Financial Markets, 2000
92. Henrik Amilon	Essays on Financial Models, 2000
93. Mattias Lundbäck	Asymmetric Information and The Production of Health, 2000
94. Jesper Hansson	Macroeconometric Studies of Private Consumption, Government Debt and Real Exchange Rates, 2001
95. Jonas Månsson	Essays on: Application of Cross Sectional Efficiency Analysis, 2001
96. Mattias Persson	Portfolio Selection and the Analysis of Risk and Time Diversification, 2001
97. Pontus Hansson	Economic Growth and Fiscal Policy, 2002
98. Joakim Gullstrand	Splitting and Measuring Intra-Industry Trade, 2002
99. Birger Nilsson	International Asset Pricing, Diversification and Links between National Stock Markets, 2002
100. Andreas Graflund	Financial Applications of Markov Chain Monte Carlo Methods, 2002

101. Therése Hindman Persson	Economic Analyses of Drinking Water and Sanitation in Developing Countries, 2002
102. Göran Hjelm	Macroeconomic Studies on Fiscal Policy and Real Exchange Rates, 2002
103. Klas Rikner	Sickness Insurance: Design and Behavior, 2002
104. Thomas Ericson	Essays on the Acquisition of Skills in Teams, 2002
105. Thomas Elger	Empirical Studies on the Demand for Monetary Services in the UK, 2002
106. Helena Johansson	International Competition, Productivity and Regional Spillovers, 2003
107. Fredrik Gallo	Explorations in the New Economic Geography, 2003
108. Susanna Thede	Essays on Endogenous Trade Policies, 2003
109. Fredrik CA Andersson	Interest Groups and Government Policy, A Political Economy Analysis, 2003
110. Petter Lundborg	Risky Health Behaviour among Adolescents, 2003
111. Martin W Johansson	Essays on Empirical Macroeconomics, 2003
112. Joakim Ekstrand	Currency Markets - Equilibrium and Expectations, 2003
113. Ingemar Bengtsson	Central bank power: a matter of coordination rather than money supply, 2003
114. Lars Pira	Staples, Institutions and Growth: Competitiveness of Guatemalan Exports 1524-1945, 2003
115. Andreas Bergh	Distributive Justice and the Welfare State, 2003
116. Staffan Waldo	Efficiency in Education - A Multilevel Analysis, 2003
117. Mikael Stenkula	Essays on Network Effects and Money, 2004
118. Catharina Hjortsberg	Health care utilisation in a developing country -the case of Zambia, 2004
119. Henrik Degrér	Empirical Essays on Financial Economics, 2004
120. Mårten Wallette	Temporary Jobs in Sweden: Incidence, Exit, and On-the-Job Training, 2004
121. Tommy Andersson	Essays on Nonlinear Pricing and Welfare, 2004
122. Kristian Sundström	Moral Hazard and Insurance: Optimality, Risk and Preferences, 2004
123. Pär Torstensson	Essays on Bargaining and Social Choice, 2004
124. Frederik Lundtofte	Essays on Incomplete Information in Financial Markets, 2005
125. Kristian Jönsson	Essays on Fiscal Policy, Private Consumption and Non-Stationary Panel Data, 2005
126. Henrik Andersson	Willingness to Pay for a Reduction in Road Mortality Risk: Evidence from Sweden, 2005

