2016-21
Anne Brink Nandrup
PhD Dissertation

Determinants of Student Achievement and Education Choice

DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS BSS • AARHUS UNIVERSITY • DENMARK

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Determinants of Student Achievement and Education Choice

PhD dissertation

Anne Brink Nandrup

Aarhus BSS, Aarhus University
Department of Economics and Business Economics

2016

PREFACE

This dissertation is the tangible result of my PhD studies at the Department of Economics and Business Economics at Aarhus University. I am grateful to the department for providing generous financial support and a truly excellent research environment. Further financial support from the Centre for Strategic Education Research, Oticon Fonden, Knud Højgaards Fond and Trygfonden's Centre for Child Reseach is gratefully acknowledged.

First and foremost, I wish to express my sincere gratitude to my supervisors Nina Smith, Maria Humlum and Marianne Simonsen. You have provided guidance and support when needed and pushed me futher than I was willing to go on my own. Thank you for all the time and effort you have put into creating an inspiring and educative atomosphere these past three years. Special thanks are also extended to Simon Calmar Andersen for a rewarding collaboration in writing the second chapter of this dissertation from which I have benefitted immensely, and to Susan Stilling for absolutely extraordinary support.

From September 2014 to January 2015, I had the privilege of visiting the Department of Economics at Princeton University, New Jersey. I am grateful to Bo Honoré, for hosting me and introducing me to the research environment as well as Princeton in general. Special thanks to Emil and Michala who made my stay truly pleasurable.

Wonderful colleagues make up for half the hardship when writing a dissertation. I am grateful to all of you for refreshing discussions and witty banter during countless lunches and coffee breaks. Morten, Louise, Bastien, Tine, Kenneth, Janne, Peter, Rune, Jonas, Palle, Simon, Jonas and Niels, thank you for insightful comments and memorable experiences. And finally, to Sanni. I hope you know the remarkable difference you have made. Everyone should have a cheerleader like you.

To my family and friends, I am horrible at asking for help when I am in over my head, but you have helped me stay upright. Thank you for your support and for laughing as well as bearing with me. To Tim, for always believing in me when I fail to do so myself and for giving me the space to do what I need. Your patience with me seems endless at times. Thank you.

Anne Brink Nandrup
Aarhus, August 2016

Updated preface

The predefence took place in Aarhus on October 11, 2016. The assessment committee consists of Peter Fredriksson, Stockholm University, Marc Gurgand, CNRS and Paris School of Economics, and Helena Skyt Nielsen, Aarhus University. I am grateful to the members of the assessment committee for their careful reading of this dissertation and their constructive comments and insightful discussion of my work. Many suggestions have been incorporated in the current version of the dissertation while others remain for future research.

Anne Brink Nandrup
Aarhus, November 2016

CONTENTS

Summary

Large discrepancies in the amount of resources allocated to the school systems exist across the OECD countries with Denmark traditionally ranking at the very top (OECD 2015). In addition to being a natural consequence of differing educational priorities, these discrepancies further reflect different productivity of school inputs such as class size and instruction time. Here, high-quality compulsory education is acknowledged as an important factor for developing and maintaining a highly qualified labor force, which is often considered the very foundation for future societal growth and welfare. In addition to public school investments, parents are often considered a major determinant for individuals' educational preferences. This dissertation examines select components of these relations.

The dissertation comprises three self-contained chapters within the economics of education and labor. Each chapter empirically investigates individuals at different stages of the Danish educational system. The first two chapters make use of quasi-experimental data as well as a field experiment combined with Danish administrative records for the purpose of identifying the causal effects of school inputs on student achievement. The third chapter employs an intergenerational framework for analyzing the educational choices of individuals with a particular focus on the transmission of social norms and gender attitudes.

The first chapter, *Do class size effects differ across grades?*, analyzes whether the short-term effects of class size are constant across grade levels in compulsory school.[1] While several studies identify negative short- and long-run effects of class size increases (e.g. Angrist and Lavy 1999, Fredriksson et al. 2013), few are concerned with the differential effects across multiple grade levels in the same setup. Using detailed Danish registers covering all students in public compulsory schools, this chapter evaluates the effects of class size on student reading and math achievement in grades 2 to 8. Identification is based on a government-imposed maximum class-size rule that creates exogenous variation in class sizes at enrollment multiples of 28; 29 students on a grade level are divided into two classes, 57 into three etc. This variation is exploited in a 'fuzzy regression discontinuity' framework in which students just above and just below thresholds are compared to identify the causal effect of a class

[1] Published in *Education Economics* 24 (2016), 84-95.

size increase. Consistent with the previous literature, significant (albeit modest) negative effects of class size are found for children at the primary school level and the effect size on math achievements differ significantly from primary to lower secondary school. In addition, larger classes do not affect girls, non-Western immigrants or socioeconomically disadvantaged students differentially compared to other students.

The second chapter, *Increasing instruction time in school does increase learning*, co-authored with Simon Calmar Andersen and Maria Humlum, empirically investigates the effect of school inputs along the dimension of instruction time.[2] Existing evidence of the effectiveness of increasing instruction time is ambiguous (see, for example, Patall et al. 2010), which is reflected in the large cross-country differences in instruction time, both in total lessons and within subjects (OECD 2015). Nonetheless, instruction time is easily manipulative and a key element in many educational reforms, for example in the US and Denmark, in an attempt to improve student achievement. On the other hand, sceptics argue that longer school days generate behavioral problems due to boredom and fatigue.

The results of this chapter is based on a large-scale cluster-randomized field experiment among Danish fourth graders, allowing us to make causal inference. Our findings suggest that increasing instruction time increases student reading achievement and that a general increase in instruction time is at least as efficient as an expert-developed, detailed teaching program with the same weekly increase in lessons despite marginal evidence of increased behavioral problems for boys.

The third and final chapter, *Closing or reproducing the gender gap? Parental transmission, gender, and education choice*, co-authored with Maria Humlum and Nina Smith, studies the persistence in the education choices of individuals in an intergenerational context. Although women have gained a footing on the labor market over the last five decades (Goldin 2014), Olivetti and Petrongolo (2016) note that the convergence has slowed considerably. The remaining gender gaps seem remarkably persistent, particularly considering the reversing gaps in education experienced by most countries. Here, occupational segregation by sex is often emphasized as a major determinant for remaining earnings differences and other gaps (Olivetti and Petrongolo 2016).

In light of a recent literature, which attempts to incorporate the sociological concepts of group identity into the decision making process of individuals, we examine the intergenerational correlation in gender-stereotypical education choice, specifically the extent to which individuals select into female-dominated fields. Using administrative records on the latest cohorts of Danish labor market entrants, we document a positive relationship between gender-stereotypical education choice and parents' stereotypical education and labor market behavior during the individual's adolescent years. Not surprisingly, same-sex parent-child associations dominate. Although, we cannot not unambiguously attribute our results to within-family trans-

[2] Published in *Proceedings of the National Academy of Science* 113(27), 7481-7484.

missions of gender norms, our results are generally consistent with such a hypothesis.

References

Angrist, J. & V. Lavy. 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114 (2): 533–575.

Fredriksson, P., B. Öckert & H. Oosterbeek. 2013. Long-Term Effects of Class Size. *Quarterly Journal of Economics* 128 (1): 249–250.

Goldin, C. 2014. A grand gender convergence: Its last chapter. *American Economic Review* 104 (4): 1091–1119.

OECD. 2015. Education at a Glance 2015: OECD Indicators. OECD Publishing, Paris. doi: 10.1787/eag-2015-52-en.

Olivetti, C. & B. Petrongolo. 2016. The evolution of gender gaps in industrialized countries. Forthcoming in *Annual Review of Economics.*

Patall E., H. Cooper & A. Allen. 2010. Extending the school day or school year. *Review of Educational Research* 80 (3): 401–436.

Dansk resumé

På tværs af de industrialiserede lande er der stor forskel på, hvor mange ressourcer der allokeres til skolesystemerne, og Danmark er traditionelt set et af de lande i verden, der rangerer allerhøjest (OECD 2015). Disse forskelle tager naturligvis afsæt i forskellige prioriteter på uddannelsesområdet men afspejler også, hvor effektivt man er i stand til at udnytte skoleinputs, såsom klassestørrelser og undervisningstid. Her anses kvaliteten af folkeskolen som et vigtigt udgangspunkt i forhold til et fremtidigt højt uddannelsesniveau i befolkningen som grundlag for vækst og velfærd. Foruden det offentliges investeringer i danske elever (i form af fx skoleinput) antages især forældre at have en afgørende rolle i udviklingen af individers uddannelsesmæssige ønsker og formåen. Denne afhandling belyser elementer af sådanne sammenhænge.

Denne afhandling består af tre selvstændige kapitler inden for uddannelses- og arbejdsmarksøkonomi, der alle omhandler empiriske analyser af individer på forskellige stadier af det danske uddannelsessystem. De to første kapitler gør brug af henholdsvis kvasi-eksperimentielle data samt et lodtrækningsforsøg kombineret med danske registerdata i forsøg på at afdække direkte årsagssammenhænge mellem skoleinputs og elevdygtighed i folkeskolen, mens det tredje benytter en mere deskriptiv tilgang til at belyse individers valg af videregående uddannelse: Fokus er her i særdeleshed på overførslen af sociale normer og kønsstereotyper fra individets forældre.

I det første kapitel *Do class size effects differ across grades?* undersøges, hvorvidt klassestørrelseseffekter på elevdygtighed varierer på tværs af klassetrin i folkeskolen.[1] En del studier påviser en negativ kausal sammenhæng mellem klassestørrelser og elevdygtighed på både kort og lang sigt (fx Angrist og Lavy 1999, Fredriksson et al. 2013) men få er i stand til at undersøge effekterne på mere end et enkelt eller få sammenhængende klassetrin. De detaljerede registerdata for alle elever i den danske folkeskole muliggør identifikation af individer, der er indskrevet på skoleårgange, der ligger lige over eller lige under tærsklerne for maksimalt tilladte klassestørrelser. Helt specifikt tillader loven klassestørrelser op til og med 28 elever, således at årgange med 29 elever deles i to klasser, mens årgange med 57 elever deles i tre. Denne eksogene variation udnyttes i et 'fuzzy regression discontinuity design' til at bestemme den

[1] Publiceret i *Education Economics* 24 (2016), 83-95.

kausale effekt af en ændring i klassestørrelse på elevdygtighed. I overensstemmelse med store dele af den øvrige litteratur viser resultaterne, at større klasser i gennemsnit sænker elevernes faglige niveau i ind- og mellemskolingen—omend effektstørrelserne er beskedne. Ydermere varierer effekten på matematikkundskaber signifikant på tværs af klassetrin, mens der ikke findes evidens for, at piger, ikke-vestlige immigranter eller dårligere stillede elever påvirkes i særlig grad i forhold til andre elevgrupper.

Afhandlingens andet kapitel *Increasing instruction time in school does increase learning*, skrevet i samarbejde med Simon Calmar Andersen og Maria Humlum, undersøger en anden dimension af offentlige investeringer i elever, nemlig undervisningstid.[2] Den eksisterende litteratur på dette område er delt (se fx Patall et al. 2010), hvilket i høj grad er reflekteret i de store nationale forskelle i undervisningstid, både totalt og inden for hvert fag (OECD 2015). Ikke desto mindre er undervisningstid relativt let at justere og udgør en betydelig del af flere reformer på skoleområdet, fx i USA og herhjemme, i et forsøg på at forbedre elevers faglige præstationer. Skeptikere fremhæver dog ofte, at en forlænget skoledag vil medføre træthedseffekter hos eleverne i form af større koncentrationsbesvær og adfærdsvanskeligheder.

Dette kapitel er baseret på et stort lodtrækningsforsøg blandt danske 4. klasser, hvilket muliggør en kausal fortolkning af vores resultater. Resultaterne viser, at øget undervisningstid fremmer elevers faglige præstationer, samt at en generel forhøjelse af antallet af undervisningstimer i dansk, på trods af marginalt flere adfærdsmæssige problemer hos drengene, er mindst ligeså gavnlig som et detaljeret undervisningsprogram udviklet af danske eksperter.

Det tredje og sidste kapitel *Closing or reproducing the gender gap? Parental transmission, gender, and education choice*, skrevet i samarbejde med Maria Humlum og Nina Smith, træder et skridt tilbage og belyser individers uddannelsesvalg i et intergenerationelt perspektiv. Selvom kvinder generelt har vundet indpas på arbejdsmarkedet igennem det sidste halve århundrede (Goldin 2014), noterer blandt andre Olivetti og Petrongolo (2016) sig, at denne trend er stagneret i de senere år. De tilbageværende kønsforskelle i fx indkomster er bemærkelsesværdigt stabile, særligt i betragtning af at kvinder i de fleste lande for længst har overhalet mænd med hensyn til uddannelseslængde—og nu blot udvider kløften. En forklaring, der ofte fremhæves, er, at tilbageværende indkomstforskelle kan forklares med, at kønnene selv i nyere tid vælger vidt forskellige uddannelser og erhverv (Olivetti og Petrongolo 2016).

Set i lyset af en ny litteratur, der forsøger at indarbejde koncepter fra sociologiens verden omhandlende gruppeidentitet og stereotyper i individers beslutningsprocesser, undersøger vi den intergenerationelle korrelation i kønsstereotypt valg af videregående uddannelse, nærmere betegnet i hvor høj grad individer vælger en uddannelse, der er domineret af kvinder. Ved brug af dansk registerdata på de seneste årgange til at indtræde på arbejdsmarkedet dokumenteres en positiv sammenhæng mellem individers valg af uddannelse og deres forældres kønsstereotype uddannelses-

[2] Publiceret i *Proceedings of the National Academy of Science* 113(27) (2016), 7481-7484.

og arbejdsmarkedsadfærd under opvæksten. Ikke uventet er sammenhængene stærkere for førstefødte og forælder–barn-par af samme køn, dvs. mor–datter og far–søn. Vi er ikke i stand til entydigt at fortolke disse korrelationer som overførsler af kønsroller i familien, men noterer os at vores resultater generelt er i overensstemmelse med en sådan hypotese.

References

Angrist, J. & V. Lavy. 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114 (2): 533–575.

Fredriksson, P., B. Öckert & H. Oosterbeek. 2013. Long-Term Effects of Class Size. *Quarterly Journal of Economics* 128 (1): 249–250.

Goldin, C. 2014. A grand gender convergence: Its last chapter. *American Economic Review* 104 (4): 1091–1119.

OECD. 2015. Education at a Glance 2015: OECD Indicators. OECD Publishing, Paris. doi: 10.1787/eag-2015-52-en.

Olivetti, C. & B. Petrongolo. 2016. The evolution of gender gaps in industrialized countries. Forthcoming in *Annual Review of Economics*.

Patall E., H. Cooper & A. Allen. 2010. Extending the school day or school year. *Review of Educational Research* 80 (3): 401–436.

CHAPTER 1

DO CLASS SIZE EFFECTS DIFFER ACROSS GRADES?

A SLIGHTLY SHORTER VERSION WAS PUBLISHED IN
EDUCATION ECONOMICS 24(1 & 2) (2016), 83–95

Anne Brink Nandrup
Aarhus University

Abstract

This paper contributes to the class size literature by analyzing whether short-run class size effects are constant across grade levels in compulsory school. Results are based on administrative data on all pupils enrolled in Danish public schools. Identification is based on a government-imposed class size cap that creates exogenous variation in class sizes. Significant (albeit modest) negative effects of class size increases are found for children on primary school levels. The effects on math achievement are statistically different across grade levels. Larger classes do not affect girls, non-Western immigrants and socioeconomically disadvantaged pupils more adversely than other pupils.

JEL Classification: I21; I28; C31
Keywords: Class size; Regression discontinuity; LATE; Compulsory schooling; Literacy; Test scores

1.1 Introduction

A primary goal of the education production function literature is to understand the technology of schooling inputs such as class size in the creation of cognitive achievement outcomes. Politically speaking, class and school sizes are also a recurrent issue. These school inputs are readily measured and in general considered easier to manipulate. Furthermore, increasing class sizes comprise large budget savings; In OECD, teachers' salaries alone constitute 62% of compulsory schooling expenditures (OECD 2012). Such cost reductions may come at a price, however. Recent work by Fredriksson, Öckert and Oosterbeek (2013) suggests significant adverse long-term effects from increasing class sizes in upper primary school.

This paper evaluates the short-run effects of class size on pupil achievement in both mathematics and reading across grades 2 to 8 for recent cohorts in compulsory school. While a vast, but inconclusive, literature is concerned with identifying short- and medium-run effects of class size increases in primary and secondary school, previous studies are often concerned with only one or a few close grades in the same setup (e.g. Angrist and Lavy (1999) on 4^{th} and 5^{th} graders, Hoxby (2000) on 4^{th} and 6^{th} graders). This is hardly sufficient seeing as compulsory education in most developed countries amounts to 9 or 10 years. The Tennessee STAR experiment (e.g. Finn and Achilles 1999) on preschool to grade 3 and Rivkin, Hanushek and Kain (2005) on grades 4 to 7 cover the greatest ranges in terms of grade levels. Unfortunately, comparisons of existing findings are complicated by varying institutional settings as well as incomparable outcome measures. Thus, the literature provides little empirical insight into the structure of class size effects across the years of compulsory schooling. This paper remedies this by employing the same identification method across grade levels in compulsory school using directly comparable measures of pupil achievement as the outcomes of interest. The sample covers all pupils in Danish public compulsory schools over a three-year period.

The effect of class size is unlikely to be constant across grade levels for several reasons which could help explain some of the mixed evidence in the literature. For example, smaller classes may in particular benefit pupils in the lower grades: Younger pupils may depend more on adult supervision and help, therefore, peer-tutoring or group work may be more effective in older grades (Blatchford and Mortimore 1994). The self-control of pupils may increase with age but also a number of other psychological and hormonal factors change as pupils mature. Mischel and Mischel (1983) show that older children can create a more favorable environment for effective self-control. As such, older pupils may be less inclined to participate in interrupting behavior. Conversely, parents may be more qualified to assist their children with homework, supplementary reading etc. in the early school years, thus effectively reducing the need for teacher one-on-one time for younger pupils.

The evidence from the existing literature is inconclusive. Based on the STAR experiment Finn and Achilles (1999) find significant positive effects of around 0.2 SD of decreasing class size by an average of 7 pupils in preschool to grade 3. However, Krueger (1999) estimates that these effects are driven primarily by the first year in a small class. Rivkin, Hanushek and Kain (2005) find significant negative (though smaller) effects of increasing Texan class sizes only in grades 4 and 5, also suggesting that the class size effects fade over time. Exploiting exogenous variation in class sizes created by administrative rules, Angrist and Lavy (1999) identify negative effect sizes in the lower end of project STAR for Israeli 5^{th} graders (nothing on 4^{th} graders), while Hoxby (2000) finds no relationship between class size and academic achievement for both 4^{th} and 6^{th} graders in Connecticut. Using the same approach on Danish data, Heinesen and Browning (2007) and Krassel and Heinesen (2014) identify modest negative class size effects in grades 8 and 10, respectively. Finally, in Norway Leuven, Oosterbeek and Rønning (2008) find no significant effect of increasing average class size in grades 7–9 on student achievement, while Fredriksson, Öckert and Oosterbeek (2013) do for average class size in grades 4–6 in Sweden. In both cases, average class size is instrumented by expected class size based on administrative rules and enrolment in the earliest grade.

Exploiting test results from the unique Danish national test system in combination with detailed register-based data, this paper identifies the effects of changes in class size on test results for three different levels of compulsory schooling: lower and upper primary and lower secondary school. Following in the footsteps of Angrist and Lavy (1999), I employ a fuzzy regression discontinuity design arising from a government imposed maximum class size rule of 28 pupils. I apply this identification strategy to data covering all public school pupils in Denmark between 2009/2010 and 2011/2012. As learning processes likely differ across linguistic and logical subjects, the effects of increased class sizes on reading and math achievement are studied separately.

Results show significant (albeit modest) negative effects of increasing class sizes in the Danish public schools where the average class size is 21 with a modal value of 23. Most effects of a class size increase in primary school are significantly negative, particularly for 6^{th} graders, whereas none of the lower secondary level estimates are significant. More importantly, I am able to reject that the results for math are similar across primary and lower secondary school. Further, I employ various robustness checks to underpin the validity of the results presented here.

1.2 Institutional setting

There are 98 municipalities in Denmark, each of which is divided into one or more school districts. The residential address of the pupils determines their school district affiliation, however, since 2006 pupils are only entitled, but not required, to attend the district school. Public schools are free and financed by local municipalities through a combination of municipality income tax and a between-municipality redistribution scheme subsidizing expenditures in low-income municipalities. Furthermore, public schools are subject to a government-imposed maximum class size rule of 28 pupils per classroom[1], but class sizes vary considerably across schools and cohorts. The school structure implies that municipalities, rather than schools independently, finance the expenditures associated with the maximum class size rule. Approximately 86% of Danish children were enrolled in public schools in 2009/2010–2011/2012.

Pupils are taught from the calendar year they turn six years old, beginning with preschool (preschool was optional before 2009). Public schools typically contain grades 0–9 (smaller schools may only contain grades 0–7). Pupils are generally divided into classrooms when they enrol in preschool (grade 0) and follow the same class throughout compulsory school with few exceptions, for example elective third language. These subjects are usually not introduced until grade 7. Teachers are subject-specific and follow classes through (parts of) compulsory school. The public school system builds on the principle that pupils cannot be tracked according to ability or social background. Consequently, there are no elite schools or classes in the public system.

There is no formal division of the levels in Danish compulsory schools, but following the literature this paper denotes three overall grade levels: lower primary school (grades 1–3), upper primary school (grades 4–6) and lower secondary school (grades 7–9).

In 2010, a national, standardised test programme was introduced to the public compulsory schools. The mandatory programme includes reading tests (grades 2, 6, and 8), math tests (grades 3 and 6), and a physics/chemistry test (grade 8)[2]. There is no math test beyond grade 6, but as physics and math are often considered to be based on somewhat similar mindsets, test results in physics/chemistry act as substitutes.

All tests are electronic and self-scoring, thus, the score itself is not influenced by the teacher. Moreover, the tests are adaptive and therefore adjust to the proficiency level of the pupil during the test session (45 min.), hence, pupils' subject-specific skills are very precisely determined compared to regular linear tests. Each test measures pupil proficiency within three separate cognitive dimensions of a subject on a Rasch-calibrated logit scale, see Beuchert and Nandrup (2014) for details. Test scores within each cognitive dimension are then combined to measure the overall proficiency of the pupil in that subject. The nature of the tests makes them qualified for comparison both across and within individuals.

[1] To accommodate potential classroom divisions outside of the summer break caused by late school transfers, up to 30 pupils are allowed per classroom *during* the school year.

[2] Further tests include reading (grade 4), English (grade 7), biology and geography (grade 8).

1.3 Data and identification

1.3.1 Data

The data set contains registry data on all pupils in the Danish school system and their test results, maintained by The Danish Ministry of Children and Education, combined with registry data on pupils and their parents, maintained by Statistics Denmark. School enrolment is registered annually in September and allows one to construct beginning-of-the-school year class sizes and grade enrolments in all schools. Test scores are obtained in January through April and are available from the school year of 2009/2010. This allows me to base the results of this paper on all mainstream classroom pupils (thus excluding pupils in full-time segregated special education classrooms and schools as well as mixed grade classrooms) in certain public school grades in the school years of 2009/2010 – 2011/2012 (4,259 school×years). Out of 965,136 observations in either grades 2, 3, 6 or 8, 71,701 observations (7.43%) are dropped from the sample because of unobserved test results[3]. The outcome of interest, pupil achievements in reading and math (physics/chemistry), is a constructed average of the test scores within each of the three cognitive dimensions (standardised across tests and years) standardised to mean zero and unit variance across tests and years.

The explanatory variable of interest is class size. However, test results may be observed up to eight months after class registration, which means that class size may have changed meanwhile. Conversely, beginning-of-the-school year class size and enrolment may be more 'exogenous' because they are less likely to be affected by parents. Also, one may argue that class size during the school year is just as important for skill accumulation. Table 1.1 shows summary statistics of the key explanatory variables in the six subpopulations. Across subsamples, grade 8 enrolment is around 10 pupils higher and correspondingly class sizes increase with approximately 0.5 pupil. Excluding schools that absorb lower secondary pupils from small schools that do not offer grades 8 and 9, class sizes and enrolment counts are similar across subsamples.

A few classes (5.61%) contain below 14 pupils while only 0.36% are larger than the 28-pupil cap size. The latter are not excluded from the sample, because I am hesitant to condition on the endogenous variable. All results in Section 1.4 are robust to the exclusion of 'too' large classes.

Other explanatory variables include: school characteristics and detailed pupil-specific information such as birth and family information and socioeconomic status (from the year of the pupil's sixth birthday). A complete list of controls including descriptive statistics hereof is found in the Appendix (Table A.1), where regression

[3] 66% of the missing test results occur in 2010 where the test system suffered from a nationwide, technical breakdown that unexpectedly cancelled two weeks of testing. These tests are unidentifiable but likely missing at random. In terms of parental characteristics, pupils with missing test results are slightly negatively selected, but they represent a relatively small fraction of the sample.

Table 1.1: Summary statistics of key variables

Variable	Reading sample Mean	SD	Math (physics/chemistry) sample Mean	SD
	Grade 2 $(N = 150,065)$		*Grade 3* $(N = 152,800)$	
Class size	21.20	4.06	21.19	4.03
Enrolment	50.65	22.21	50.06	21.70
Std. test result	0.01	0.99	0.01	0.99
	Grade 6 $(N = 153,810)$		*Grade 6* $(N = 153,846)$	
Class size	21.37	3.91	21.36	3.86
Enrolment	50.25	21.23	50.35	21.27
Std. test result	0.04	0.96	0.03	0.97
	Grade 8 $(N = 141,938)$		*Grade 8* $(N = 140,975)$	
Class size	21.87	3.58	21.86	3.57
Enrolment	60.97	22.38	60.92	22.45
Std. test result	0.05	0.94	0.02	0.98

results demonstrate that the included controls indeed are relevant predictors of pupil achievement.

1.3.2 Identifying variation

The causal effect of class size is rather difficult to study because the majority of class size variation is the result of choices made by parents, school administrators, teachers and politicians on a local or national level. Thus, class size is potentially correlated with other determinants of pupil achievement. As originally suggested by Angrist and Lavy (1999), this paper uses exogenous variation in class sizes created by the 28-pupil rule as an instrument for the endogenous class size variable. Following the authors, the expected class size, assuming cohorts are divided into classes of equal size of grade g in school s, is given by eq. (1.1):

$$f_{gs} = {}^{e_{gs}}/\left(floor\left({}^{(e_{gs}-1)}/_{28}\right)+1\right) \tag{1.1}$$

where e_{gst} denotes the grade level enrolment and $floor(n)$ the largest integer less than or equal to n. Eq. (1.1) reflects that subject to the 28-pupil rule enrolments of up to 28 pupils are assigned to one class while enrolments between 29 and 56 are divided into two classes of 14.5–28 pupils each, etc.

Figure 1.1 illustrates the relationship between grade enrolment and expected (solid line) and mean observed (dots) class sizes in the full estimation sample. f_{gst} is shown to generally be a strong predictor of actual class size, at least for the three lower thresholds.[4] Here, the probability of treatment (assignment to a smaller class) is higher just to the right of the cutoffs than just to the left. This is a necessary feature in a regression discontinuity context, however treatment is not guaranteed for all pupils.

[4] The pattern is largely consistent across grades with a somewhat poorer fit for the eighth grade.

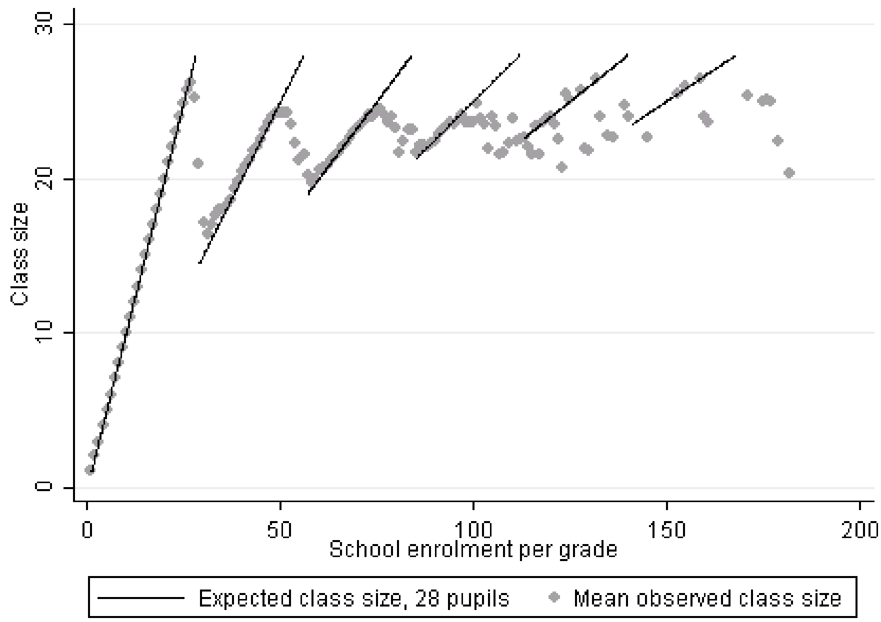

Figure 1.1: Expected and mean observed class size by enrolment, the full estimation sample. The sample includes mainstream class room pupils in grades 2, 3, 6, and 8 in public schools (2009/2010–2011/2012). Expected class size by enrolment is based on the 28-pupil rule.

1.3.3 Is the regression discontinuity design valid at the school level?

In a regression discontinuity context, random assignment of treatment intensity may be undone by administrator sorting when the assignment rule is public knowledge. In Denmark, the decision-making authority regarding school districts and school catchment areas lies with the municipality. Thus, municipalities are entitled to change the school catchment areas and school districts if deemed necessary. In practice this entitlement is implemented differently across Danish municipalities, and there is only very few examples of yearly school district revisions. Because of discontinuities in the enrolment count of Swedish schools, Fredriksson, Öckert and Oosterbeek (2013) are compelled to focus on school district enrolment rather than on the school level. The discontinuities arise as Swedish legislation encourages adjustment of school catchment areas within school districts to utilise demography and school resources optimally. Also, Urquiola and Verhoogen (2009) document an extreme case of bunching based on class size caps in Chilean subsidized private schools. To examine the Danish setting, Figure 1.2 illustrates the distribution of grade 1 enrolments in the school year of 2009/2010. As municipalities can only adjust school catchment areas before cohorts enrol into schools and preschool was not mandatory until 2009, grade

1 enrolment is the most relevant distribution to examine. By visual inspection, there are no clear indications of bunching below the thresholds caused by administrator sorting. Furthermore, the free school choice should at least partly offset this.

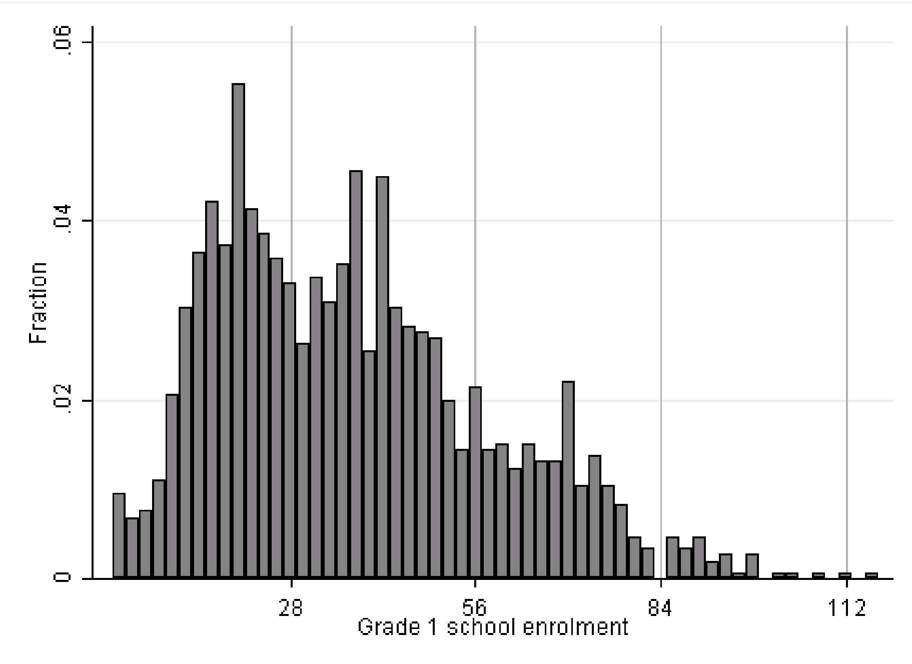

Figure 1.2: Distribution of grade 1 school enrolment, 2009/2010. The sample includes mainstream classroom 1^{st} graders in Danish public schools. Vertical lines indicate thresholds created by multiples of the 28-pupil rule.

1.3.4 Estimation strategy

Exploiting the exogenous variation in class sizes induced by the 28-pupil rule, it is possible to interpret the effects of class size on pupil achievement causally (Angrist and Lavy 1999). The model is estimated by two-stage least squares (2SLS):

$$y_{icgs\tau} = \boldsymbol{X}_{icgs}\boldsymbol{\alpha}_2 + \alpha_1 CS_{cgs\tau} + \varphi_\tau + h_\tau\left(e_{gs}\right) + \varepsilon_{icgs\tau} \tag{1.2}$$

$$CS_{cgs\tau} = \boldsymbol{X}_{icgs}\boldsymbol{\gamma}_2 + \boldsymbol{above}_{gs}\boldsymbol{\gamma}_1 + \phi_\tau + q_\tau\left(e_{gs}\right) + \upsilon_{icgs\tau} \tag{1.3}$$

where $y_{icgs\tau}$ denotes the standardised test result of individual i in class c of grade g at school s at enrolment segment τ. $\boldsymbol{X}_{icgs}$ is a vector of controls (including pupil, parental and school characteristics and cohort dummies). CS_{cgs} denotes observed class size and the residuals, $\varepsilon_{icgs\tau}$ and $\upsilon_{icgs\tau}$, are idiosyncratic. (1.2) and (1.3) include segment fixed effects, φ_τ and ϕ_τ, to accommodate different patterns around the separate enrolment thresholds[5]. Also, the coefficients of the second order enrolment polynomials, $h_\tau\left(e_{gs}\right)$ and $q_\tau\left(e_{gs}\right)$, are allowed to vary by segments.

Class size is instrumented by the binary indicators $\boldsymbol{above}_{gs}$ equalling one for grade enrolments above thresholds and zero otherwise: $above28 = \mathbf{1}\,[28 < e \leq 42]$, $above56 = \mathbf{1}\,[56 < e \leq 60]$ etc. This setup highlights the quasi-experimental identification strategy and ignores the smooth variation in the expected class size between thresholds. Heinesen and Browning (2007) argue that this is the most appropriate specification because only variation in the instrument around thresholds is used. By allowing the enrolment polynomials to vary by segment, I follow Fredriksson, Öckert and Oosterbeek (2013) and effectively consider each threshold as a different experiment. Further, analyses that focus on variation only around the thresholds are conducted. I decide on an interval of ± 4 pupils to avoid extraordinary demands on the data. The three lower thresholds are chosen because of the greater predictive power of the instruments. This strategy is preferred as it to a greater extent removes possible effects of permanent schools characteristics that is not captured by the enrolment controls.[6]

The coefficient α_1 is of primary interest. It captures a weighted average treatment effect to a unit change in class size for the unknown subpopulation of pupils whose treatment status is affected by the instrument in a setting where class size effects are heterogeneous and non-linear (Angrist and Imbens 1995). The weights are proportional to the number of pupils who, because of the 28-pupil rule, are induced to attend a smaller class. Identification arises given independence and monotonicity assumptions: Independence requires that treatment is as good as randomly assigned, which is closely related to the exclusion restriction that requires instruments not to affect test results other than through their effect on class size. Monotonicity requires that the group of 'defiers' is empty, i.e. that for all schools class size given enrolment above the government-imposed threshold is never larger than it would have been had the enrolment been below the threshold. Both assumptions are non-testable but generally a stronger instrument causes the IV estimand to be less sensitive to violations (Angrist and Imbens 1995).

Better schools likely face increased demand, thus, enrolment and instruments are potentially related to pupil achievement for reasons other than class size. This

---

[5]Each segment consists of ± 14 pupil intervals around threshold τ: $\varphi_\tau = \mathbf{1}\left[e_{gs} \in \overline{e}_\tau \pm 14\right]$, where $\overline{e}_\tau = \{56, 84, 112, 140\}$. The first segment also includes enrolments below 15 pupils: $\varphi_{28} = \mathbf{1}\left[e_{gs} \leq 42\right]$. Where the narrower bandwidths around the lower thresholds are used, segment τ consists of $\varphi_\tau = \mathbf{1}\left[e_{gs} \in \overline{e}_\tau \pm 4\right]$, where $\overline{e}_\tau = \{28, 56, 84\}$.

[6]Results are driven by smaller schools. Firstly in the ± 4 pupil interval sample, because only observations around the three lower thresholds are included. Secondly, even when all thresholds are included, the instruments cause a much greater difference in class size around the first (13.5 pupils) compared to the second and third (9 and 6.75, respectively), and the weighted average causal treatment effect places larger weight on observations that are more affected by the instruments (Angrist and Imbens 1995, Angrist and Lavy 1999).

relation is, however, expected to be a more or less smooth function of enrolment and highlights the need for including sufficient controls for enrolment effects (Angrist and Lavy 1999). The most popular choice of controls is a second order polynomial in enrolment (e.g. Krassel and Heinesen 2014). Analogous to the extensive specification analysis in Fredriksson, Öckert and Oosterbeek (2013), I find that the results are largely stable across enrolment specifications. Conceptually, I prefer flexibility, thus, the full sample results are based on a more flexible specification where the slopes of the second order polynomial in enrolment are allowed to vary across segments.[7] Likewise, results do not change when settling for a piecewise linear trend in enrolment with slopes identical to the slopes of f_{gst} (see Angrist and Lavy 1999).

In Denmark, a free school choice effectively reduces school transfer costs. This is potentially problematic as parents may be more inclined to exploit the 28-pupil rule and undo the random assignment of class sizes. However, if parents are not able to *precisely* manipulate the assignment variable, the variation in treatment near the thresholds should be randomized (Lee and Lemieux 2010). Intuitively, parents can roughly predict class size based on district size, but as treatment depends on the enrolment of all other children in the school cohort, it would be risky to actively choose schools based on enrolments in small intervals around thresholds—particularly as school transfers always involve costs (disruptions, loss of peers etc.).

Generally speaking, parents can evade the 28-pupil rule in two ways: when enrolling their child into compulsory school and by school transfers during the school year. In 2008–2011 more than 26% of the Danish public schools had a different number of classes on the first grade level compared to the year before. Thus, before entering compulsory school parents may have difficulties anticipating the class size based on previous years. However when pupils transfer schools during the school years, class sizes in the receiving schools are already observed. There are three reasons why this is less problematic. Firstly, identification is based on beginning-of-the-school year class sizes, thus sorting during the school year does not affect the results (recall footnote 1). Secondly, for schools with grade enrolments in small intervals around the thresholds it would still be risky to predict class size. Finally, choice of school is presumably based on many other factors than just class size. In the end, the school headmaster decides which class to enrol a new pupil in, given that there are several classes on the grade level. It seems unlikely that parents would select their children into very small public schools, usually located in the countryside, to be certain

[7] In the published paper, the tables erroneously denote that full sample regressions include a second-order polynomial in enrolment interacted with both segments and thresholds. However, as in Fredriksson, Öckert and Oosterbeek (2013) this specification is too flexible and absorbs the variation created by the instruments. Fredriksson, Öckert and Oosterbeek (2013) settle for a linear function of enrolment interacted with both segments and thresholds instead, which is also slightly too flexible in the setup presented here (and thus omitted); the F-test for excluded instruments are reduced to around 10 while the absolute size of the full sample estimates presented here are somewhat increased. Nonetheless, I base my main conclusions on the preferred specifications on the sample of ±4 pupils around the three lower thresholds as discussed above, which are unchanged.

of a small class. Besides larger transportation costs and potentially poorer family characteristics of classmates, countryside schools are generally associated with less flexibility and less specialisation and diversity of teachers.

Table 1.2 tests the significance of the coefficient of being above a threshold when regressing selected baseline variables separately on a pooled version of the instrument (for completeness, Table A.2 additionally presents the estimated coefficients to shed light on the magnitude of differences in means above and below thresholds).[8] Column (1) shows regression results for all pupils. Merely a few covariates are unrelated to the pooled instrument; thus, one has reason to suspect that treatment is not randomly assigned across thresholds. However, in a smaller ±4 pupil interval around the three lower cutoffs, few of the predetermined characteristics are related to the instrument (column (2)). Also, when regressing all baseline covariates on being above a cutoff for ±4 pupils around the three lower cutoffs, all coefficients are jointly insignificant (p-value .195). Across subsamples, control variables are generally insignficantly related to placement above a cutoff for ±4 pupils around the three lower cutoffs in all execpt the 3^{rd} grade math sample (see the lower panel of Table A.2).

Because of limited data on other school inputs, estimated class size effects should be interpreted as 'total policy effects' (Todd and Wolpin 2003). I.e. the *ceteris paribus* effect of a class size increase plus an indirect effect through the responses of other inputs. Although one is usually interested in total policy effects, these estimates provide little insight into the education production function.

1.4 Results

This section quantifies the effect of class size on math and reading achievement using the empirical approach outlined in Section 1.3.4. All reported standard errors are clustered to account for group structures in the residuals within grade enrolments.[9]

1.4.1 Main results

Table 1.3 presents 2SLS estimates of the class size effect on pupil achievement in reading and math across compulsory school. Results in the full sample are obtained using a flexible enrolment specification where a second order polynomials of enrolment is interacted with segments and thresholds to fully account for enrolment effects.[10]

[8] For simplicity, p-values are from regressions on a pooled binary indicator for being above *any* threshold. Results carry through for regressions on each *above*-indicator separately (available on request).

[9] Clustering by enrolment count is suggested by Lee and Card (2008) and performed in Fredriksson, Öckert and Oosterbeek (2013). This yields 136 clusters in the full estimation sample, considerably less than when clustering on the school grade by year level where the instrument varies. Thus, standard errors are slightly larger but the difference is modest.

[10] A specification analysis using OLS and 2SLS with only a smooth second order polynomial in enrolment both with f_{gst} and the *above*-dummies as instruments is summarized in the Appendix (columns (1)–(3), Table A.4).

Table 1.2: Balancing of covariates

Baseline covariate	(1) p-value, *above* cutoff All pupils	(2) p-value, *above* cutoff ±4 pupils around lower cutoffs
Mother's education:		
—Basic	.026	.488
—Vocational	.148	.369
—Higher	.000	.137
Father's education:		
—Basic	.012	.272
—Vocational	.713	.040
—Higher	.000	.614
Mother's log-earnings	.000	.187
Mother's age	.000	.082
Father's log-earnings	.006	.645
Father's age	.001	.088
Single mother	.012	.138
Number of siblings	.013	.913
Female	.879	.574
Western immigrant	.065	.593
Non-Western immigrant	.002	.075
Birth weight	.161	.275
Gestational age	.040	.017
First-born	.055	.358
Second-born	.848	.781
Multiple born	.639	.817
No. of observations	893,434	188,061

Notes. The *above* cutoff indicator equals 1 if the grade enrolment exceeds a threshold created by the 28-pupil rule up to +14 pupils (+4 pupils around the three lower cutoffs in (2)). Columns report the p-values for t-tests of the significance of the pooled instrument by separate OLS regressions of each variable listed on the instrument. All regressions additionally include year and enrolment segment fixed effects and indicator variables of degree of urbanization of the school municipality. Columns (1) further include linear and square controls for grade enrolment interacted with separate thresholds. Table A.2 presents an extended balancing analysis. Standard errors are adjusted for clustering by enrolment count.