127. Björn Ekman	Essays on International Health Economics: The Role of Health Insurance in Health Care Financing in Low- and Middle-Income Countries, 2005
128. Ulf G Erlandsson	Markov Regime Switching in Economic Time Series, 2005
129. Joakim Westerlund	Essays on Panel Cointegration, 2005
130. Lena Hiselius	External costs of transports imposed on neighbours and fellow road users, 2005
131. Ludvig Söderling	Essays on African Growth, Productivity, and Trade, 2005
132. Åsa Eriksson	Testing and Applying Cointegration Analysis in Macroeconomics, 2005
133. Fredrik Hansen	Explorations in Behavioral Economics: Realism, Ontology and Experiments, 2006
134. Fadi Zaher	Evaluating Asset-Pricing Models in International Financial Markets, 2006
135. Christoffer Bengtsson	Applications of Bayesian Econometrics to Financial Economics, 2006
136. Alfredo Schclarek Curutchet	Essays on Fiscal Policy, Public Debt and Financial Development, 2006
137. Fredrik Wilhelmsson	Trade, Competition and Productivity, 2006
138. Ola Jönsson	Option Pricing and Bayesian Learning, 2007
139. Ola Larsson	Essays on Risk in International Financial Markets, 2007
140. Anna Meyer	Studies on the Swedish Parental Insurance, 2007
141. Martin Nordin	Studies in Human Capital, Ability and Migration, 2007
142. Bolor Naranhuu	Studies on Poverty in Mongolia, 2007
143. Margareta Ekbladh	Essays on Sickness Insurance, Absence Certification and Social Norms, 2007
144. Erik Wengström	Communication in Games and Decision Making under Risk, 2007
145. Robin Rander	Essays on Auctions, 2008
146. Ola Andersson	Bargaining and Communication in Games, 2008
147. Marcus Larson	Essays on Realized Volatility and Jumps, 2008
148. Per Hjertstrand	Testing for Rationality, Separability and Efficiency, 2008
149. Fredrik NG Andersson	Wavelet Analysis of Economic Time Series, 2008
150. Sonnie Karlsson	Empirical studies of financial asset returns, 2009
151. Maria Persson	From Trade Preferences to Trade Facilitation, 2009
152. Eric Rehn	Social Insurance, Organization and Hospital Care, 2009
153. Peter Karpestam	Economics of Migration, 2009
154. Marcus Nossman	Essays on Stochastic Volatility, 2009
155. Erik Jonasson	Labor Markets in Transformation: Case Studies of Latin America, 2009

156. Karl Larsson	Analytical Approximation of Contingent Claims, 2009
157. Therese Nilsson	Inequality, Globalization and Health, 2009
158. Rikard Green	Essays on Financial Risks and Derivatives with Applications to Electricity Markets and Credit Markets, 2009
159. Christian Jörgensen	Deepening Integration in the Food Industry – Prices, Productivity and Export, 2010
160. Wolfgang Hess	The Analysis of Duration and Panel Data in Economics, 2010
161. Pernilla Johansson	From debt crisis to debt relief: A study of debt determinants, aid composition and debt relief effectiveness, 2010
162. Nils Janlöv	Measuring Efficiency in the Swedish Health Care Sector, 2010
163. Ai Jun Hou	Essays on Financial Markets Volatility, 2011
164. Alexander Reffgen	Essays on Strategy-proof Social Choice, 2011
165. Johan Blomquist	Testing homogeneity and unit root restrictions in panels, 2012
166. Karin Bergman	The Organization of R&D - Sourcing Strategy, Financing and Relation to Trade, 2012
167. Lu Liu	Essays on Financial Market Interdependence, 2012
168. Bujar Huskaj	Essays on VIX Futures and Options, 2012
169. Åsa Ljungvall	Economic perspectives on the obesity epidemic, 2012
170. Emma Svensson	Experimenting with Focal Points and Monetary Policy, 2012
171. Jens Dietrichson	Designing Public Organizations and Institutions: Essays on Coordination and Incentives, 2013
172. Thomas Eriksson	Empirical Essays on Health and Human Capital, 2013
173. Lina Maria Ellegård	Political Conflicts over Public Policy in Local Governments, 2013
174. Andreas Hatzigeorgiou	Information, Networks and Trust in the Global Economy - Essays on International Trade and Migration, 2013
175. Gustav Kjellsson	Inequality, Health, and Smoking, 2014
176. Richard Desjardins	Rewards to skill supply, skill demand and skill match-mismatch: Studies using the Adult Literacy and Lifeskills survey, 2014
177. Viroj Jienwatcharamongkhol	What Drives Exports? Empirical Evidence at the Firm Level, 2014
178. Anton Nilsson	Health, Skills and Labor Market Success, 2014
179. Albin Erlanson	Essays on Mechanism Design, 2014
180. Daniel Ekeblom	Essays in Empirical Expectations, 2014