Whereas segment fixed effects control sufficiently for enrolment effects when the sample is limited to close intervals around thresholds. Columns (1) and (2) present results for the full sample, while (3) and (4) include only ±4 pupil intervals around the three lower cutoffs. Specifications (1) and (3) do not include other baseline covariates.

If the 28-pupil rule produces experimental variation in class size, the 2SLS estimates should be robust to the inclusion of controls—they should only improve the precision of the estimates. Particularly the coefficients including only ±4 pupils around thresholds, where treatment is more likely to be random cf. Section 1.3.3, are robust to the inclusion of controls.[11] Also, F-statistics of excluded instruments clearly reject the null of weak instruments.[12]

[11] To accommodate a serious threat to identification caused by potential measurement error in the enrolment variable, analyses excluding observations at the cutoffs (i.e. 28, 29, 56, 57 etc.) are also conducted. Results and conclusions are robust to this exclusion (available on request).

[12] The upper panel of Table A.3 presents the first stage for the preferred specification on the sample with ±4 pupils around the three lower thresholds including all controls. Placement above a threshold is generally significantly related to a smaller class size, although less convincingly around the third threshold. Evidence from the reduced form estimates (lower panel) further associate placement above the two lower

Table 1.3: 2SLS estimates, class size effects in grades 2–8

Outcome variable	(1) All pupils	(2)	(3) ±4 pupils around lower cutoffs	(4)
2^{nd} grade reading score	−.0015	−.0071***	−.0047	−.0101***
	(.0035)	(.0030)	(.0045)	(.0036)
F-test (excl. instruments)	41.50	43.14	49.02	53.77
No. of observations	150,065		31,061	
3^{rd} grade math score	−.0026	−.0044	−.0073	−.0059
	(.0039)	(.0029)	(.0056)	(.0040)
F-test (excl. instruments)	27.80	28.08	28.97	28.66
No. of observations	152,800		33,404	
6^{th} grade reading score	.0001	−.0044	−.0123***	−.0125***
	(.0039)	(.0029)	(.0045)	(.0037)
F-test (excl. instruments)	33.30	33.87	36.59	35.80
No. of observations	153,810		31,543	
6^{th} grade math score	−.0033	−.0078**	−.0144***	−.0149***
	(.0039)	(.0034)	(.0050)	(.0041)
F-test (excl. instruments)	33.39	33.92	36.89	35.73
No. of observations	153,846		31,543	
8^{th} grade reading score	.0027	−.0014	−.0060	−.0078
	(.0059)	(.0044)	(.0080)	(.0067)
F-test (excl. instruments)	17.87	18.42	26.75	29.36
No. of observations	141,938		30,424	
8^{th} grade physics-/chemistry score	.0039	−.0004	.0008	−.0014
	(.0044)	(.0039)	(.0055)	(.0048)
F-test (excl. instruments)	16.39	16.44	29.21	31.96
No. of observations	140,975		30,086	
2^{nd} order polynomial of enrolment	YES	YES	NO	NO
Enrolment interacted w/ segments	YES	YES	NO	NO
All controls	NO	YES	NO	YES

Notes. The estimates are based on pupils in mainstream classrooms in the Danish public schools in 2009/2010 – 2011/2012. Specifications (3) and (4) only include pupils enrolled in schools with a grade enrolment of ±4 around the three lower cutoffs: 28, 56, and 84. In addition to the control variables listed in the table, all specifications include fixed effects for enrolment segments. Controls include the remaining covariates from Table A.1. Standard errors adjusted for clustering by grade enrolment are in parentheses, *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Although modest in magnitude, all coefficients in columns (2) and (4) of Table 1.3 are negative. Thus, increasing class size seems to harm pupil achievement in both reading and math, though only significantly in primary school where a marginal class size increase lowers reading and math scores with around .01 SD (generally, the largest effects are found in the narrow sample of ±4 pupils around thresholds). Compared to this, results of an OLS specification suggest compensatory allocation of class size. The OLS estimates of class size on reading scores vary between insignificant −.001 and significant .006 of a standard deviation (on 2^{nd} and 8^{th} grade reading results, respectively) (see column (1), Table A.4). These effect sizes are in the lower end compared to other studies finding significant class size effects: findings from the STAR experiment suggest effect sizes of around .20 SD of lowering class sizes

thresholds with positive pupil achievement.

by 7 pupils (e.g. Finn and Achilles 1999), Fredriksson, Öckert and Oosterbeek (2013) report an effects size of .033 SD from increasing average class size in grades 4–6 by one pupil and Angrist and Lavy (1999) suggest an effect size of around .023 SD based on a recalculation of class means in grade 5. However, in a recent study on Danish 10^{th} graders, Krassel and Heinesen (2014) report effect sizes of .008 SD on GPA, corresponding to the results presented here. Contrary to Rivkin, Hanushek and Kain (2005) I find the largest significant negative effects of class size in grade 6, these are not decreasing until grade 8.

Table 1.4 presents the results in a setting where class size is interacted with grade levels, thus allowing one to study the significance of the differences in class size effects directly. Here, the main effect pertains to pupils in the upper primary school (grade 6). Specifically, grade indicators are interacted with class size and instrument as well as enrolment control functions and segment fixed effects.

Table 1.4: 2SLS estimates, class size effects in grades 2–8, interaction specification

Independent variable	(1) 2SLS, reading score	(2)	(3) 2SLS, math score	(4)
Interaction	−.0039	.0022	.0032	.0091
(lower primary school)	(.0045)	(.0058)	(.0044)	(.0056)
Main effect	−.0034	−.0119***	−.0075**	−.0148***
	(.0031)	(.0036)	(.0035)	(.0041)
Interaction	.0023	.0044	.0078	.0133**
(lower secondary school)	(.0045)	(.0064)	(.0049)	(.0061)
No. of observations	446,113	93,028	447,621	95,033
All pupils	YES		YES	
±4 pupils around cutoffs		YES		YES
All controls	YES	YES	YES	YES

Notes. Table note (1.3) applies. The lower primary school interaction term pertains to grade 2 in columns (1) – (2) and grade 3 in columns (3) – (4). The main effect pertains to grade 6 while the lower secondary school interaction term denotes the differential class size effect from grade 6 to 8. In addition to the control variables listed in the table, specifications (1) and (3) include segment fixed effects interacted with school level and linear and squared controls for grade enrolment into schools interacted with both segment and school level. Specifications (2) and (4) only include segment fixed effects interacted with school level. Instruments are interacted with school levels as well. Standard errors adjusted for clustering by grade enrolment are in parentheses,*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table 1.4 shows that increased class sizes in upper primary school generally decrease test results. However, the difference across school levels is only significant for lower secondary math achievements.

The results of Tables 1.3 and 1.4 are robust to alternative specifications of class size. Replacing beginning-of-school year class size with a three-year average, to accommodate the hypothesis that not only contemporaneous class size may affect skill accumulation, does not change the conclusions (see column (4), Table A.4). Here, enrolment and instrument specifications are based on information two years prior to the outcome measure. Standard errors are slightly increased, which causes more imprecise estimates. Likewise, replacing class size, enrolment and instruments with the corresponding grade 1 information (only feasible for grades 2 and 3 results due to data limitations) does not change the magnitude of the estimates (available on request).

Importantly, given monotonicity identification arises only when the 28-pupil rule binds. Though such a strategy is unnecessarily expensive (and less credible), some municipalities may choose to operate under a lower class size cap than 28. To accommodate this potential pitfall, I use a bandwidth of ± 14 pupils around the two lower thresholds to estimate the discontinuities in grade 1 class sizes on the basis of different maximum class size rules. This strategy is applied separately to each municipality and provides evidence of municipalities certainly abiding by the 28-pupil rule. The exercise leaves 48 municipalities, but results are unchanged although of greater magnitude (see column (5), Table A.4).

1.4.2 Heterogeneity

To examine whether class size effects are heterogeneous, Table 1.5 presents results of the ± 4 pupil sample around the three lower cutoffs where class size is interacted with gender, parental income and immigrant status. Here, for example, gender is interacted with the class size and the instruments as well as enrolment segment.

In line with e.g. Fredriksson, Öckert and Oosterbeek (2013) and Krassel and Heinesen (2014) also on Scandinavian data, Table 1.5 reveals little evidence of systematic effects of class size across pupil characteristics. Neither girls nor immigrants from non-Western countries (or descendants hereof) are more adversely affected by increased class sizes. If anything, 2^{nd} grade reading skills of non-Western immigrants seem to be slightly improved by a larger class. A general concern is that children from disadvantaged backgrounds are more adversely affected by a decrease of school resources. However, it seems that schools (and teachers in particular) are observant of these children when the class size is large, preventing them from falling further behind. Correspondingly, children from low-earnings families[13] are not more adversely affected by larger class sizes in Denmark. A similar pattern emerges when interacting class size with the education level of the parents (omitted here). Interestingly, class size effects on math achievement of children from high-earning families appear to be more pronounced.

1.5 Conclusion

This paper extensively analyses the effects of class size across grade levels in compulsory school. Previous studies are primarily concerned with class size effects of close grade levels, therefore little evidence of how the class size effects behave across

[13] The 'highest earnings' variable is defined as the highest earnings of the pupil's mother and father. When parents are divorced, the income of the mother is used.

Table 1.5: 2SLS estimates, heterogeneous effects of class size

Outcome variable	(1) Gender Main	Interact. (*female*)	(2) Immigrant status Main	Interact. (*non-Western*)	(3) Highest earnings quartile Interact. $(1^{st}\,Q)$	Main	Interact. $(4^{th}\,Q)$
2^{nd} grade reading scores [N = 31,061]	−.0061 (.0086)	−.0000 (.0043)	−.0085 (.0079)	.0348*** (.0127)	.0053 (.0083)	−.0099 (.0084)	.0001 (.0043)
3^{rd} grade math scores [N = 33,404]	−.0244 (.0203)	.0003 (.0039)	−.0236 (.0185)	.0111 (.0220)	.0139* (.0072)	−.0278 (.0201)	.0107 (.0080)
6^{th} grade reading scores [N = 31,542]	−.0254* (.0141)	−.0032 (.0043)	−.0268** (.0130)	.0035 (.0190)	−.0001 (.0092)	−.0254** (.0131)	−.0019 (.0050)
6^{th} grade math scores [N = 31,543]	−.0262* (.0144)	.0044 (.0049)	−.0231* (.0140)	.0042 (.0101)	.0040 (.0049)	−.0203 (.0142)	−.0156** (.0077)
8^{th} grade reading scores [N = 30,424]	.0200 (.0189)	−.0074 (.0058)	.0205 (.0152)	.0163 (.0167)	.0157* (.0089)	.0095 (.0183)	.0154 (.0106)
8^{th} grade physics scores [N = 30,086]	−.0043 (.0139)	−.0101 (.0076)	−.0078 (.0127)	.0021 (.0172)	−.0014 (.0105)	−.0079 (.0129)	−.0021 (.0075)

Notes. Estimates are based on schools with a grade enrolment of ±4 pupils within the three lower thresholds. Regressions include controls from Table A.1 and segment fixed effects interacted with interaction terms. The instruments are interacted with interaction terms as well. Standard errors adjusted for clustering by grade enrolment are in parentheses, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

grades. To gain insight, I employ a well-known fuzzy RD design approach exploiting exogenous variation in class sizes based on a government-imposed class size cap. The results are based on administrative data of all Danish pupils in public compulsory schools and reveal significantly negative (albeit modest) impacts of class size increases at the primary school level but not at the lower secondary level. Thus, the findings suggest that marginal class size increases in grade 8 may not be harmful to the learning environment, whereas pupils in grades 2 and 6 may in particular benefit from a class size decrease. However, the beneficial impact is modest in absolute values (around .01 SD). As such, other initiatives, for example introducing a second teacher in the classroom or increasing instruction time of key subjects alone or combined, may be more cost-effective compared to mere class size reductions. Furthermore, larger class sizes do not seem to increase inequality; girls, non-Western immigrants and socioeconomically disadvantaged pupils are not more adversely affected than other pupils.

Acknowledgements

The author thanks Helena Skyt Nielsen, Nina Smith, Maria Knoth Humlum, seminar participants at Aarhus University, participants at the 2014 IWAEE, Catanzaro, Italy, and especially Colin Green and two anonymous referees for very helpful discussions and comments. Financial support from the Danish Council for Strategic Research [CSER, 09-070295] is gratefully acknowledged.

1.6 References

Angrist, Joshua, and Guido Imbens. 1995. Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity. *Journal of the American Statistical Association* 90 (430): 431–442.

Angrist, Joshua, and Victor Lavy. 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics* 114 (2): 533–575.

Beuchert, Louise, and Anne Nandrup. 2014. The National Tests—A Practical Guide. *Economics Working Paper* No. 2014–25, Aarhus University.

Blatchford, Peter, and Peter Mortimore. 1994. 'The Issue of Class Size for Young Children in Schools: What Can We Learn from Research? *Oxford Review of Education* 20 (4): 411–428.

Finn, Jeremy, and Charles Achilles. 1999. Tennessee's Class Size Study: Findings, Implications, Misconceptions. *Educational Evaluation and Policy Analysis* 21 (2, Summer Special Issue: *Class Size: Issues and New Findings*): 97–109.

Fredriksson, Peter, Bjørn Öckert, and Hessel Oosterbeek. 2013. Long-term Effects of Class Size. *The Quarterly Journal of Economics* 128 (1): 249–250.

Heinesen, Eskil, and Martin Browning. 2007. Class Size, Teacher Hours and Educational Attainment. *Scandinavian Journal of Economics* 109 (2): 415–438.

Hoxby, Caroline. 2000. The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics* 115 (4): 1239–1285.

Krassel, Karl, and Eskil Heinesen. 2014. Class-size Effects in Secondary School. *Education Economics* 22 (4): 412–426.

Krueger, Alan. 1999. Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114: 497–537.

Lee, David, and David Card. 2008. Regression Discontinuity Inference with Specification Error. *Journal of Econometrics* 142 (2): 655–674.

Lee, David, and Thomas Lemieux. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48 (2): 281–355.

Leuven, Edwin, Hessel Oosterbeek and Marte Rønning. 2008. Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway." *The Scandinavian Journal of Economics* 110 (4): 663–693.

Mischel, Harriet, and Walter Mischel. 1983. The Development of Children's Knowledge of Self-Control Strategies. *Child Development* 54: 613–619.

OECD. 2012. Education at a Glance 2012: Highlights. *OECD Publishing*, Paris. doi: 10.1787/ eag_highlights-2012-en.

Rivkin, Steven, Eric Hanushek and John Kain. 2005. Teachers, Schools and Academic Achievement. *Econometrica* 73 (2): 417–458.

Todd, Petra, and Kenneth Wolpin. 2003. On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal* 113 (485): F3–F33.

Urquiola, Miguel, and Eric Verhoogen. 2009. Class-Size Caps, Sorting, and the Regression-Discontinuity Design. *The American Economic Review* 99 (1, June): 179–215.

Appendix

Table A.1: Sample means and OLS regression of outcome on controls

	Mean	SD	OLS Coeff.	OLS SE
Outcome				
Average standardized test result	.027	.971		
Instruments				
Above 28-threshold	.198			
Above 56-threshold	.207			
Above 84-threshold	.061			
Above 112-threshold	.006			
Above 140-threshold	.001			
Controls				
Class size	21.468	3.854		
Enrolment	53.689	22.430		
Mother's education				
—None/missing	.045			
—Basic	.257		0.002	(0.010)
—Vocational	.367		0.055***	(0.010)
—Higher	.331		0.315***	(0.010)
Father's education				
—None/missing	.075			
—Basic	.236		0.023**	(0.009)
—Vocational	.410		0.076***	(0.010)
—Higher	.278		0.342***	(0.009)
Mother's logearnings	9.843	4.738	0.006***	(0.000)
Mother's age	34.763	7.445	0.014***	(0.000)
Father's logearnings	10.189	4.901	0.005***	(0.000)
Father's age	36.160	10.469	0.004***	(0.000)
Single mother	.155		−0.084***	(0.003)
Number of siblings	1.262	.866	0.011***	(0.002)
Girl	.492		0.032***	(0.003)
Western immigrant (or descendant hereof)	.020		0.016*	(0.009)
Non-Western immigrant (or descendant hereof)	.099		−0.324***	(0.006)
Birth weight (g)	3298.906	1006.609	0.000***	(0.000)
Length of gestation (days)	199.637	126.030	−0.000***	(0.000)
Born in the first quarter of the year	.242			
—second quarter	.252		−0.040***	(0.003)
—third quarter	.262		−0.090***	(0.003)
—fourth quarter	.235		−0.070***	(0.003)
First-born	.420		0.288***	(0.004)
Second-born	.371		0.125***	(0.003)
Born third or later	.209			
Multiple-born	.038		0.074***	(0.007)
Age indicators (*omitted here*)	–		–	
Tested in 2010	.316		0.006	(0.006)
Tested in 2011	.342		0.005	(0.006)
Tested in 2012	.342			
Municipality with smaller cities (below 10,000)	.169			
—with a large city	.518		0.027***	(0.008)
—in the capital area	.313		0.027***	(0.008)
Observations	893,434		893,434	
R^2			0.12	

Notes. In addition to the control variables listed in the table, the regression includes segment fixed effects and linear and squared controls for grade enrolment into schools interacted with segments. Standard errors adjusted for clustering by grade enrolment are in parentheses, $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A.2: Balancing of covariates

Baseline covariate	(1) All pupils Coef., *above* cutoff	(1) All pupils p-value	(2) ±4 pupils around lower cutoffs Coef., *above* cutoff	(2) ±4 pupils around lower cutoffs p-value
Mother's education:				
—Basic	0.002	0.026	0.004	0.488
—Vocational	0.001	0.148	0.003	0.488
—Higher	−0.005	0.000	−0.009	0.137
Father's education:				
—Basic	0.002	0.012	0.005	0.272
—Vocational	0.000	0.713	−0.006	0.040
—Higher	−0.004	0.000	−0.003	0.614
Mother's log-earnings	−0.000	0.000	−0.001	0.187
Mother's age	−0.000	0.000	−0.000	0.082
Father's log-earnings	−0.000	0.006	−0.000	0.645
Father's age	−0.000	0.001	−0.000	0.088
Single mother	0.003	0.012	0.010	0.138
Number of siblings	0.001	0.013	0.000	0.913
Female	0.000	0.879	0.001	0.574
Western immigrant	−0.004	0.065	−0.005	0.593
Non-Western immigrant	0.009	0.002	0.026	0.075
Birth weight	−0.000	0.161	−0.000	0.275
Gestational age	−0.000	0.040	−0.000	0.017
First-born	−0.001	0.055	0.002	0.358
Second-born	−0.000	0.848	0.001	0.781
Multiple born	−0.001	0.639	0.003	0.817
No. of observations	893,434		188,061	
p-value for F-test of joint significance of baseline variables on being *above* cutoffs:				
All subsamples	0.000		0.195	
—2^{nd} grade reading	0.078		0.366	
—3^{rd} grade math	0.181		0.010	
—6^{th} grade reading	0.199		0.870	
—6^{th} grade math	0.054		0.725	
—8^{th} grade reading	0.541		0.342	
—8^{th} grade physics	0.170		0.117	

Notes. The *above* cutoff indicator equals 1 if the grade enrolment exceeds a threshold created by the 28-pupil rule up to +14 pupils (+4 pupils around the three lower cutoffs in (2)). In the top panel, columns report coefficients (left) and p-values (right) for t-tests of the significance of the pooled instrument by separate OLS regressions of each variable listed on the instrument. The bottom panel reports the p-values for F-tests of the joint significance of the variables listed in the table. All regressions additionally include year and enrolment segment fixed effects and indicator variables of degree of urbanization of the school municipality. Columns (1) further include linear and square controls for grade enrolment interacted with separate thresholds. Standard errors are adjusted for clustering by enrolment count.

Table A.3: Reduced form and 2SLS estimates, ±4 pupils around three lower thresholds

Model	(1) 2^{nd} grade reading	(2) 3^{rd} grade math	(3) 6^{th} grade reading	(4) 6^{th} grade math	(5) 8^{th} grade reading	(6) 8^{th} grade physics
First stage estimates, outcome: class size						
*Above*28	−6.6192***	−6.1054***	−6.6986***	−6.6018***	−5.9147***	−6.1237***
	(0.5282)	(0.8146)	(0.7159)	(0.7246)	(0.6837)	(0.6796)
*Above*56	−1.5923***	−2.6370***	−2.0697***	−2.0914***	−1.7183***	−1.7269***
	(0.6150)	(0.5284)	(0.5686)	(0.5574)	(0.4964)	(0.4671)
*Above*84	−0.3883	−1.0630**	−1.1935***	−1.2566***	0.2638	0.1334
	(0.8059)	(0.5380)	(0.3904)	(0.3476)	(0.4463)	(0.4656)
Second stage (2SLS) estimates, outcome: pupil test scores						
Class size	−0.0101***	−0.0059	−0.0125***	−0.0149***	−0.0078	−0.0014
	(0.0036)	(0.0040)	(0.0037)	(0.0041)	(0.0067)	(0.0048)
F-test (excl. instruments)	53.77	28.66	35.80	35.73	29.36	31.96
Reduced form estimates, outcome: pupil test scores						
*Above*28	0.0674**	0.0909***	0.0108	0.0338	0.0937***	−0.0038
	(0.0269)	(0.0200)	(0.0423)	(0.0257)	(0.0292)	(0.0301)
*Above*56	0.0154	0.0117	0.0581***	0.0197	0.0363	0.0199
	(0.0362)	(0.0333)	(0.0203)	(0.0230)	(0.0438)	(0.0277)
*Above*84	−0.0164	−0.0088	−0.0848**	0.0005	0.0487	−0.0179
	(0.0371)	(0.0471)	(0.0399)	(0.0489)	(0.0301)	(0.0447)
Observations	31,061	33,404	31,543	31,543	30,424	30,086

Notes. The estimates are based on mainstream classroom pupils in public schools with a grade enrolment of ±4 pupils within the three lower thresholds in 2009/2010–2011/2012. *Above* threshold indicators (instruments for class size) equal 1 if school grade enrolment exceeds 28, 56 or 84, respectively. In the reduced form estimations are outcomes regressed on the instruments directly. Regressions include controls from Table A.1 and segment and year fixed effects. Standard errors adjusted for clustering by enrolment count are in parentheses, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4: Auxiliary results: Specification analysis and robustness checks

Outcome variable	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS, 3yr class size	(5) 2SLS, rule abiding
2^{nd} grade reading score	−.0010	−.0094***	−.0088***	–	−.0105**
	(.0014)	(.0030)	(.0029)		(.0049)
F-test (excl. instruments)	–	94.60	51.86		19.53
No. of observations		150,065			15,736
3^{rd} grade math score	−.0010	−.0065**	−.0065**	−.0127	−.0046
	(.0015)	(.0029)	(.0029)	(.0085)	(.0058)
F-test (excl. instruments)	–	92.77	25.43	17.12	20.97
No. of observations		152,800		32,231	16,753
6^{th} grade reading score	.0033***	−.0043	−.0040	−.0073	−.0189***
	(.0012)	(.0031)	(.0030)	(.0047)	(.0045)
F-test (excl. instruments)	–	82.74	26.23	21.83	20.99
No. of observations		153,810		30,810	16,793
6^{th} grade math score	.0013	−.0097***	−.0087***	−.0127**	−.0242***
	(.0017)	(.0031)	(.0031)	(.0057)	(.0046)
F-test (excl. instruments)	–	82.54	25.12	21.44	21.45
No. of observations		153,846		30,809	16,710
8^{th} grade reading score	.0056***	−.0011	−.0019	−.0048	−.0068
	(.0014)	(.0058)	(.0053)	(.0124)	(.0050)
F-test (excl. instruments)	–	30.86	10.15	7.91	35.70
No. of observations		141,938		29,601	15,254
8^{th} grade	.0024	−.0008	.0008	−.0065	−.0061
physic/chemistry score	(.0015)	(.0054)	(.0049)	(.0149)	(.0080)
F-test (excl. instruments)	–	33.05	10.35	7.95	35.97
No. of observations		140,975		29,275	15,031
Instrument					
Expected class size (f_{gs})		√			
Binary *above* indicators			√	√	√
Enrolment specifications:					
2^{nd} order polynomial	√	√	√		
Interacted w/ segments	√				
±4 pupils around three lower cutoffs				√	√
Full set of controls	√	√	√	√	√

Notes. The estimates are based on pupils in mainstream classrooms in the Danish public schools in 2009/2010–2011/2012. Columns (1)–(3) serve as a specification analysis; column (1) presents the results from a standard OLS regression of eq. (1.2), columns (2) and (3) present results from 2SLS specifications. In (2) the expected class size function from eq. (1.1) is used as the instrument while including a second order polynomial of enrolment but without interacting with segments. The same enrolment specification is used in (3), only class size is instrumented by the binary indicators presented in eq. (1.3). The robustness checks in specifications (4) and (5) only include pupils enrolled in schools with a grade enrolment of ±4 around the three lower cutoffs: 28, 56, and 84. In (4) enrolment and instruments are based on enrolment two years before the outcome is observed. Class size is a three-year average. (5) includes only municipalities that likely abide by the 28-pupil rule. In addition to the enrolment specifications listed in the table, all regressions include segment fixed effects. All specifications include a full set of controls (see Table A.1). Standard errors adjusted for clustering by grade enrolment are in parentheses, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Chapter 2

Increasing Instruction Time in School Does Increase Learning

Published in Proceedings of the National Academy of Science 113 (27) (2016), 7481–7484

Simon Calmar Andersen
Aarhus University and TrygFonden's Centre for Child Research

Maria Knoth Humlum
Aarhus University

Anne Brink Nandrup
Aarhus University

Abstract

Increasing instruction time in school is a central element in the attempts of many governments to improve student learning, but prior research—mainly based on observational data—disputes the effect of this approach and points out the potential negative effects on student behavior. Based on a large-scale, cluster-randomized trial, we find that increasing instruction time increases student learning and that a general increase in instruction time is at least as efficient as an expert-developed, detailed teaching program that increases instruction with the same amount of time. These findings support the value of increased instruction time.

Keywords: Education; Randomized controlled trial; School performance; School resources

2.1 Introduction

All governments responsible for school systems must consider the amount of instruction time that they should provide. The time that students spend in the classroom varies by a factor of two across the OECD (Organization of Economic Co-operation and Development) countries both in total compulsory instruction time and within specific subjects, such as reading, writing, and literature (OECD 2014). These international differences have generated sustained debates about whether students benefit from having more instruction or on the contrary, whether governments can cut spending on instruction time without negatively impacting student achievements (Patall et al. 2010). Increased instruction time has been an element in many educational reforms in the United States, Europe, and Japan (Patall et al. 2010, Meyer and van Klaveren 2013, Kikuchi 2014).

However, the existing evidence of the effectiveness of increasing instruction time is deficient. A review of the research before 2009 concludes that there seems to be a neutral to small positive effect of extending school time on achievements; however, most studies are based on weak designs, and the effect remains disputed. Skeptics argue that longer school days generate behavioral problems caused by fatigue and boredom (Patall et al. 2010). More recently, studies based on observational data found positive effects on student achievements (Kikuchi 2014, Jensen 2013, Parinduri 2014), and studies of the impact of increased instruction time in combination with other interventions (e.g., more effective teachers, data-driven instruction, ability tracking, or improved pedagogy) also found positive effects (Fryer 2014, Cortes and Goodman 2014). A randomized trial conducted in The Netherlands does not find significant effects of increased instruction time (Meyer and van Klaveren 2013). Nonetheless, it should be noted that this trial had substantial noncompliance and was based on only seven schools, which seems low-powered (Schochet 2008).

Other than the methodological limitations of the existing evidence, there are two potential explanations for the lack of strong evidence for the effect of increasing instruction time. One explanation relates to the students, and the other relates to the teachers. First, to benefit from more instruction time, students may need to be motivated to sacrifice short-term pleasures to pay attention to the teaching and thereby, achieve long-term gains (Duckworth and Seligman 2005). This exercise, however, requires self-control. Self-control has been shown to be a scarce resource that is exhausted when used. When that happens, it is harder for students to control their thoughts, fix their attention, and manage their emotions, and they may become more aggressive (Baumeister et al. 2007). Thus, extending the school day may be ineffective, because students' self-control is depleted, and they may have more trouble managing their emotions, become more aggressive, become hyperactive, and/or conflict with their classmates. Furthermore, previous research has shown that boys have less self-control capacity than girls (Duckworth and Seligman 2006), and immigrant children and children with low socioeconomic status also tend to have less self-control (Raver 2004). [This finding does not imply that immigrants have less self-control or lower academic achievement, because they have a different cultural background. Their achievement may be strongly related to the lower socioeconomic status of immigrants on average. Also, there may be significant heterogeneity among immigrants with non-Western backgrounds.] More formally, children can be expected to maximize their learning in school relative to the effort that it requires. Because the cost of effort as well as the relationship between effort and learning may be different for boys and girls as well as immigrants and natives, it is worth examining the effect of increasing instruction time separately for each of these groups, even if the power of the study does not allow strong conclusions based on subgroup analyses.

Second, the effect of increasing school resources is likely to depend on how teachers spend the additional time, which relates to the instructional regime in the school (that is, the set of rules for how to regulate the interplay between assessment and instruction) (Raudenbusch 2008). We compare two opposite instructional regimes. One type has formalized instruction in a teaching program. On the one hand, this format may have several advantages. If teachers do not know how to teach effectively or if they are satisfied with some level of student achievements and therefore, not motivated to use the additional time effectively, providing a detailed teaching program for the additional instruction time may help increase the effectiveness of this time. On the other hand, this instructional regime with a high level of formalized instruction leaves less room for assessment of the individual student's responsiveness to the intervention. Therefore, it may not tailor the instruction sufficiently to the needs of the different groups of students. We compare this regime with another type that has no formalized prescriptions for how the instruction should take place. This high discretion treatment leaves more room for individual assessments and thereby, more tailored instruction. This tradeoff between high discretion, allowing frontline bureaucrats to use their expertise, and low discretion, ensuring a more specific policy implementation, is a classic but topical dilemma (Weber 1922, Bawn 1995, Huber and Shipan 2002). However, there is very little evidence on whether high or low discretion affects policy outcomes (Carpenter et al. 2012).

2.2 Results

To *(i)* improve the methodological quality of the evidence on increasing instruction time, *(ii)* compare two different instructional regimes on how to regulate the use of additional time, and *(iii)* compare how they affect different groups of students in terms of reading skills and behavioral problems that may come with depleted self-control, we use a large-scale, cluster-randomized trial involving 90 schools and 1,931 fourth grade students in Denmark. Instruction time in reading, writing, and literature was increased by 3 hours (four lessons) weekly over 16 weeks, corresponding to a 15% increase in the weekly instruction time (correspondingly reducing the students' spare time). The cost was approximately US $182 per student. In the first treatment condition, there were no requirements in terms of how teachers should spend the extra time. This instructional regime with high levels of teacher discretion allows teachers to accommodate their teaching to the specific needs of the students in the classroom across a broad range of outcomes. Conversely, a more detailed, expert-developed teaching program may better ensure high-quality teaching. In the second treatment condition, teachers had the same increase in instruction time but were required to follow a detailed program developed by national experts and aimed at improving general language comprehension. We compare the two treatments with a control group continuing with the same instruction time as usual. SI.A Methods, has further details about the treatments.

To measure student achievement in reading, we use a national standardized, online, self-scoring adaptive reading test used by all of the schools in the country (Wandall 2011). The test is based on three subscales: language comprehension, decoding, and reading comprehension. The study was designed to test the effect on the combined measure, but we also examine the effects on the three subscales. Different versions of the test are developed for second and fourth grades. We use the results of the second grade test as a baseline test (along with a third grade math test) and the fourth grade test as an outcome measure. To measure behavioral problems, we use the student responses to the Strengths and Difficulties Questionnaire (SDQ), which is based on five subscales (emotional symptoms, conduct problems, peer relationship problems, hyperactivity/inattention, and prosocial behavior), of which the first four can be combined for a total difficulty score (Goodman 1997, Goodman and Goodman 2009). The SDQ responses are used as a second outcome (SI Methods has details about the assessments). Outcomes were measured at the end of the intervention period (see Fig. S.1).

We find no significant differences in means of the baseline covariates between the control group and the treatment group without a teaching program. A formal F test of the null hypothesis that all baseline covariates are the same for the no teaching program treatment and the control group is not rejected ($p = 0.71$). The teaching program treatment group differs significantly from the control group on two baseline covariates (of 28), and the F test rejects the null hypothesis that all baseline covariates are the same ($p < 0.01$). SI.A Methods and Table S.1 has more details about the baseline balance. Table S.2 presents a formal attrition analysis.

The estimated treatment effects on reading are presented in Fig. 2.1, Panel A. Increasing instruction time without a teaching program increases student achievement in reading by 0.15 SD ($p = 0.02$) compared with the control group. The effect of increasing instruction time with a teaching program is small and insignificant. However, the estimated effects on the reading subscales suggest that the teaching program (aimed at improving general language comprehension) has a statistically significant effect on the subscale language comprehension of 0.14 SD ($p = 0.03$), which would be expected, but not on the other two subscales (Table S.3). The general increase in instruction time with no teaching program has significantly positive effects on both language comprehension and decoding. The differences between the two treatment groups are not statistically significant. It should be noted, however, that the statistical power of the trial does not allow us to detect minor differences between the treatment groups.

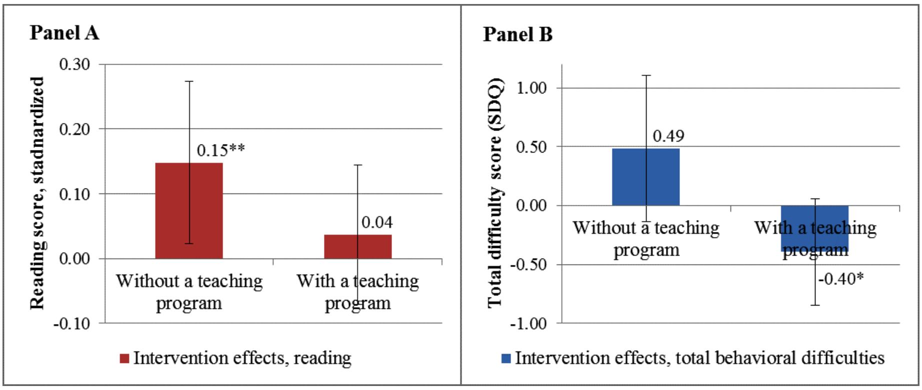

Figure 2.1: Effects of increasing instruction time.
This figure illustrates the effects of increasing instruction time on *(A)* student achievement in reading and *(B)* behavioral difficulties. Stratum indicators and baseline achievement in reading and math are included as controls. Error bars reflect 95% confidence intervals. SEs are corrected for clustering at the classroom level. $*p < 0.1$; $**p < 0.05$.

Fig. 2.1, Panel B shows that the teaching program significantly reduces student behavioral problems. Instruction time without a teaching program may increase behavioral problems, but the overall results are insignificant (see Table S.4 for details).

The study is not powered for multiple tests of effects on groups, but because of the concerns that increasing instruction time will be less beneficial for students with less capacity for selfcontrol, it is worth examining results for subgroups—although it should be emphasized that these results are exploratory. Subgroup analyses presented in Table S.5 suggest that students of non-Western origin do not seem to benefit from any of the two interventions. Increased instruction time without a teaching program may cause increased behavioral difficulties for boys, whereas it has no effect on girls. Conversely, increased instruction time with a teaching program seems to reduce behavioral difficulties for girls.

2.3 Discussion

These results suggest that increasing instruction time does increase average student achievement. An effect of 0.15 SD in reading means that a student at the median of the reading score distribution moves to the 44th percentile. This effect is substantial compared with the relatively cheap intervention. The effect size per US $1,000 per student would be approximately 0.82 SD. The results suggest that governments cannot reduce instruction time without the risk of adversely affecting student achievement, but this interpretation naturally depends on the generalizability of the results. First, for reasons explained above, the marginal effect of instruction time may be decreasing. OECD figures show that, in Denmark, where this study took place, instruction time in reading, writing, and literature at age 10 (which is the median age in this study) is above the average of OECD countries (Fig. S.2). The total compulsory instruction time is very close to the OECD average (OECD 2014). If marginal returns to instruction time are decreasing, the many countries that provide the same or less instruction time than Denmark would have at least as large of effects. The effects for countries providing more instruction time are more uncertain.

Second, the effects of increasing instruction time may depend on the instructional regime (Raudenbush 2008). Decision-makers face a tradeoff between a high-discretion program allowing teachers to use their assessment of the individual students to tailor instruction vs. a more detailed teaching program that—based on the best available evidence—regulates the instruction more. The latter primarily affected one subscale of reading. The benefits of a detailed, expertdeveloped program may be outweighed by a narrower focus on a specific learning domain, and more detailed instructional regulation curtails teachers' opportunities to differentiate their teaching to the needs of different students in the classroom. It is worth emphasizing that implementation survey data suggest that teachers generally had a positive attitude toward the teaching program: 96% of teachers reported that they used the material to some or a large extent, 83% of the teachers found that the teaching program was useful, and 88% believed that it was beneficial for the whole class.

The high-discretion instructional regime with no teaching program had a significant average treatment effect. This finding does not prove that high-discretion programs will be better than a more detailed, evidence-based program. However, it does suggest that there are some benefits of giving teachers good opportunities to differentiate their instruction. Survey data show that 90% of the teachers report that they used (parts of) the increased instruction time for working more with existing materials, and at the same time, 90% used new materials, which supports the notion that they use the high-discretion regime to accommodate their teaching. The effect of increasing instruction time may also depend on other factors, such as the educational level of the teachers or other available school resources. No single study will, therefore, settle the debate. However, the very low regulated treatment tested in the no teaching program condition makes the results relatively applicable to other contexts.

However, the exploratory results of the high-discretion no teaching program raise two concerns. First, boys did benefit from the intervention in terms of their reading skills, but they may also have experienced increased behavioral problems. Boys have been found to have less self-control (Duckworth and Seligman 2006), and therefore, making them work longer during the day may exhaust their self-control and thereby, create behavioral problems. Second, non-Western students seemed to show no or very little benefit of the intervention. Nevertheless, 73% of the teachers report that they also believed that the intervention benefitted bilingual students. This issue points to the other aspect of the instructional regime than the instruction, namely the assessment of the students' progress. Teachers did not seem to notice if the non-Western students did not benefit from the instruction. Therefore, it might be that the effect of a high-discretion regime on the instruction side would be even more effective combined with more regulation on the assessment side, thereby making teachers more aware of how their students respond to their teaching.

It should be emphasized that these considerations should be seen as hypotheses for future research, because the power of the trial does not allow strong inference about the differences between the student groups and treatment conditions. The results do confirm, however, that increasing instruction time in an instructional regime with little formalization has positive average treatment effects on the reading skills of the students.

2.4 Methods

Participants. The randomized, controlled trial was approved and funded by the Danish Ministry of Education and Aarhus University. All schools have volunteered to participate in the trial. Parents of students were informed about the content of the trial beforehand and told how to withdraw their child from the trial if they wished. All interventions were implemented for students in grade 4 in the fall of 2013. Danish public schools that expected to have at least 10% bilingual students in grade 4 in the school year 2013/2014 were eligible to participate in the trial.

The participating schools were fully reimbursed for the costs associated with participation in the trial.

Procedure. In March of 2013, the Minister of Education sent an email to all municipalities in Denmark informing them about the upcoming randomized trial, eligibility criteria, and enrollment procedures. The municipalities were invited to enroll all of their eligible schools in the trial; 126 schools enrolled in the trial. We estimate that this constitutes about 37% of the eligible schools.

The trial was a two-stage, cluster-randomized trial with three treatment arms. Fig. S.3 shows a diagram of the flow of schools and students participating in the trial. The two levels of randomization were school and classroom. First, administrative records from the school year of 2011/2012 were used to divide schools into strata based on the share of students of non- Western origin in grade 2 and the average score on the national reading test in grade 2. Each stratum contained four schools, and allocation to one of four experimental conditions (the two treatment arms, the control group, and a third treatment not related to instruction time and not analyzed here) was random within the stratum. Second, one classroom in each school was selected to participate in the trial by simple randomization. Randomization assures an unbiased distribution of baseline characteristics between experimental conditions, although some imbalance will occur in any finite sample. We find no substantially large imbalances between control and treatment groups (among 28 baseline student characteristics reported in Table S.1, none of the mean values were significantly different across the treatment group without a teaching program and the control group, and only two means were significantly different across the treatment group with a teaching program and the control group). However, some minor imbalance occurs in baseline reading achievement between the control group and the teaching program group. To be conservative and because of a strong expected relationship between baseline achievement and outcomes, the effect estimates presented here are controlled for baseline achievement in reading and math. SI.A Methods and SI.B Results has more details about the balance of the experimental conditions and the robustness of the results to different model specifications.

Although the trial included three treatment arms, the focus of this study is the two interventions that involved an increase in instruction time. Thus, we exclude schools that received the third treatment from our analyses. In the interventions that involved an increase in instruction time, the classrooms received four extra (45-min) lessons per week for 16 weeks. In the first treatment arm, classrooms received extra instruction time in Danish with high teacher discretion (i.e., without a teaching program). Thus, the teachers were not provided with any explicit teaching material. In the second treatment arm, classrooms received extra instruction time in Danish with low teacher discretion (i.e., teachers were provided with very detailed teaching material containing texts and classroom exercises for each week of the intervention).

Analysis. We estimate each of the two treatment effects separately (i.e., including only observations in the relevant treatment arm and the control group). The empirical analysis presented is based on a linear regression model that includes a treatment indicator, stratum fixed effects, and baseline test scores. The stratum fixed effects are included to take into account that treatment assignment was random within strata. We account for the hierarchical structure of the data (students within classrooms) by clustering SEs at the classroom level. Similar results are found using hierarchical linear modeling (Table S.6), which would be expected based on simulations comparing hierarchical linear or multilevel models with models using clustered SEs (Green and Vavreck 2008). Because of attrition, outcome data are missing for some students. Test score attrition does not correlate with the treatment assignment, but participants in the control group were less likely to respond to the postintervention survey containing the SDQ outcomes (Table S.2). We adjust for baseline achievement to present the more conservative estimates. Effect sizes are generally larger when we do not include baseline achievement (Table S.3). We estimate the intention to treat effect, which is an estimate of the effect of assigning students to increased instruction time that does not impose assumptions about noncompliance. The intention to treat effect is of immediate relevance to policymakers, because it reflects the average treatment effect taking into account that not all students will comply with a policy that increases instruction time. For instance, students may transfer to a private school or other schools with no increase in instruction time. Additional details about the trial and the analyses are in SI.A Methods and SI.B Results.

Acknowledgements

We thank Michael Rosholm for supporting the setup of the project, Rambøll for data collection assistance and support, and Jessica Goldberg for helpful comments and discussions. The implementation and evaluation of the randomized trial were funded by the Danish Ministry of Education.