181. Sofie Gustafsson	Essays on Human Capital Investments: Pharmaceuticals and Education, 2014
182. Katarzyna Burzynska	Essays on Finance, Networks and Institutions, 2015
183. Mingfa Ding	Corporate Ownership and Liquidity in China's Stock Markets, 2015
184. Anna Andersson	Vertical Trade, 2015
185. Cecilia Hammarlund	Fish and Trips in the Baltic Sea - Prices, Management and Labor Supply, 2015
186. Hilda Ralsmark	Family, Friend, or Foe? Essays in Empirical Microeconomics, 2015
187. Jens Gudmundsson	Making Pairs, 2015
188. Emanuel Alfranseder	Essays on Financial Risks and the Subprime Crisis, 2015
189. Ida Lovén	Education, Health, and Earnings – Type 1 Diabetes in Children and Young Adults, 2015
190. Caren Yinxia Nielsen	Essays on Credit Risk, 2015
191. Usman Khalid	Essays on Institutions and Institutional change, 2016
192. Ross Wilson	Essays in Empirical Institutional Analysis, 2016
193. Milda Norkute	A Factor Analytical Approach to Dynamic Panel Data Models, 2016
194. Valeriia Dzhamalova	Essays on Firms' Financing and Investment Decisions, 2016
195. Claes Ek	Behavioral Spillovers across Prosocial Alternatives, 2016
196. Graeme Cokayne	Networks, Information and Economic Volatility, 2016
197. Björn Thor Arnarson	Exports and Externalities, 2016
198. Veronika Lunina	Multivariate Modelling of Energy Markets, 2017
199. Patrik Karlsson	Essays in Quantitative Finance, 2017
200. Hassan Sabzevari	Essays on systemic risk in European banking, 2017
201. Margaret Samahita	Self-Image and Economic Behavior, 2017
202. Aron Berg	Essays on informational asymmetries in mergers and acquisitions, 2017
203. Simon Reese	Estimation and Testing in Panel Data with Cross-Section Dependence, 2017
204. Karl McShane	Essays on Social Norms and Economic Change, 2017
205. Elvira Andersson	From Cradle to Grave: Empirical Essays on Health and Economic Outcomes, 2017
206. Yana Pryymachenko	Heavy Metal Exposure in Early Life - Health and Labour Market Perspectives, 2017
207. Alemu Tulu Chala	Essays on Banking and Corporate Finance, 2017
208. Jim Ingebretsen Carlson	Essays on economic behavior, focusing and auctions, 2018
209. Jörgen Kratz	Essays on Matching, 2018

210. Anna Welander Tärneberg	Essays on Health in Developing Countries, 2018
211. Osmis Areda Habte	Essays on competition and consumer choice, 2018
212. Thomas Hofmarcher	Essays in Empirical Labor Economics, 2019
213. Hjördis Hardardottir	Time and inequality – A study of individual preferences, 2019
214. Erik Grenestam	Essays in Applied Microeconomics, 2019
215. Sara Moricz	Institutions, Inequality and Societal Transformations, 2019
216. John Källström	Mobility in Science, 2019
217. Mehmet Caglar Kaya	Essays on Corporate Growth and Corporate Credit Risk, 2020
218. Dinh-Vinh Vo	Essays on risk spillover and information transmission in the financial markets, 2020
219. Kristoffer Persson	Essays on Expectations - Information, Formation and Outcomes, 2020
220. Polina Knutsson	Empirical Studies on Firm and Labor Market Dynamics, 2020
221. Sanna Ericsson	Reaching For Equality: Essays in Education and Gender Economics, 2020
222. Yana Petrova	Essays on Panel Data with Multidimensional Unobserved Heterogeneity, 2020
223. Pol Campos-Mercade	Incentives in Education and Moral Behavior in Groups, 2020
224. Staffan Lindén	Essays on expectations and financial markets, 2020
225. Dominika Krygier	Essays on systemic risk and financial market volatility, 2021
226. Sara Mikkelsen	Family matters: Essays in Applied Microeconomics, 2021
227. Hampus Poppius	Quantitative Studies on Pricing and Consumer Behavior, 2021
228. Danial Ali Akbari	Das Human-Kapital: Emerging Patterns in the Class Structure, 2021
229. Matthew Collins	Essays on instruction time, grades and parental investments in education, 2022

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

Lund University
Department of Economics
ISBN 978-91-8039-150-4
ISSN 0460-0029