2.5 References

Altman D. 1985. Comparability of randomized groups. *Statistician* 34 (1): 125–136.

Baumeister RF, Vohs KD, Tice DM. 2007. The strength model of self-control. *Curr Dir Psychol Sci* 16 (6): 351–355.

Bawn K. 1995. Political control versus expertise: Congressional choices about administrative procedures. *Am Polit Sci Rev* 89 (1): 62–73.

Bruhn M, McKenzie D. 2009. In pursuit of balance: Randomization in practice in development field experiments. *Am Econ J Appl Econ* 1 (4): 200–232.

Carpenter D, Chattopadhyay J, Moffitt S, Nall C. 2012. The complications of controlling agency time discretion: FDA review deadlines and postmarket drug safety. *Am J Pol Sci* 56 (1): 98–114.

Cortes KE, Goodman JS. 2014. Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement. *Am Econ Rev* 104 (5): 400–405.

Duckworth AL, Seligman MEP. 2005. Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol Sci* 16 (12): 939–944.

Duckworth AL, Seligman MEP. 2006. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *J Educ Psychol* 98 (1): 198–208.

Fryer RG. 2014. Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Q J Econ* 129 (3): 1355–1407.

Goodman R. 1997. The Strengths and Difficulties Questionnaire: A research note. *J Child Psychol Psychiatry* 38 (5): 581–586.

Goodman A, Goodman R. 2009. Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J Am Acad Child Adolesc Psychiatry* 48 (4): 400–403.

Goodman R, Meltzer H, Bailey V. 1998. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *Eur Child Adolesc Psychiatry* 7 (3): 125–130.

Green DP, Vavreck L. 2008. Analysis of cluster-randomized experiments: A comparison of alternative estimation approaches. *Polit Anal* 16 (2): 138–152.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat* 6 (2): 65–70.

Huber JD, Shipan C. 2002. Deliberate Discretion: The Institutional Foundations of Bureaucratic Autonomy. *Cambridge Univ Press*, Cambridge, United Kingdom.

Jensen VM. 2013. Working longer makes students stronger? *Educ Res* 55 (2): 180–194.

Kézdi G. 2004. Robust standard error estimation in fixed-effects panel data models. *Hung Stat Rev* 9: 95–116.

Kikuchi N. 2014. The effect of instructional time reduction on educational attainment. *J Jpn Int Econ* 32: 17–41.

Krueger AB, Zhu P. 2004. Another look at the New York City school voucher experiment. *Am Behav Sci* 47 (5): 658–698.

Meyer E, van Klaveren C. 2013. The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands. *Econ Educ Rev* 36: 1–11.

OECD. 2014. Education at a Glance. *OECD Publishing*, Paris.

Parinduri RA. 2014. Do children spend too much time in schools? Evidence from a longer school year in Indonesia. *Econ Educ Rev* 41: 89–104.

Patall E, Cooper H, Allen A. 2010. Extending the school day or school year. *Rev Educ Res* 80 (3): 401–436.

Puma M, Olsen R, Bell S, Price C. 2009. What to Do When Data Are Missing in Group Randomized Controlled Trials. *National Center for Education Evaluation and Regional Assistance*, US Department of Education, Washington, DC.

Pøhler L, Sørensen S. 2010. Nationale Test og Anden Evaluering af Elevens Læsning. *Dafolo Forlag*, Copenhagen, Denmark, 1st Ed.

Schochet PZ. 2008. Statistical power for random assignment evaluations of education programs. *J Educ Behav Stat* 33 (1): 62–87.

Raver CC. 2004. Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Dev* 75 (2): 346–353.

Raudenbush SW. 2008. Advancing educational policy by advancing research on instruction. *Am Educ Res* J 45 (1): 206–230.

Statistics Denmark. 2014. Statistical Yearbook 2014. *Statistics Denmark*, Copenhagen.

The Danish Ministry of Education. 2010. Minimumstimetallet fra Skoleåret 2010/2011 og Frem. *The Danish Ministry of Education*, Copenhagen.

Wandall J. 2011. National tests in Denmark – CAT as a pedagogic tool. *J Appl Test Tech* 12 (Special Issue): 3.

Weber M. 1922. Wirtschaft und Gesellschaft. *Mohr*, Tuebingen, Germany.

Supporting Information

This supporting information chapter contains supplementary material for the article *Increased Instruction Time in School Does Increase Learning.* SI.A Methods details the intervention procedure, applied measures and collected data. SI.B Results presents the formal results from primary and supplementary analyses; Section B.1 provides results from preliminary raw differences in means, while Section B.2 contains the primary results presented in the main article including heterogeneity analyses for student subpoplations. Section B.3 contrasts our primary findings to results from a two-level model. Finally, Section B.4 presents descriptive evidence of teacher attitudes and time use in the two interventions with increased instruction time.

SI.A Methods

A.1 Procedure

In the autumn of 2013, a large-scale randomized, controlled trial in primary schools in Denmark was implemented. The Danish Ministry of Education funded the trial. A primary concern was to evaluate the interventions that could improve the outcomes—both academic and behavioral—of bilingual students with Danish as a second language. Therefore, only schools with a relatively high concentration of bilingual students were eligible.

Fig. S.1 shows a timeline for the implementation of the trial and the relevant measurements. In March of 2013, all of the municipalities in Denmark received an email from the Minister of Education with information about the randomized trial. The email contained a brief description of the interventions and study design. The municipalities were given a little more than 2 weeks to sign up and informed that all of the schools that expected to have at least 10% bilingual students in grade 4 in the following school year (August to June) were eligible to participate. The municipalities were encouraged to sign up all of their eligible schools, meaning that only public schools were eligible for participation. In 2013, about 81% of all Danish primary and lower secondary school students attended public schools (Statistics Denmark 2014).

Based on administrative records for grade 3 students in the 2012/2013 school year, roughly 332 of 1,209 schools were eligible to participate, 126 of which enrolled in the trial. It is estimated that about 37% of the eligible schools chose to enroll in the trial.

After the enrollment deadline, the randomization was performed. The trial was designed as a two-stage, cluster-randomized trial with three treatment arms (two instruction time treatments with a teaching program (TP) and without a teaching program (NOTP) and a third treatment not related to instruction time and not analyzed here). There were two levels of randomization: school and classroom. It was a block-randomized trial. The enrolled schools were divided into blocks—which we denote "strata"—based on administrative records on the share of students of non-Western origin in grade 2 in the school year 2011/ 2012 and the average performance of their prospective grade 4 students in the national grade 2 reading test. Specifically, schools were first placed in three groups based on their share of non- Western students: 8 schools with a share between 0.55 and 1, 56 schools with a share between 0.19 and 0.55, and, finally, 60 schools with a share between 0 and 0.19. Two schools with less than 10% of non-Western origin were randomly excluded from the analysis. Within each of three groups, the schools were ranked according to their average performance on the reading test and divided into strata accordingly. Each stratum consisted of four schools, and by random assignment, each of these schools was placed in one of three treatment arms or the control group. In each school, one grade 4 classroom was randomly selected for participation in the trial by simple randomization. However, only classrooms that were expected to have at least 10% bilingual students were included in the randomization.

The number of schools participating in the trial ended up being determined by the number of schools that could be recruited: 31 schools were eventually allocated to each of the experimental conditions. The number of participating schools limits the statistical power of the study. Notably, although we are able to detect a significant effect of the treatment with NOTP compared with the control group, we are not able to detect a significant difference between the two treatment groups (because the mean outcome in the other treatment group is higher than in the control group). Because of the limited power of the study, we are not able to conclude whether this lack of difference between the two treatments is caused by the fact that there is no substantial difference or the fact that we cannot estimate the difference precisely enough.

Although the trial consisted of three different interventions, we focus on the two that increased instruction time. We exclude schools that received the third treatment from our main analyses; they are only included as a robustness check in Table S.3.

The Danish Ministry of Education reimbursed the participating schools for all of the costs associated with participation in the trial given that they implemented the intervention that they were allocated and participated in the data collection. All schools (in both the treatment and control conditions) were paid US $352 to have one classroom participate in the data collection. Schools in the treatment conditions were additionally paid US $3,924 to implement the treatment in one classroom.

A smaller number of schools in the three treatment arms participated in case studies during the implementation of the trial. In June of 2013, information meetings were held that provided some practical information for the schools allocated to one of three treatment arms. School representatives were later divided according to their treatment arm and provided with treatment-specific information. On these occasions, school representatives had the opportunity to pose questions related to the implementation of the interventions.

The first intervention was simply extra instruction time in Danish with high teacher discretion (i.e., NOTP). The classrooms in this treatment arm were given four extra (45-min) lessons per week for 16 weeks during the autumn of 2013. If possible, the regular Danish teacher taught the lessons.[1] The teachers did not receive any explicit teaching material, only a 27-page report entitled "Evidence-based teaching strategies—an idea catalogue with evidence-based strategies for the planning of reading instruction of bilingual children." The report was prepared by researchers and national experts in the instruction of bilingual children and based on an extensive, systematic literature review of research on effective methods and teaching strategies for teaching bilingual children with Danish as a second language. The teachers received the material for inspiration but were not required to use it. Data from implementation surveys show that the extent to which the teachers used the report was limited, and only 17% found it useful (SI.B Results, section B.4).

The second intervention was an expert-developed TP coupled with extra instruction time with lower teacher discretion. The classrooms in this treatment group were also given four extra (45-min) lessons per week for 16 weeks during the autumn of 2013. If possible, the regular Danish teacher taught the lessons. The teachers were provided with very detailed teaching material containing different texts and classroom exercises for each week of the intervention. The material was developed and reviewed by national experts in language instruction with the aim of improving student performance—that of bilingual students with Danish as a second language in particular. The schools received four copies of the teaching material and were told not to distribute it during the trial. Teachers were also asked not to use the material in other classrooms than the one selected for participation in the trial. The teachers largely complied with the program. All but one teacher reported that they have used the material to some or great extent (SI.B Results, section B.4).

The control group schools were required to participate in the data collection to the same extent as those in the three treatment arms. The only difference was that the treatment arm schools were also subject to two implementation surveys during the autumn that were aimed at investigating whether schools implemented the interventions as intended. All of the schools were reimbursed for the costs of participating in the data collection.

When the schools enrolled in the trial, they were informed that it would run for two rounds. The first is described here and took place in the fall of 2013. The schools that were allocated to the control group in the first round would be allocated to a treatment arm in the second round. This feature was meant to encourage the control group schools to remain in the trial and may also have reduced any observer effects specific to being allocated to a treatment or control group. Furthermore, differences between the two treatment groups should not be of such observer effects.

In the 2012/2013 school year (the year before implementing the trial), the average

[1] In Denmark, teachers are subject-specific but follow the same classroom across several grades. On average, the teachers in the trial had taught the students for 1.5 years before the trial started.

number of hours of instruction at the grade 4 level was 818 hours, corresponding to 20 hours per week or about 27 lessons. The average number of hours of Danish instruction was 191 hours, corresponding to about five hours per week or about six lessons. The average number of hours of math instruction was 124 hours, corresponding to about three hours per week or about four lessons. The remaining lessons at the grade 4 level would mainly be made up of lessons in English, history, religious studies, science, physical education, and creative subjects. The exact number of lessons can vary across schools and classrooms, but the Danish Ministry of Education sets a minimum for the number of instruction hours (The Danish Ministry of Education 2010). On average, the intervention increased the yearly number of Danish lessons by 25%, whereas the total weekly instruction time was increased by 15% during the intervention. Information from implementation surveys confirms the intervention details.

Increasing the instruction time necessarily occurs at the expense of other spare time activities. Of the participating students, 57% were enrolled in after-school care in the autumn of 2013. The primary function of these care facilities is pedagogical rather than academic. The facilities are heavily subsidized, but the parents cover part of the costs.

Fig. S.2 presents the average yearly instruction time in reading, writing, and literature for 10 year-olds in OECD countries. Denmark is in the upper end of the distribution, markedly above the OECD average. Thus, the intervention increased the instruction time from an already relatively high level of lessons within the subject. Comparing the total instruction time for 10 year-olds in compulsory education, Denmark is slightly below the OECD average, with around 800 hours per year (OECD 2014).

A.2 Measures

The primary outcomes of interest are student achievement in reading and measures of behavioral responses. Academic achievement was assessed using the Danish National Tests, an official, standardized testing system introduced in 2010. Public school students are subject to mandatory reading tests every second year from grade 2 onward. The mandatory testing period is carried out in each spring term (January to April) but with the extra feature that teachers can sign up their classes for up to two voluntary tests at the same level in the fall term before and/or after the mandatory tests. This feature is exploited to evaluate the reading achievements of participating students after the intervention. In 2013, the voluntary test period ran from October 21 to December 13. To ensure that students were only evaluated after the intervention period, participating schools were asked to book the voluntary test slots on the last day of testing, December 13, before other schools were allowed to sign up. We use the results from the corresponding mandatory second grade reading and mandatory third grade math tests as baseline achievement measures.

The National Tests are adaptive, meaning that the difficulty level of the questions is continuously updated to match the estimated ability of the student throughout the test (Raudenbush 2008). Both student ability and difficulty level are measured on a Rasch calibrated logit scale. The test questions are randomly drawn from a large item bank. Although continuously updating the item bank is an important part of adaptive testing (Pøhler and Sørensen 2010), all of the included test items are subject to pilot testing and strict analysis before inclusion in the item bank (Wandall 2011, Pøhler and Sørensen 2010). During the reading test session, the student's reading skills are evaluated within three separate cognitive domains simultaneously: language comprehension, decoding, and reading comprehension (Wandall 2011). The main outcome analyzed is an average standardized measure of the three achievement scores for each student. First, achievement scores are standardized to zero mean and unit variance within each cognitive domain. Second, a simple average is constructed across domains, and third, this measure is standardized once more. Because the analyzed test results are from the voluntary tests, the population of test takers in the fall of 2013 is likely to be a selected sample of students. Therefore, we use the distribution of the 2013 mandatory reading test for fourth graders in the standardizations (i.e., the cohort one year ahead of our intervention sample).[2] Generally, the means of test scores within a specific domain increase slightly over time, whereas the SDs remain more or less constant. We also include analyses based on the standardized achievement measures for the separate domains as outcomes.

To measure student behavior and determine the possible behavioral effects of the intervention, the SDQ was administered to students after the intervention.[3] The SDQ is a validated survey instrument (Goodman 1997, Goodman and Goodman 2009). We administered the one-sided, self-rated SDQ for 11–17 year-olds (Goodman et al. 1998) in an electronic version approved by Youthinmind. Most of the students participating in the trial were actually only 10 years old at the time of the survey, but we faced a variety of tradeoffs in the choice of survey instrument and found this to be the preferred feasible option. The SDQ is a relatively brief, 25-item questionnaire. The questionnaire contains items on both problematic and positive behavior. All items are answered by marking one of three boxes, indicating whether the respondent finds the relevant statement "not true," "somewhat true," or "certainly true." The respondent is asked to consider their situation for the past six months when answering the questionnaire. The items are divided into five different subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. The first four can be combined to a total difficulties score ranging from 0 to 40; higher scores indicate higher levels of difficulty. Prosocial behavior

[2] Baseline test results are obtained by means of mandatory testing and standardized based on their own distributions.

[3] A student survey was also administered at the beginning of the intervention period. Because this survey was conducted after randomization, which is not ideal, we do not use these data as baseline. Auxiliary analyses confirm that the propensity to answer the presurvey differed across treatment conditions.

is evaluated separately on a 0–10 scale, with higher scores indicating better social behavior.

The postsurveys were administered electronically, and the teachers were responsible for ensuring that all students answered the survey. The teachers were encouraged to use a lesson and provide all of the students with access to a computer. They were allowed to help students understand the questionnaire items if they were unsure about wordings. The postintervention surveys were distributed at the end of the intervention period. Participants were asked to respond before December 20.

We conduct analyses both with and without student-specific covariates. The students in the participating classrooms were linked to administrative registers maintained by Statistics Denmark using unique personal identifiers. The data on individuals include information about their birth, country of origin, previous test results, and parental information (Table S.1 shows a full list of covariates).

To avoid confounding factors caused by parental responses to treatment allocation, all of the information related to parents and family structure is measured in the year of the student's sixth birthday, well before the trial. We were unable to match approx. 1% of the participating students' identifiers to register data. Where information on student characteristics and baseline achievement is missing, a dummy variable adjustment approach is used as recommended for group-randomized, controlled trials as recommended by Puma et al. (2009).

A.3 Sample Characteristics, Comparabiliy, and Attrition

One hundred twenty-six schools signed up for the trial. Fig. S.3 presents a flow diagram of the trial. To match the four experimental conditions, 124 schools were included in the stratified randomization. Before the stratified randomization, two schools with less than 10% non-Western students were randomly excluded. These two schools are not included in the analyses; two schools withdrew from the trial before implementation, and 31 schools were allocated to a third treatment group aimed at improving the math teachers' ability to integrate Danish as a second language in the instruction. These schools are not included in the main analyses. Of the remaining 91 participating schools, one school participated fully in the intervention but with another classroom than assigned by the randomization and was, therefore, excluded from the sample, because we cannot measure the outcomes for the intended classroom. On the student level, four students who had been linked to invalid identifiers were also excluded. Finally, 20% of the sample was lost because of missing test scores (16% for SDQ scores).

Descriptive characteristics of students are presented in Table S.1. Here, the comparability of the sample across groups is also shown. The differences in means indicate that the randomization was successful in creating uniform samples across the intervention and control groups.

A few differences are observed across the treatment and control groups, but they are no more than what is to be expected by construction. Students in the treatment group with a TP were significantly less likely to be missing baseline reading scores compared with those in the comparison group. The pretests on student reading achievement are measured before randomization. Moreover, the fathers of students in the TP sample are marginally younger than those of students in the control group. Although the differences inmeans of the baseline achievement scores in math and reading are not significant, evidence from Table S.1 suggests that the TP students may be a slightly positively selected sample compared with the students in the other groups.[4]

In the postintervention surveys, the treated students answered two questions about their participation in the extra lessons. Here, 4% of the students responded that they had formally been exempted from participation in the TP treatment, whereas 11% were formally exempt from participation in the NOTP treatment. Overall, 80% of the students assigned to either treatment reported that they had participated in all lessons. Around 90% reported having been present for more than one-half of the extra lessons. The results of our primary analyses are robust to excluding students who reported not having been present for a considerable fraction of the intervention.

As illustrated in Fig. S.3, 20% of the students did not take the reading test in December, whereas 16% of the students were lost because of attrition in postintervention survey responses. In total, five whole classrooms did not complete the reading tests after the intervention, whereas six did not complete the postintervention survey.[5] Balance analyses across groups suggest that samples and treatment groups remain largely similar in terms of observables. A formal attrition analysis on both outcomes is presented in Table S.2. In specifications 1 and 3 in Table S.2, the probability of not attending the reading test (answering the postintervention survey) is regressed only on indicators for treatment groups and stratum fixed effects. Specifications 2 and 4 in Table S.2 further include available student characteristics as listed in Table S.1. In terms of the postintervention reading test, specifications 1 and

[4]An F test confirms that mean baseline characteristics jointly differ across TP and control groups ($p = 0.00$; p values corrected for clustering at the school level). The F statistic tests the joint hypothesis that the coefficients of all covariates are equal to zero in a model regressing the treatment indicator on all covariates. In other words, if the mean of just one variable is significantly different between the treatment and control groups, the null hypothesis does not hold. Therefore, for the TP group, the significant F test confirms what can also be seen from the balance table, namely that a few variables are significantly related to treatment assignment. However, we fail to reject that differences in the baseline characteristics of the NOTP and control groups are jointly zero ($p = 0.71$; p values corrected for clustering at the school level). Thus, other than not being significantly related to the treatment indicator individually, the covariates are also not jointly significant. It should be noted, however, that small imbalances in variables strongly related to the outcome—such as baseline achievement—are more important to adjust for than statistically significant differences in variables weakly related to the outcome (Altman 1985, Bruhn and McKenzie 2009).

[5]As a consequence of the pairwise (four-way) stratification strategy, when outcomes are missing for an entire school, the other school in the stratum does not contribute to the point estimate of the intervention effect. Thus, the estimation sample is selected on attrition. Estimated effect sizes in models without strata indicators are larger than in the main specification with strata indicators (compared with raw differences in means in Table S.3).

2 in Table S.2 reveal that there are no differences in the attrition rates across control and intervention groups. Disadvantaged students were less likely to attend the test; poorer performance on the second grade baseline reading test, missing the baseline tests, or having a single mother are strong predictors of missing the postintervention reading test. Additional analyses (not presented) suggest that NOTP students who do not attend the test may be marginally disadvantaged in terms of baseline attendance rates and reading scores compared with the control group students.

When considering the self-reported behavioral outcomes in specifications 3 and 4 in Table S.2, students from both treatment groups are significantly less likely to attrite compared with the control students. Again, poorer baseline reading scores, not attending the baseline math test, and having a single mother are all associated with a higher probability of not answering the postintervention survey. We suspect that control group teachers (who are less interested in participating in the data collection, because they did not receive any treatment and may, therefore, not understand the purpose of the data collection) may cause the significantly higher attrition rate of the control group.

Because the trial focuses on students from schools with a considerable percentage of bilingual students with Danish as a second language, the sample is slightly negatively selected in terms of academic achievement, parental characteristics, and so forth compared with the general population of students in public schools. Thus, the distribution of grade 2 reading scores from the school year of 2011/2012 (one of the stratifying variables) of participating schools is to the left of that of the general population of students. Consistent with the participation requirement, 25% of the third grade students from 2012/2013 (who would attend grade 4 in the autumn of 2013) in the participating schools are of non-Western origin; the corresponding fraction is 8% for the average public school. Consequently, students in the participating schools have slightly less advantageous parental backgrounds, except that fewer come from single-mother homes. There were, however, no considerable differences in the instruction time. The average public school had slightly less yearly instruction time in Danish (190 hours) than the participating schools (193 hours) but with a smaller variance.

Auxiliary analyses show that, compared with other eligible public schools in Denmark, there are no significant differences in the baseline scores of the students or the yearly instruction time in Danish (193 hours for both). Still, there are significantly more students of non-Western origin in the participating schools compared with in the eligible ones (on average 25% vs. 20%), and their mothers are generally slightly less educated.

SI.B Results

B.1 Preliminary Analyses

Our preliminary analyses estimate the raw intervention effects as measured by the difference in the raw means between the two treatment groups and the control group. The results presented in columns 1 and 2 in Table S.3 suggest that increasing instruction time has a significant positive effect of more than 0.2 SDs ($p = 0.02$ and $p = 0.06$, respectively) on student achievement for both interventions, whereas the effects on total behavioral difficulties are inconclusive (columns 1 and 2 in Table S.4). These differences in raw means for reading achievement are somewhat larger than the results of our primary analyses (SI.B Results, section B.2), where stratum fixed effects are included in the models. Including stratum indicators reduces the estimated effect sizes to around 0.15 SD ($p = 0.06$ and $p = 0.01$, respectively), presumably reflecting a slight imbalance in terms of unequal proportions of treatment and control students across blocks. Although the ex ante probability of being assigned to an experimental condition is the same across strata, differing class sizes cause the ex post proportions of students in these conditions to differ. However, including stratum fixed effects ensures that only similar schools in terms of the stratification variables (baseline achievement in reading and share of non-Western students at the grade level) are compared in the analysis, thus more accurately—and more conservatively—depicting the true intervention effects. Lastly, we cannot reject that the intervention effects are the same when estimated with and without stratum indicators ($p = 0.24$ NOTP; $p = 0.62$ with TP). We, therefore, include stratum indicators in our primary analyses as suggested in the works in Bruhn and McKenzie (2009) and Krueger and Zhu (2004).

B.2 Primary Analyses

In our primary analyses, we estimate the effects of the interventions using ordinary least squares regression with SEs corrected for clustering within schools (in SI.B Results, section B.3, we also present evidence of the robustness of our results by reproducing them in a multilevel model to account for the nested structure of the data). The model specifies that student achievement is a function of baseline achievement and other student-specific characteristics. Thus,

$$Y_{ijk} = \alpha_0 + Treatment_j \alpha_1 + \mathbf{PreTest_i} \boldsymbol{\alpha_2} + \mathbf{x_i} \boldsymbol{\alpha_3} + Stratum_k + u_{ijk}, \tag{S.1}$$

where Y_{ijk} denotes the outcome of interest for student i in school j of stratum k. The model is estimated separately for the two treatment groups to increase flexibility (i.e., each treatment group is compared with the control group separately)[6]; α_1

[6] We fail to reject that a pooled specification, where indicators of both treatments are included simultaneously, produces different intervention effects. Eq. S.1, however, has the advantage that it is less sensitive to potential imbalances in baseline covariates that may occur by chance across experimental conditions.

represents the intention to treat (ITT) effect of the relevant treatment. The primary outcomes are student academic achievement in reading and total behavioral difficulties (SDQ). All models include stratum fixed effects ($Stratum_k$). The stratum fixed effects are parameterized by including a dummy variable for each stratum (i.e., 30 stratum indicators). We estimate the above model both including and excluding baseline achievement and other baseline student characteristics. The vector $\mathbf{PreTest_i}$ contains the baseline achievement in reading (grade 2) and math (grade 3) as well as corresponding missing indicators. The vector $\mathbf{x_i}$ includes child covariates (indicators for student age, quarter of birth, gender, birth order, being of non-Western origin[7] and having a single mother), parental covariates (age, log of earnings, and a set of indicators for educational attainment), and relevant missing indicators (Table S.1 shows a complete list of covariates).

Table S.3, panel A presents the intervention effects on student reading achievement. Our preferred specifications 4 and 7 in Table S.3 correspond to the results presented in the text and include stratum fixed effects and the baseline achievement measures. Specifications 3 and 6 in Table S.3 include only stratum indicators, whereas specifications 5 and 8 in Table S.3 also include covariates. Inclusion of covariates has little effect on the estimated effects. There are 58 and 57 clusters in the NOTP and TP samples, respectively. The average cluster sizes are 18 in both samples, and the median is 18 in the NOTP sample and 20 in the TP sample. Only nine clusters in total contain less than 12 students; 50 clusters of roughly equal size are often enough for cluster-robust SEs to provide valid inference (Kézdi 2004).

Participating schools were asked to sign up for the reading test on the last day of testing (December 13). Thus, the estimated ITT effects presented here may be interpreted as an immediate ITT. As noted in the text, the results indicate considerable and significant effects of increasing instruction time (Table S.3, panel B). Particularly, four extra high-discretion Danish lessons per week (NOTP) benefit students in terms of improved reading achievement.

Although the results for increased instruction time NOTP are robust to the inclusion of baseline achievement and other covariates, the results for increasing instruction time with the low discretion program are not. Controlling for baseline achievement, the overall effect of the TP treatment becomes small and statistically insignificant.[8] The marked drop in the estimated effect size from specification 6 to 7 in Table S.3 suggests that the distribution of student baseline achievement across the control and treatment groups is slightly skewed. Overall, the TP treatment group students simply performed slightly better at the outset, although the differences in

[7] Non-Western origin includes students who are either immigrants or second generation immigrants from non-Western countries. Definitions of immigrants and non-Westerns countries are provided by Statistics Denmark.

[8] The drop in the point estimate is unlikely to be driven by attrition because of missing outcomes. The gap in estimated effect sizes is similar when including an imputed version (nonstochastic regression imputation) of the missing fourth grade test results as an outcome.

the means of the baseline achievements in reading and math were not significant. The text focuses on the specifications including the baseline achievement measures, which are more conservative.

To further elaborate on this issue, we test the effects of increasing instruction time against the effect of the third intervention, which aimed to enhance the skills of the classroommath teachers. The math teachers participated in a 5-day course designed by the Danish Ministry of Education and received a written proposal of an ideal lesson plan to better qualify them to integrate Danish as a second language in the math classes. This treatment is not expected to affect—or is expected to at least affect only to a very limited extent—student achievement in reading. At the same time, the distributions of baseline achievement measures among the students in the TP treatment group are more comparable with those of the students in the math treatment group than the control group, thus, implying that the math treatment group is a plausible comparison group.[9] Strictly speaking, the estimated effects will reflect how much increased instruction time increases student achievement relative to an upgrading of the math teacher. The results of this robustness analysis are shown in Table S.3, panel B. The estimated effects in specifications 3–5 are on the same order of magnitude as those in Table S.3, panel A, and they are still robust to the inclusion of baseline achievement measures. However, the inclusion of baseline achievement in specification 7 in Table S.3 causes a less dramatic drop in the estimated effect than would be expected if the TP group is somewhat positively selected relative to the control group.

Furthermore, we analyze the intervention effects on the three domains of the reading test: language comprehension, decoding, and reading comprehension. Table S.3, panel C presents the results where the three domains are considered as separate outcomes. All of the specifications include the baseline achievement to take into account the sensitivity of the effect of the TP intervention. The results are robust to the inclusion of covariates. The curriculum in the TP intervention was developed with a particular focus on language comprehension. Interestingly, language comprehension seems to be the only domain significantly affected by the TP intervention, whereas the NOTP intervention seems to impact a wider range of cognitive reading domains.[10]

Explorative analyses of whether the effects of increased instruction time change across identifiable subgroups of students are conducted based on gender and country of origin. Table S.5 presents the results of our main specification by subsamples: boys, girls, students of non- Western origin, and students of Western and Danish origin. Where information of the relevant characteristic is missing, the observation is dropped from the subsamples; thus, the sample sizes do not correspond to the results in Tables S.3 and S.4. Unadjusted *p* values are shown. To control for the familywise error rate associated with testing multiple hypotheses simultaneously,

[9]Mean baseline reading score for the math treatment group is -0.103, and the corresponding mean baseline math score is −0.169.

[10]This analysis should be considered exploratory only; *p* values are not corrected for multiplicity.

we apply the simple stepdown method proposed by Holm (1979). Adjusted p values are shown. Although more powerful than the standard Bonferroni method , this correction assumes independence of the individual p values, and thus, we consider it conservative. We block 24 hypotheses in Table S.5 into two families based on the outcome of choice [i.e., reading scores (Table S.5, panel A) and SDQ scores (Table S.5, panel B)]. Using two-sided tests and a familywise error rate of 0.1, we can still reject that increased instruction time NOTP has no effect on Danish students and students of Western origin and that the extra lessons with a TP have no effect for girls. In the SDQ block, the null hypotheses of no treatment effect of increased instruction time with the TP are rejected for girls.

B.3 Analyses Based on a Two-Level Hierarchical Model

We investigate whether the nested structure of the data significantly affects our results in Table S.3. A standard two-level hierarchical model of student achievement is considered to account for the sampling of students within classrooms. As before, all of the models include stratum fixed effects, and analyses both with and without baseline achievement and student-specific covariates were conducted. For the full models, the student level (level 1) includes all of the baseline achievement measures and covariates. The school-level model (level 2) includes indicators for treatment assignment and random effects for individual schools. Formally, we consider the following models.

Level 1 model (students):

$$Y_{ijk} = \beta_{0j} + \mathbf{PreTest_i}\boldsymbol{\beta_1} + \mathbf{x_i}\boldsymbol{\beta_2} + Stratum_k + v_{ijk},$$

Level 2 model (schools):

$$\beta_{0j} = \gamma_0 + Treatment_j\gamma_1 + r_j.$$

In the models, γ_1 measures the ITT effect of each treatment in separate specifications. The model is estimated separately for the two treatment groups to increase flexibility. A more detailed description of the variables can be found in relation to the description of the model used for the primary analyses.

The results from the two-level hierarchical models are presented in Table S.6. The results are very similar to the results presented in Table S.3, which is in line with previous findings that hierarchical linear models and linear models with clustered SEs produce similar results (Green and Vavreck 2008).

B.4 Analyses of Teachers' Use of the Additional Instruction Time

During the intervention period, two rounds of implementation surveys were administered to the two treatment groups; one in the last half of September and one in

November (Fig. S.1). In the following section, we focus on results from the latter one, which is thought to better evaluate the overall implementation of the interventions.

All teachers in the NOTP treatment respond that they are the regular Danish teacher of the participating class; 77% of the teachers (two nonrespondents) report the same in the TP sample.

In the remainder of this section, the implementation survey sample sizes are 30 for the NOTP sample (one nonrespondent) and 24 for the TP sample (four nonrespondents).

The NOTP teachers. The NOTP teachers were free to spend the extra time as they pleased. This feature is reflected in the survey data; 90% of the responding teachers reported that they, to some or a great extent, spent the extra lessons covering the same material that they had already planned—but in greater detail. At the same time, most teachers also experimented with new material. In total, 90% responded that they, to some or a great extent, tried out new material and methods; 40% reported that they, to some or a great extent, spent the extra lessons on separate projects or bundled the lessons for project days.

The NOTP teachers received a 27-page report for inspiration on how to support bilingual students, but they were not required to use it in the treatment; 40% of the teachers reported that they used this material to a small extent, whereas 40% reported not to have used it at all. Only 17% of the teachers found the report useful, whereas 47% found it useless or did not use it at all. The teachers generally had a positive attitude toward the intervention; 77% of the NOTP teachers agree that the high-discretion intervention was profitable for the class as a whole, whereas 73% agreed that the bilingual students profited. Eightythree percent responded that they enjoyed teaching the extra lessons.

The TP teachers. The TP teachers were presented with a very detailed TP to use. The survey data confirm a high compliance rate. In the implementation surveys, 96% of the responding TP teachers (all but one) stated that they had used the TP material to some or a large extent; 83% responded that they found the material useful. The remaining 17% were indifferent.

Also here, the teachers generally had a positive attitude toward the program; 88% agreed that the intervention with the TP was profitable for the class as a whole, whereas 83% agreed that it was profitable for bilingual students. Sixty-seven percent of the TP teachers enjoyed teaching the program in the extra lessons.

SI Figures

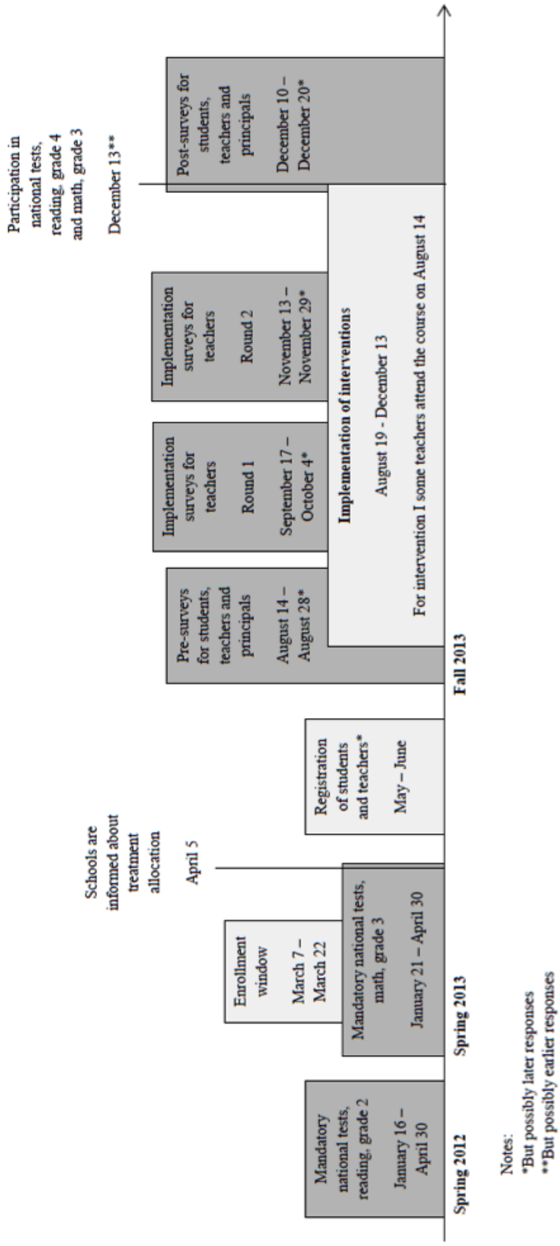

Figure S.1: Timeline of the intervention with increased instruction time

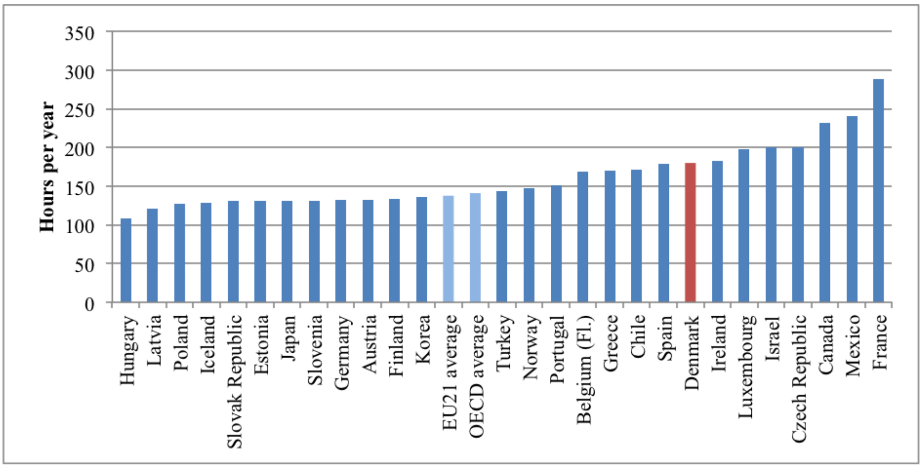

Figure S.2: Yearly instruction time in reading, writing and literature at age 10
This figure illustrates the total yearly instruction time in reading, writing and literature for 10 year-olds across OECD countries (OECD 2014).

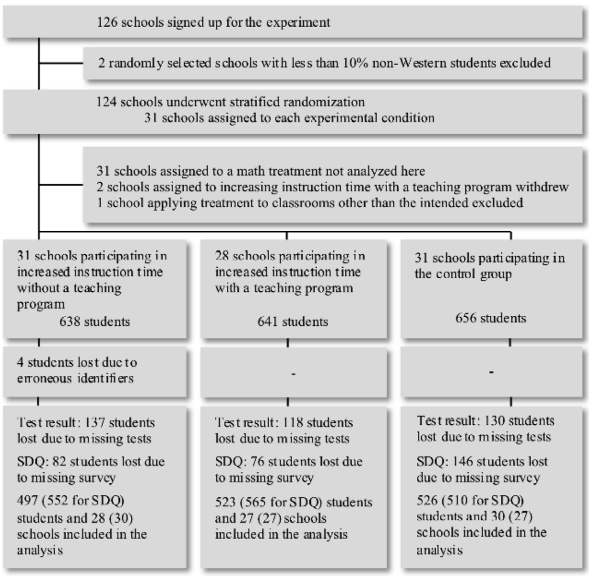

Figure S.3: Flow diagram of schools and students participating in the randomized trial.

SI Tables

Table S.1: Balancing of student characteristics

	(1) Without a teaching program		(2) With a teaching program		(3) Control group		Difference (1)-(3)	Difference (2)-(3)	
	Mean	SD	Mean	SD	Mean	SD			
Students									
Test score, reading grade 2	-0.18	1.07	-0.14	0.94	-0.23	1.01	0.05	0.10	
Missing test score, reading grade 2	0.04		0.02		0.05		-0.01	-0.03	***
Test score, math grade 3	-0.18	1.08	-0.07	0.98	-0.19	0.97	0.00	0.11	
Missing test score, math grade 3	0.06		0.04		0.08		-0.02	-0.04	
Non-Western background	0.29		0.22		0.30		-0.01	-0.08	
Girl	0.50		0.49		0.51		-0.01	-0.02	
Single-mother family	0.27		0.24		0.23		0.04	0.02	
No. of siblings	1.46		1.32		1.48		-0.02	-0.16	
First-born	0.40		0.40		0.40		0.00	0.00	
Age 9 or younger	0.29		0.26		0.28		0.01	-0.02	
Age 10	0.64		0.68		0.65		-0.01	0.02	
Age 11	0.07		0.07		0.07		0.01	0.00	
Born in the first quarter	0.24		0.25		0.22		0.02	0.03	
—second quarter	0.22		0.23		0.26		-0.03	-0.03	
—third quarter	0.29		0.30		0.27		0.02	0.03	
—fourth quarter	0.24		0.22		0.25		-0.01	-0.03	
The mothers									
Logearnings	8.82	5.40	9.46	5.04	8.72	5.43	0.10	0.74	
Age	36.00	5.28	35.60	5.12	35.62	5.38	0.38	-0.02	
None or missing education	0.10		0.06		0.09		0.01	-0.03	
High school	0.30		0.34		0.34		-0.05	-0.01	
Vocational education	0.31		0.31		0.28		0.03	0.03	
Higher education	0.29		0.29		0.28		0.01	0.01	
The fathers									
Logearnings	9.84	5.03	10.32	4.71	9.80	5.06	0.04	0.52	
Age	39.11	6.14	38.36	5.67	39.07	6.38	0.05	-0.71	*
None or missing education	0.12		0.09		0.10		0.01	-0.01	
High school	0.30		0.33		0.34		-0.04	-0.01	
Vocational education	0.32		0.31		0.32		-0.01	-0.01	
Higher education	0.27		0.27		0.24		0.03	0.04	
N schools (total = 90)	31		28		31		62	59	
N students (total = 1931)	634		641		656		1290	1297	

Notes. Observations with missing information are excluded from the table unless otherwise indicated. When covariates are included in regressions, relevant indicators for missing covariates are always included. All information related to parents and family is registered when students are six years old; t tests are corrected for clustering at the school level. $^{*}p < 0.1$, $^{***}p < 0.01$.

Table S.2: Attrition analysis

	(1) No test Coeff.	S.e.	(2) No test Coeff.	S.e.	(3) No SDQ Coeff.	S.e	(4) No SDQ Coeff.	S.e.
No teaching program	-0.001	(0.063)	-0.002	(0.061)	-0.098 **	-0.049	-0.096 **	(0.047)
With a teaching program	-0.006	(0.062)	0.000	(0.059)	-0.120 **	-0.052	-0.114 **	(0.050)
Students								
Test score, reading grade 2			-0.039 ***	(0.015)			-0.028 **	(0.012)
Missing test score, reading grade 2			0.174 ***	(0.059)			0.058	(0.048)
Test score, math grade 3			0.028 *	(0.014)			-0.006	(0.012)
Missing test score, math grade 3			0.153 **	(0.063)			0.123 **	(0.050)
Non-Western background			-0.032	(0.022)			-0.019	(0.022)
Missing immigrant information			0.000	(0.002)			-0.003	(0.002)
Missing immigrant information			-0.021	(0.131)			-0.165	(0.136)
Age 10			0.008	(0.019)			-0.011	(0.021)
Age 11			0.053	(0.048)			0.019	(0.040)
Girl			0.015	(0.018)			0.009	(0.017)
Single-mom family			0.046 **	(0.021)			0.051 **	(0.022)
No. of siblings			-0.002	(0.009)			-0.005	(0.009)
First-born			0.007	(0.017)			0.027	(0.019)
Born in the second quarter			-0.024	(0.025)			-0.008	(0.024)
- third quarter			-0.027	(0.028)			-0.032	(0.023)
- fourth quarter			-0.005	(0.032)			-0.027	(0.026)
Missing family information			0.028 *	(0.014)			-0.006	(0.012)
The mothers								
Logearnings			0.000	(0.002)			-0.003	(0.002)
Age			0.003	(0.002)			-0.004	(0.002)
High school			0.009	(0.040)			0.002	(0.044)
Vocational education			0.015	(0.039)			-0.018	(0.046)
Higher education			-0.005	(0.045)			-0.008	(0.048)
The fathers								
Logearnings			-0.001	(0.002)			-0.002	(0.002)
Age			-0.001	(0.001)			0.003	(0.002)
High school			-0.001	(0.044)			-0.016	(0.048)
Vocational education			0.004	(0.045)			-0.036	(0.047)
Higher education			0.040	(0.044)			-0.004	(0.047)
Constant	0.102 **	(0.042)	0.054	(0.103)	0.225 ***	(0.073)	0.253 **	(0.125)
Observations	1,931		1,931		1,931		1,931	
Adjusted R-squared	0.155		0.178		0.108		0.126	
Stratum FE	YES		YES		YES		YES	
Covariates	NO		YES		NO		YES	

Notes. The propensity of not attending the test/answering the survey with SDQ is modeled by a linear probability model. Coefficients for selected covariates are displayed. Indicators for missing information and stratum fixed effects are also included. Cluster-robust standard errors are in parentheses; $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S.3: Intervention effects on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw diff. in means		Primary results					
Panel a Intervention effects	Reading grade 4	Reading grade 4	Reading grade 4	Reading grade 4	Reading grade 4	Reading grade 4	Reading grade 4	Reading grade 4
Without a teaching program	0.247**		0.146*	0.148**	0.149**			
	(0.106)		(0.075)	(0.064)	(0.061)			
With a teaching program		0.206*				0.157**	0.037	0.036
		(0.109)				(0.060)	(0.055)	(0.052)
Constant	–0.563***	–0.563***	–1.192***	–0.331*	–0.092	–1.509***	–0.587***	–0.841***
	(0.078)	(0.078)	(0.301)	(0.166)	(0.295)	(0.038)	(0.116)	(0.291)
Observations	1,023	1,049	1,023	1,023	1,023	1,049	1,049	1,049
Adjusted R-squared	0.015	0.010	0.086	0.573	0.581	0.107	0.584	0.589
Stratum indicators	NO	NO	YES	YES	YES	YES	YES	YES
Baseline achievement	NO	NO	NO	YES	YES	NO	YES	YES
Covariates	NO	NO	NO	YES	NO	NO	YES	

Panel b Robustness check	(1) Reading grade 4	(2) Reading grade 4	(3) Reading grade 4	(4) Reading grade 4	(5) Reading grade 4	(6) Reading grade 4
Without a teaching program	0.181***	0.208***	0.219***			
	(0.064)	(0.051)	(0.046)			
With a teaching program				0.143*	0.090	0.096*
				(0.077)	(0.057)	(0.055)
Constant	–0.995***	–0.301*	–0.652**	–1.335***	–0.618***	–1.241***
	(0.191)	(0.156)	(0.300)	(0.082)	(0.101)	(0.292)
Observations	1,072	1,072	1,072	1,098	1,098	1,098
Adjusted R-squared	0.084	0.608	0.623	0.078	0.611	0.625
Stratum indicators	YES	YES	YES	YES	YES	YES
Baseline achievement	NO	YES	YES	NO	YES	YES
Covariates	NO	NO	YES	NO	NO	YES

Panel c Learning domains of the reading score	(1) Language comp.	(2) Decoding	(3) Reading comp.	(4) Language comp.	(5) Decoding	(6) Reading comp.
Without a teaching program	0.152**	0.167***	0.072			
	(0.059)	(0.055)	(0.062)			
With a teaching program				0.139**	0.006	–0.048
				(0.062)	(0.049)	(0.052)
Constant	–0.624***	–0.114	–0.136*	–0.838***	–0.365***	–0.348*
	(0.192)	(0.183)	(0.074)	(0.056)	(0.076)	(0.201)
Observations	1,023	1,023	1,023	1,049	1,049	1,049
Adjusted R-squared	0.444	0.459	0.464	0.452	0.462	0.482
Stratum indicators	YES	YES	YES	YES	YES	YES
Baseline achievement	YES	YES	YES	YES	YES	YES
Covariates	NO	NO	NO	NO	NO	NO

Notes. Panel a presents the results of our main analyses. Panel b presents the results when comparing the increasing instruction time interventions to the third (math) treatment arm. Panel c presents the results from our primary specification on the three domains of student reading achievement. See the table for included covariates. Cluster-robust standard errors are in parentheses; $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S.4: Intervention effects on student total behavioral difficulties and strengths

Panel a	(1) Raw diff. in means	(2)	(3)	(4)	(5) Primary results	(6)	(7)	(8)
Total behavioral difficulties	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.
Without a teaching program	0.081		0.563*	0.487	0.500			
	(0.461)		(0.333)	(0.317)	(0.320)			
With a teaching program		-0.286				–0.562**	–0.395*	–0.545**
		(0.441)				(0.231)	(0.232)	(0.220)
Constant	9.625***	9.625***	7.718***	5.819***	6.476**	10.623***	8.363***	10.352***
	(0.325)	(0.325)	(1.994)	(1.564)	(2.604)	(0.576)	(0.967)	(2.609)
Adjusted R-squared	–0.001	–0.000	0.026	0.101	0.105	0.037	0.113	0.129
Panel b	Strengths	Strengths	Strengths	Strengths	Strengths	Strengths	Strengths	Strengths
Without a teaching program	-0.102		–0.050	–0.043	–0.004			
	(0.160)		(0.129)	(0.129)	(0.129)			
With a teaching program		0.044				0.048	0.057	0.051
		(0.154)				(0.093)	(0.093)	(0.091)
Constant	7.725***	7.725***	8.846***	8.923***	8.913***	8.464***	8.553***	8.713***
	(0.119)	(0.119)	(0.121)	(0.125)	(0.531)	(0.220)	(0.246)	(0.550)
Adjusted R-squared	0.000	-0.001	0.041	0.042	0.059	0.06	0.064	0.070
Observations	1,062	1,075	1,062	1,062	1,062	1,075	1,075	1,075
Stratum indicators	NO	NO	YES	YES	YES	YES	YES	YES
Baseline achievement	NO	NO	NO	YES	YES	NO	YES	YES
Covariates	NO	NO	NO	NO	YES	NO	NO	YES

Notes. Panel a presents the results on the total difficulty score, whereas panel b shows the results on the prosocial score. See the table for included covariates. Cluster-robust standard errors are in parentheses; $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S.5: Heterogeneous intervention effects on student reading achievement (Panel a) and total behavioral difficulties (Panel b) by gender and country of origin

Panel a Reading achievement	(1) Boy	(2) Girl	(3) Non-Western	(4) Western
Without a teaching program	0.125	0.143	0.013	0.186
	(0.076)	(0.086)	(0.094)	(0.065)
p-value	*0.107*	*0.103*	*0.886*	*0.006*
Adj. p-value	[> 0.50]	[> 0.50]	[> 0.50]	[0.047]
Observations	505	507	310	710
With a teaching program	–0.095	0.163	–0.097	0.074
	(0.071)	(0.060)	(0.090)	(0.063)
p-value	*0.190*	*0.009*	*0.286*	*0.240*
Adj. p-value	[> 0.50]	[0.062]	[> 0.50]	[> 0.50]
Observations	523	515	290	757
Panel b Total behavioral difficulties	Boy	Girl	Non-Western	Western
Without a teaching program	0.954	0.019	0.363	0.543
	(0.435)	(0.529)	(0.528)	(0.386)
p-value	*0.033*	*0.971*	*0.495*	*0.165*
Adj. p-value	[0.196]	[> 0.50]	[> 0.50]	[> 0.50]
Observations	519	526	308	745
With a teaching program	0.477	-1.322	0.147	–0.593
	(0.424)	(0.468)	(0.535)	(0.269)
p-value	*0.265*	*0.007*	*0.785*	*0.032*
Adj. p-value	[> 0.50]	[0.054]	[> 0.50]	[0.224]
Observations	531	532	275	797
Stratum indicators	YES	YES	YES	YES
Baseline achievement	YES	YES	YES	YES
Covariates	NO	NO	NO	NO

Notes. The intervention effects of the two treatments are estimated in separate regressions. All specifications include stratum indicators and baseline achievement in reading and math. Observations with missing information on the relevant characteristic is dropped. Cluster-robust standard errors are in parentheses. p-values are in italic, p-values adjusted for multiplicity are in brackets.

Table S.6: Intervention effects on student reading achievement (Panel a) and total behavioral difficulties (Panel b) based on two-level hierarchical models

Panel a Intervention effects	(1) Reading, grade 4	(2) Reading, grade 4	(3) Reading, grade 4	(4) Reading, grade 4	(5) Reading, grade 4	(6) Reading, grade 4
Without a teaching program	0.150**	0.167***	0.166***			
	(0.066)	(0.055)	(0.052)			
With a teaching program				0.157**	0.041	0.039
				(0.061)	(0.049)	(0.046)
Constant	-1.187***	-0.321**	-0.111	-1.509***	-0.594***	-0.829***
	(0.172)	(0.144)	(0.270)	(0.164)	(0.131)	(0.273)
Observations	1,023	1,023	1,023	1,049	1,049	1,049
Number of groups	58	58	58	57	57	57
Panel b Intervention effects	(1) Difficulties	(2) Difficulties	(3) Difficulties	(4) Difficulties	(5) Difficulties	(6) Difficulties
Without a teaching program	0.563	0.487	0.500			
	(0.355)	(0.340)	(0.340)			
With a teaching program				-0.562	-0.395	-0.545
				(0.364)	(0.350)	(0.347)
Constant	7.718***	5.819***	6.476***	10.623***	8.363***	10.352***
	(1.015)	(0.998)	(2.177)	(0.985)	(0.974)	(2.263)
Observations	1,062	1,062	1,062	1,075	1,075	1,075
Number of groups	57	57	57	54	54	54
Stratum indicators	YES	YES	YES	YES	YES	YES
Baseline achievement	NO	YES	YES	NO	YES	YES
Covariates	NO	NO	YES	NO	NO	YES

Notes. Random intercepts are included in all specifications. Panel a shows the results on reading achievement, while Panel b shows the results on the total difficulty score. See the table for included covariates. Standard errors are in parentheses $^{***}p < 0.01$, $^{**}\ p < 0.05$, $^{*}\ p < 0.10$.

CHAPTER 3

CLOSING OR REPRODUCING THE GENDER GAP? PARENTAL TRANSMISSION, GENDER AND EDUCATION CHOICE

Maria Knoth Humlum
Aarhus University

Anne Brink Nandrup
Aarhus University

Nina Smith
Aarhus University

Abstract

Over the last decade, the economic literature has increasingly focused on the importance of gender identity and slow-moving social norms in an attempt to explain the persistence in remaining gender gaps. At the same time, transmissions of beliefs, culture and preferences have been used to explain intergenerational correlations of economic outcomes. This paper attempts to combine the two approaches. Using Danish register data on the latest cohorts of labor market entrants, we document a significant relationship in the gender stereotypical nature of educational choices across generations. Although to some extent picking up genetics and/or transmissions of skills, our results suggest that if parents exhibit gender stereotypical labor market behavior, children of the same sex are more likely to choose a gender stereotypical education with stronger associations for sons. Analysis accounting for unobserved family heterogeneity confirms these results. We propose that these slow-moving gender norms attenuate the final closing of the gender gap.

JEL Classification: I23; J16; J24
Keywords: Intergenerational transmission; Gender identity; Social norms

3.1 Introduction

In most Western countries, women have successfully gained on men's labor market position over the last 50 years. As a result, the gender gap in human capital investment, labor force participation and wage has diminished. Economists have long focused on the contributors to this gender convergence, including educational attainment (Blau and Kahn 2006), technological progress in household appliances (Coen-Pirani et al. 2010) as well as in the labor market such that physical strength is de-emphasized (Weinberg 2000), medical advances to control fertility (Goldin and Katz 2002), reduced discrimination (e.g. Altonji and Blank 1999), and availability of childcare (Waldfogel 1998). However, marked differences in pay, promotional patterns and types of activities performed by men and women still exist (Olivetti and Petrongolo 2016, World Economic Forum 2016). In a recent paper, Goldin (2014a) argues that the last chapter of the grand gender convergence is in the pipeline. Specifically, Goldin (2014a) suggests increasing labor market flexibility with respect to remuneration to and timing of working hours as a means to eliminate the remainder of the gender wage gap. Such change has already come about in several sectors, for example technology and health, while other sectors, such as corporate and legal, continue to lack behind. While we do not dispute the importance of labor market flexibility for female careers, in this paper we suggest that the gender convergence may be attenuated in its final stages by slow-moving norms. We approach this issue by considering the persistence in educational choices across generations.

Goldin (2014a) focuses solely on the gender gap in wages for cohorts of US women, suggesting that the majority of the gender wage gap occurs within sectors rather than between. Thus, changing the gender mix of occupations will not close the gender wage gap. While this is an important distinction, Blau and Kahn (2006) find that occupational segregation is one of the most important contributors to the gender differences in pay. Using Norwegian data Kirkeboen et al. (*forthcoming*) note that the gender pay gap is relatively small as compared to the earnings differences by educational fields. As such, we believe that the determinants of educational choice continue to be of the utmost importance. Further, perceived limitations in the educational choice set may result in suboptimal utilization of the labor force.

Seminal works by Akerlof and Kranton (2000, 2002) incorporate the sociological and psychological concept of identity into a standard utility framework. They define one's self-image or identity as belonging to a social category, which contains a set of appropriate behaviors prescribing how one ought to behave. Identity then influences educational and labor market outcomes, as deviating from the prescribed behavior is costly. For example, in the gender identity framework the social categories 'male' and 'female' and associated prescriptions such as "men care about prestige and career" and "women take care of children" would motivate men and women to choose different educational fields to avoid losses in identity. When these prescriptions are passed on through generations, identity norms will be sticky (Alesina et al. 2013). Indeed, one factor that has been emphasized as shaping particularly female labor market behavior over the last decades is the transmission of social norms within neighborhoods and across generations (e.g. Fernandez 2013, Fernandez and Fogli 2009, Fernandez et al. 2004, Blau et al. 2012).

This paper approaches the gender gap in occupation choice by examining the intergenerational correlation in gender-stereotypical choice of education, specifically the degree to which individuals select into female-dominated educational fields. We consider the correlation of educational rather than occupational characteristics for three reasons. First, the choice of education is the first major decision individuals make concerning their future labor market career. Education paths are often chosen prior to starting a family or entering the labor market. Thus, for a particular set of skills, we argue that choice of education, although indubitably an important determinant of subsequent occupation, is more immediately related to the preferences of real world individuals, whereas later labor market outcomes, such as occupation, promotion and earnings, to a greater extent reflect fertility and marriage decisions in addition to labor market conditions and employer discrimination (Oguzoglu and Ozbeklik 2016). Second, although female labor force participation rates have been used in many studies (Fortin 2005) as a proxy for gender attitudes we do not consider it a satisfactory measure in for example the Nordic countries. Here, most women have been full-time labor market participants for decades; thus, having a mother in the labor force is a weak signal of household gender norms or even maternal comparative advantages. Third, the occupational segregation is as large (or even larger) in Denmark as in other countries (Dolado et al. 2002, Datta Gupta et al. 2008). A recent report on the Danish labor market suggests that while the educational segregation for university college programs (professional Bachelor's) has decreased somewhat in the recent decade, the educational segregation in higher university has increased (SFI 2016). Consequently, there is little hope that closing the gender gap in labor force participation or educational attainment will close the remaining gaps in occupational positions and earnings.

As demonstrated in Fig. 3.1, women graduating in decidedly female-dominated fields on average earn less compared to women with degrees in male-dominated fields. Of course, this is absent important wage determinants such as ability or motivational sorting (a growing literature considers the wage differences across field of study, see for example Altonji et al. 2015, Kirkeboen et al. *forthcoming*). Further, reminiscent of Bertrand et al. (2015) and Goldin (2014b) we suggest that, while previously somewhat neglected, male stereotypes are equally important factors of the remaining gender gaps. In particular, while women have assumed certain previously male-dominated educations, for example sociology and veterinary and agricultural sciences, there are no examples of a converse pattern.[1] However unlike for women, men may face reduction in expected earnings by entering female-dominated fields as many female-labelled occupations are in the public sector and on average pays less.

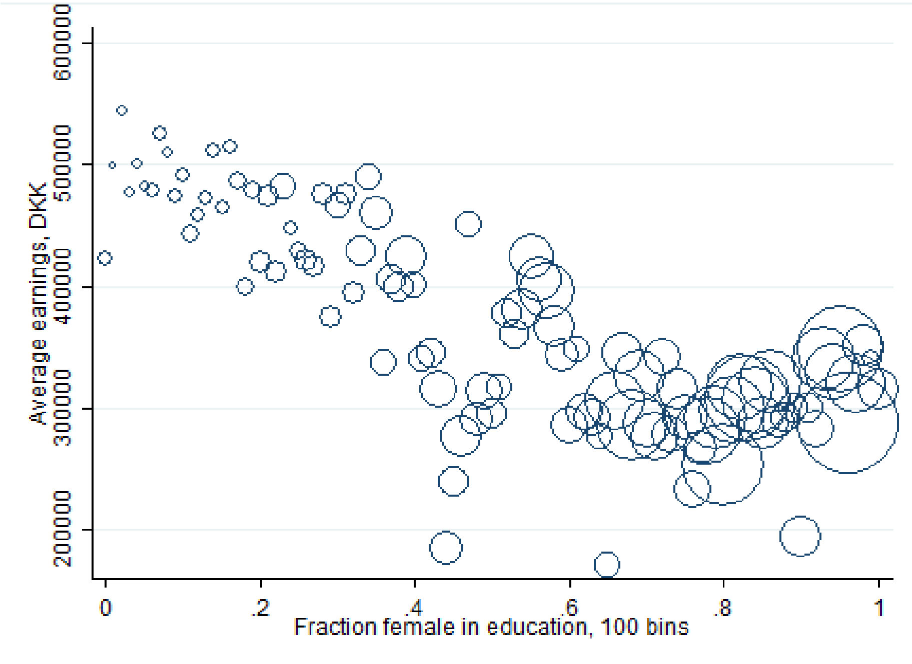

Figure 3.1: Working women's average annual earnings by fraction of female graduates in education, 2010. The sample includes all working women with at least a BA in 2010. The fraction females in education is measured as the share of female graduates in the year of enrollment. Markers are weighted by bin size.

In line with the intergenerational literature on occupational choice (for example, Corak and Piraino, 2011, on employers; Dunn and Holz-Eakin, 2000, and Linquist et al., 2015, on entrepreneurship), we begin by documenting significant positive intergenerational correlations in the degree to which parents' and their children's field of study are dominated by females. We then move on to consider the gender-stereotypical nature of other parental labor market behaviors as determinants for their children's education choice. In particular, children's choices tend to reflect the behavior of their same-sex parent. These findings are in line with intergenerational transmission of social norms, but as identity affiliations by construction are difficult to study empirically, we cannot exclude that our correlations pick up other kinds of transmissions. We therefore consider a wide range of complementary analyses

[1] For example, physiotherapy has seen a marked increase in male graduates from the 1970s to the 2000s; however, with more than 70% female graduates on average in the 2000s it is still decidedly female.

and competing hypotheses as to why gender differences in educational preferences may persist across generations. In this, we take advantage of the rich administrative records on the family affiliations, labor market and education outcomes.

The remainder of this paper is organized as follows. Section 3.2 places the intergenerational transmission of education in the context of the relevant literature and presents potential transmission mechanisms. Section 3.3 introduces the data and estimation procedure and Section 3.4 presents the results and discusses the hypotheses tested. Relevant measures are presented continually throughout sections. Finally, Section 3.5 discusses and concludes.

3.2 Remaining gender gaps and intergenerational transmissions

A vast economic literature on gender trends has explored the rise in female labor force participation following World War II; see e.g. Goldin (2014a) and Olivetti and Petrongolo (2016) for recent and comprehensive descriptions.[2] However, where Goldin (2014a) considers the diminishing gender wage gap as evidence of an impending final chapter of the gender convergence, Olivetti and Petrongolo (2016) note that the remaining gender gaps in several labor market outcomes, including college major choice, are remarkably persistent—particularly when considering the reversing gaps in educational attainment in many countries. Specifically, the general progress in gender trends has slowed considerably since the mid-1990s (Blau and Kahn 2006, Olivetti and Petrongolo 2016).

To various extents, gender differences remain in almost all labor market outcomes, such as wage, working hours and occupation. The reversing gap in length of educational attainment seems to be the grand exception. Women in most industrialized countries simply go to school longer compared to men (World Economic Forum 2016) while obtaining better grades (with the exception of math-based subjects, Fryer and Levitt 2010). Still, marked educational and occupational segregation exists. In contrast to Goldin (2014a), we argue that although job flexibility is important for promoting female careers, differential parental transmissions and slow-moving gender norms may continue to uphold gender-stereotypical labor market behavior if the labor market is already highly gender-segregated. This concerns both male and female gender stereotypes alike.[3]

[2] Goldin (2014a) on the US, Olivetti and Petrongolo (2016) on several industrialized countries.

[3] Relatedly, Pan (2015) analyzes the dynamics of occupational segregation and develops a social interaction model to explain "tipping" occupations, i.e. occupations that rapidly become predominantly female once the share of females in an occupation exceeds a given threshold. Fig. A.2 suggests some degree of persistence in female-dominated fields of study despite the large influx of women into college in recent decades, however, within specific programs the variation is likely greater. If gender identity is transmitted from parents to children in their adolescence, and the share of females in parents' education indeed captures these gender norms, we would generally not expect later changes in the gender composition of their education to influence the educational choices of their children. On the other hand, a great subsequent influx of women into parent's occupations may change their gender attitudes.

Fortin (2005), contributing to a relatively recent and growing body of literature, proposes that the stagnating gender convergence coincides with renewed interests in traditional family and religious values. While changing societal attitudes towards female work have been associated with women's substantial labor market gains (Fernandez et al. 2004, Fernandez 2013, Fogli and Veldkamp 2011), Alesina et al. (2013) find evidence of very slow-moving gender norms dating back to pre-industrial societies, thus, suggesting that the remainder of the gender gaps may be highly persistent. Relatedly, Fortin (2005) notes that in contrast with discriminatory beliefs about women in the labor market, which decreased somewhat during the 1990s, the views on women's traditional role seem more stable across time.

A prominent literature on intergenerational mobility documents considerable positive correlations between educational and occupational outcomes of parents and children (e.g. recent and comprehensive reviews by Björklund and Salvanes 2011, Black and Devereux 2011) where in particular the same-sex correlations appear strong. Although possibly also arising from transmissions of comparative advantages in skills, this evidence is in line with inheritable social norms. In addition, Hellerstein and Morrill (2011) find that the intergenerational transmission of occupations from fathers to daughters has increased with women's rising labor force participation.

Much of the literature concerning intergenerational evidence of occupation and education choice focus on the transmission of economic resources and human capital from parents to children, including information, skill and network transfers (e.g. Lentz and Laband 1989; Laband and Lentz 1992; Dunn and Holz-Eakin, 2000; Black and Devereux 2011; Corak and Piraino 2011), which are potentially closely linked to transmissions of parental attitudes and diffucult to separate from these. In a study of Northwestern University sophomores, Zafar (2013) finds that one of the greatest determinants of individuals' college major choices is to gain their parents' approval. Such transmissions are likely to cause positive associations in education and occupation choice across generations, although from an income-maximizing perspective they alone cannot explain the dominant same-sex associations demonstrated in the empirical literature (e.g. Thomas 1994). A number of studies, however, have shown that parents tend to invest more in their same-sex children (Lundberg 2005), thus, potentially producing larger correlations along the same-sex dimension. Although, Grönqvist et al. (*forthcoming*) demonstrate that labor market outcomes for children of both sex are equally strongly related to both the mothers' and fathers' cognitive and non-cognitive skills.

Fernandez (2013) and Fogli and Veldkamp (2011) construct models of information transfers in close geographical neighborhoods. Specifically, the long-term consequences of female labor force participation is transmitted between women in the same geographical neighborhood, however, mothers and mothers-in-law may constitute particularly strong transmitters of information regardless of residential closeness. For example, Fernandez et al. 2004 present evidence that wives of men whose mothers participated in the labor force are more likely to work themselves, although the authors propose that the operating channel is the change in norms of these men and not information spillovers among women.

The economics literature has long been concerned with why men and women select into different types of education. Arising from the classical human capital model, Polachek (1981) suggests that women self-select into occupations in which human capital depreciates relatively slower, i.e. in which the wage penalty of long-term absences from the labor market, for example in connection with childbirth, is lower. Given that human capital depreciation structures evolve slowly across time, this mechanism would lead to positive intergenerational correlations in occupational choices along the same-sex dimension as documented in the literature without explicit transmissions from parents to children.

Yet another strand of literature has focused on the differences in psychological traits to explain the gender gap in occupational choices.[4] Bonin et al. (2007) empirically demonstrate that more risk-averse individuals tend to sort into occupations with more stable earnings, which are on average lower paid due to compensating wage differentials for risk-averse agents. Women on average exhibit more risk-averse behavior than men (e.g. Reuben et al. *forthcoming*). Maestripieri et al. (2009) empirically link individuals' testosterone levels to risk aversion and further present evidence that career choice is related to testosterone levels. Olivetti and Petrongolo (2016) argue that service-sector jobs match female preferences and household roles well, and that females have a comparative advantage in this sector. Relatedly, Zafar (2013) shows that gender differences in college major are explained by differences in preferences rather than in skill levels. Antecol and Cobb-Clark (2013) determine that entry into male-dominated fields is related to traditional 'masculine' psychosocial traits such as impulsivity, independence and non-emotion. Humlum et al. (2012) find that identity-related social and career factors, on which men and women load differently, are related to planned field of study. Finally, Reuben et al. (*forthcoming*) conclude that major choice, as defined by four broad fields, is unrelated to their experimental measures of competitiveness and overconfidence (which differs systematically between genders), but that these may operate through the choice of specific major within the broad categories.

Throughout the economics literature, there is overwhelming evidence that men and women differ systematically with respect to attitudes toward risk, competition, negotiation, social preferences and other psychological attributes, e.g. the Big Five (Bertrand 2011).[5] Although many personality traits are likely genetically inherited, for example testosterone levels and risk aversion in Maestripieri et al. (2009), or

[4] Building on Altonji and Blank (1999), Bertrand (2011) surveys these newer explanations of why men and women exhibit different labor market behavior including psychological attributes and societal norms.

[5] Much of the literature is based on lab experiments involving hypothetical or small stake choices, thus, Bertrand (2011) calls for additional evidence linking the experimentally observed differences to real labor market outcomes.

transmitted from parents, they may be exacerbated by societal norms. Interestingly, evidence suggesting that many systematic gender differences arise as a product of societal factors rather than pure genetics has surfaced over the last decades. For example, in response to Guiso et al. (2008,) who demonstrate a negative relationship between the gender gap in math and societal gender equality indicators, Fryer and Levitt (2010) find that the gender gap in math develops during the school years for all strata of society, but not in Muslim countries where same-sex classrooms and schools are prevalent. Likewise, Booth and Nolen (2012a, b) and Booth et al. (2014) note that girls from single-sex schools behave more like boys in terms of risk attitudes and willingness to compete. Nevertheless, Lindquist et al. (2015) find that post-birth factors (interpreted as nurture) account for twice as much as pre-birth factors (nature) when decomposing the intergenerational association in entrepreneurship using the Swedish adoption registers.

Following the seminal paper by Akerlof and Kranton (2000), a literature has focused on gender identity as a leading determinant of gender differences in occupational and educational preferences. The authors propose a model in which identity enters directly into the utility function. Thus, one's identity, encompassing for example the social categories 'man' and 'woman' and the prescriptions of how one ought to behave in these categories, may change predictions about one's actions because of a desire to conform with these societal modes of behavior. Relatedly, Fortin (2005) shows that women's gender role attitudes across 25 OECD countries are highly predictive of their labor market outcomes. And, based on the prescription that "a man should earn more than his wife", Bertrand et al. (2015) demonstrate that wives with a potential to earn more than their husbands distort their labor supply to reduce earnings while they increase time spend on household chores. The authors interpret this as a mean of appearing less threatening. Adhering to gender categories may further drive the observed differences in behavior within same-sex and coed school environments. Specifically, girls may reinforce stereotypical behavior to appear attractive when boys are present, while boys may have a preference for assertiveness to attract the opposite sex and reduce threats from competitors (Booth and Nolen 2012a, b).

While Akerlof and Kranton (2000) treat identity as given, Bordalo et al. (*forthcoming*) offer an explanation of how group identity is formed. Still relying of the concept of group affiliation, the individual's self-identity here is constituted by her stereotype of the group to which she belongs. Because stereotypes exaggerate differences between groups, expected payoffs are distorted by individuals' incorrect beliefs about own ability in comparison to members' of other groups, thus, influencing behavior. This relates to Goldin (2014b)'s pollution theory of discrimination, where female hiring in male-dominated occupations leads to reduced occupational prestige in the opinion of outsiders because of asymmetric information about the value of the individual female's characteristics. In Bordalo's et al. (*forthcoming*) framework, however, the transmission of non-stereotypical information is often ineffective in changing stereotypes as individuals overreact to stereotype-confirming information and fails to update their beliefs in the face of non-stereotypical information.

Somewhat overlapping with the theory of role model identification, group identity affiliations potentially generate positive same-sex associations in education and occupation choice of parents and children if gender identity and social norms are inherited from parents (Blau et al. 2013, Fernandez et al. 2004, Fernández and Fogli 2009, Johnston et al. 2014), which also results in slow-moving social norms (Alesina et al. 2013). Consequently, the remaining occupational and, in particular, educational gender gaps may be highly persistent. Where Ruef et al. (2003) suggest that role models are typically of the same sex, gender stereotypes may be transferred from representative of the opposite sex or even the surrounding society (although the theory of same-sex role modeling extends far beyond the transfer of stereotypes only). As the mechanisms underlying intergenerational transfers are often intertwined conceptually, it is not surprising the empirical predictions of the theory of same-sex role modeling are difficult to distinguish from those of differing parental skilll transfers based on same-sex preferences.

Traditionally, the literature has focused on the changes in female gender roles to close the gender gaps. Studies by Maccoby (1998) support this in suggesting that the pressure of conforming to gender identity is greater for girls than boys, and Johnston et al. (2014) demonstrate that gender attitudes are transmitted from mothers to both daughters and sons, although, only daughters' labor market outcomes appear to be affected by this. However, reminiscent of Goldin (2014b) and Bordalo et al. (*forthcoming*), upholding male gender stereotypes may be an important factor in closing the remaining gender gaps.[6]

3.3 Data and estimation procedure

3.3.1 Empirical methodology

Ideally, we wish to assess the influence of parental gender norms on child education choice. Unfortunately, (the transmission of) social norms and the associated stereotypes are inherently difficult to study since the concept of group identity is not directly observable. Ethically or even practically, it would be infeasible to attempt to identify the effect of parental genders norms on child education choice using lab or field experiments. First, choice of education is a consequential life decision to manipulate. Second, information on attitudes and stereotypes has to be observed for both the parent and child generations when children choose their educational

[6] The influx of females into higher education has increased relatively more than that of males, and women have gained footing in many male-dominated fields. Fields such as nursing, teaching and humanities are still predominantly female, while engineering and business are distinctively male (Goldin 2014a, see Figures A.1 and A.2 for Danish data).

field (preferably the one, they end up completing). Thus, we settle for addressing the transmission of social norms and in particular gender norms as one of several possible channels through which educational and occupational gender seggregation persists.

We exploit the detailed nature of the Danish administrative registers to collect information on actual education and labor market behavior of the parents of entire population cohorts, see Sections 3.3.2 and 3.4.3. In terms of a gender norm interpretation, our approach is supported by the findings of Fortin (2005) who documents very high correlations between the self-reported gender attitudes of particularly women and their observed labor market behavior, however due to the difficulty of pinning down an unambiguous interpretation of what the measures pick up we will refrain from making strong conclusions.[7]

We begin by regressing our measure of gender-stereotypical education choice for individual i on our measure of gender-stereotypical education choice for his parents based on the following reduced form model for individual i:

$$FF_i = \alpha_0 + \alpha_1 FFmom_i + \alpha_2 FFdad_i + \alpha_3 X_i + u_i, \tag{3.1}$$

where FF denotes the share of females in the chosen education of the individual, his mom, or his dad, X_i are child and family characteristics presented in Section 3.3.4 and u_i is the OLS regression error term. α_1 and α_2 then determine the mother–child and the father–child intergenerational correlation in female-dominated educational choices, respectively. Other transmissions affecting educational preferences, e.g. from peers and siblings, and other kinds of parental transfers not captured by the share of female graduates in the field of study or in the controls in X_i will be in u_i. As such, the results from eq. (3.1) are partial correlations rather than causal effects and should only be interpreted as such.

Choice of educational field is considered a major determinant of labor market success in adulthood, but as previously discussed this choice likely reflects the more immediate preferences and self-image of the individual compared to later labor market outcomes (Ozumoglu and Ozbeklik 2016). We therefore use education choice characteristics as our main outcome variable. Acknowledging that determinants for education choice and influence of gender stereotypes potentially operate through different channels for men and women (Blau et al. 2013, Johnston et al. 2014), we go on to estimate eq. (3.1) separately by gender. We wish to test the hypothesis that parental education choice affect the education choices of their children and, thus, whether α_1 and α_2 differ significantly from zero.

[7] Several papers study transmissions of self-reported gender roles or self-stereotyping using surveys and retrospective questionnaires (for example, Johnston et al. 2014). However, survey measures of self-reported gender roles and self-stereotyping may suffer from different types of measurement problems. In particular, the respondent may not answer truthfully if gender norms and identity are considered a controversial area (Eriksson et al. 2016). Therefore, the degree to which children would pick up or respond to these self-reported measures is uncertain.

The intergenerational correlation in our measure of stereotypical education choice likely picks up a range of factors related to share of females graduates across generations other than gender norms, for example inherited and acquired comparative advantages in certain skills. Through a wide range of analyses, we attempt to pin down channels through which the intergenerational correlation in female-dominated education choice most likely runs. Direct information on most of these channels are not available in the Danish registers, however, they do to some extent have testable implications, which we will utilize. We futher include the wide range of family and parental characteristics in our regressions (see Table 3.2) to reduce influence from confounding factors. Potential mechanisms may be placed into three broad categories, namely, societal factors, genetic transfers and parental transfers. Societal factors include labor market conditions in terms of for example remuneration and work hour schemes. Genetic transfers encompass all shared biological factors between parents and children. Parental transfers include direct transfers of human and financial capital, networks, information, and social norms (with social norms potentially transmitted indirectly through parental role modeling as well). Although we are not able to unequivocally pin down the interpretation of the intergeneration correlation in female-dominated education choice, our analyses render some channels more probable while dismissing others.

After thoroughly addressing the intergenerational correlation in female-dominated education choice, in Section 3.4.3 we move on to consider briefly the influence of other parental labor market outcomes, that may reveal information about parental gender norms, on child education choice.

3.3.2 Measuring gender-stereotypical education choice

We measure gender-stereotypical education choice for cohort members, FF_i, as the share of female graduates in the education program (see also Antecol and Cobb-Clark 2013, Eriksson 2015) in the year the individual enrolled in the education program. Thus, our outcome variable depicts the gender composition of the program as observed by the individual when he or she applied to school. For cohort members, we consider only Bachelor's level programs, see Section 3.3.3. We measure degree of gender-stereotypical education choice for both men and women as share of female graduates, therefore, the more stereotypical for men the less female-dominated (and consequently, more male-dominated).

We observe the main explanatory variables at age 15 of the cohort member including obtained education. For the older parents in the sample, information about year of enrollment is incomplete. Thus, we measure parents' gender-stereotypical education choice as the share of female graduates obtaining the degree at age 30 of the parents, not restricting the level of the degree. If parents did not obtain a formal education, the fraction of females in the highest attained general education level (compulsory school or high school) is applied. Not surprisingly, the top female-dominated education programs in the 1970s include dental assistant, pharmaconomist and midwife, while various mechanics, craftsmen and officer (army, navy and air force) programs top the male-dominated list with next to no share of female graduates.

The extent to which a given field is female- or male-dominated reflects, at least in part, the required, underlying skill set (for example, math skills). Skills are often inheritable if not by nature then by nurture and would generate positive correlations in education share of females across generations in itself. We address this issue, first, by including controls for high school ability and, second, by contrasting the magnitude of the correlations in families with various sibling compositions (see Section 3.4.2), and third, by exploiting the Danish administrative registers to construct alternative measures of household gender norms. We present these measures in detail in Section 3.4.3.

3.3.3 Sample and descriptives

From the birth registers, we identify all 1,133,658 children (cohort members) in Denmark born in 1970–1986. We further restrict the estimation sample to individuals for whom we can identify parents and parental country of origin. To obtain a homogenous sample of young adults and avoid e.g. integration aspects, we exclude children whose parents are not both of Danish ancestry, i.e. where neither parent is both born in Denmark and a Danish citizen. Further, educational outcomes for the parent generation are generally more unreliable and to a larger extent missing for immigrants. This leaves 949,862 observations.

Our analysis is primarily based on individuals obtaining at least a Bachelor's degree (BA)[8] at age 28. We choose this sample to resemble the college major choice literature. Data on educational attainment is available until 2014. To avoid truncation issues for the younger individuals in the sample, we observe the educational outcomes at age 28 for all individuals. This cutoff should leave plenty of time to finish a BA even including a couple of gap years which are common for Danish students; compulsory school is generally completed at age 15. A three-year high school education (or alternatively two-year plus grade 10) qualifies for admission in most BA programs. As seen from Table 3.1, 27% of cohort members obtain at least a BA at or before age 28. As in most Western countries, female graduates dominate the BA programs, although with large differences between fields; sixty-two percent of

[8] A Bachelor's degree (3-year) from a university or a Professional Bachelor's degree (3–4.5 years) from a university college. BA education is defined as the first six digits in the 8-digit educational code used by Statistics Denmark, i.e. fields such as different business programs with a lot of common factors are categorized as one. Following the Bologna process, formal Bachelor's programs were only established in the early 1990s, with exceptions, such as medicine and dentistry in the mid-2000s. Thus, for individuals enrolling in Bachelor's programs (educational code: 60xxxxxx) before 1994, we use the fraction females graduating from the corresponding Master's program (educational code: 65xxxxxx). We use the same approach for individuals enrolling in programs for architecture before 2005, medicine before 2003, dentistry before 2008, theology before 2003, as pharmacists before 2006 and as land surveyor before 2008. Correspondingly, the data contains 237 different BAs or higher.

Table 3.1: Sample overview

	Observations		Percent
Individuals born of Danish ancestry 1970–1986 with matched parents	949,862		100.0%
First choice BA at or before age 28	312,741		32.9%
Male	123,708	39.6%	
Female	189,033	60.4%	
Completed a BA at or before age 28	260,355		27.4%
Male	99,649	38.3%	
Female	160,706	61.7%	

Notes. Summary educational statistics of cohort members in the estimation samples. First choice BA is defined as the first enrollment choice if that was at a BA level not conditional on completion.

the individuals obtaining a Bachelor's degree are female. Table 3.1 further includes information of the cohort members' first enrollment in a BA education. First choice of enrollment likely reflects educational preferences with less consideration of the individual's skill level. Therefore, we include first enrollment if that was a BA program without conditioning on completion in our supplemental analyses.

Table 3.2 provides a detailed description of the cohort members in the BA estimation sample. Not surprisingly in light of the considerable overweight of women in the BA programs, males in the sample are on average of slightly higher 'quality' than females: better high school GPAs, and higher earning, higher educated parents. Focusing on the primary variables instead, we note a marked difference in the gender composition of the chosen BA programs between cohort member men and women. Men on average obtain their degrees in fields with almost 30 percentage points less females compared to women. The gender compositions of the mothers' and fathers' obtained education are very similar across the male and female samples, while the difference in gender compositions for mothers' and father's education is roughly the same.

We are unable to match the fraction of females graduating in the year the individual enrolls for 3,983 observations (1.5% of the BA sample) due to newly established or periodic educational fields. Consequently, the BA estimation sample includes 256,372 observations in total.

3.3.4 Control variables

We collect all controls other than parents' female-dominated education choice in the vector X_i. The wide range of variables in X_i are included to capture cohort, region or family characteristics that may jointly affect parental labor market behavior and cohort-member educational behavior. Besides the variables listed in Table 3.2, X_i includes a full set of indicators for cohort, mother's and father's year of birth and residential municipality (at age 15). Specifically, birth year fixed effects for children and parents control for the general rise in the ratio of female to male BA graduates across years. Family and parental controls are measured at age 15 for the child, the last year of compulsory education and, thus, the last year before tracking in the Danish education system.[9] Of particular importance is the individual's high school GPA, which controls for ability and is likely correlated with both parental abilities and future labor market behavior for the cohort member. The decision to attend high school is in itself likely based on future educational expectations and is as such not exogenous. It is, however a prerequisite for enrolling in most BA programs. The older high school registers are limited to two types of schools (STX and HF), with single-course HF included from 1992 and HTX and HHX from 2001 onwards. Further, only passed GPAs (above 5.5, equivalent to D by US standards) are recorded, and high school GPA is therefore missing for a relatively large fraction of individuals (15%). Where information on non-primary individual controls is missing, a dummy variable adjustment approach is used.

We further control for parental educational attainment (compulsory, high school, vocational and higher education), logearnings, mother's and/or father's work hour, number of siblings (as identified by the same mother) and in particular older siblings of the same and opposite sex. We also include dummies for whether the mother or father died between age 15 and year of entry in tertiary education.

3.4 Results

The empirical analysis first considers the intergenerational correlation in gender-stereotypical education choice, focuses second on differential patterns by family structure and sibling composition to lure out potential channels, and third on alternative measures of gender attitudes in parental labor market behavior.

3.4.1 Intergenerational correlation in gender-stereotypical choice of education

We begin by documenting the size of the correlation between the share of females in parents' education and the share of females in their daughters' and sons' choice of

[9] Although, parental characteristics at age 15 may in part reflect behavioral response to the child itself, for our purpose we prefer to use the later measures to capture the household norms at the age when the child faces the first actual educational choices and to a large extent has developed an autonomous 'persona' presumably reflected in the parents' behavior. For example, Burt and Scott (2002) confirm that gender role attitudes extend back into early adolescence. An additional reason for using age 15 instead of a younger age (e.g. age 5) is that it is possible to get register information on parents for a larger sample. Sensitivity checks using variables measured at age 5 (i.e. a smaller data set of all cohorts born in 1975–1986) yield very similar results.

Table 3.2: BA estimation sample descriptives

Variables	Male Mean	Male SD	Female Mean	Female SD
Primary variables				
Frac. female in BA education	0.424	0.243	0.709	0.208
Frac. female in mother's education	0.666	0.197	0.665	0.195
Frac. female in father's education	0.331	0.267	0.323	0.269
Controls				
Birthweight < 2500	0.037		0.046	
Born in first quarter	0.258		0.254	
—second quarter	0.277		0.274	
—third quarter	0.250		0.253	
—fourth quarter	0.215		0.219	
Firstborn	0.476		0.468	
Multiple born	0.021		0.020	
No. of siblings (by mother)	1.412	0.853	1.431	0.875
No. of older brothers	0.357	0.596	0.364	0.605
No. of older sisters	0.335	0.577	0.348	0.589
High school GPA	8.588	0.922	8.431	0.929
Parents (child age 15)				
Mother's logearnings	10.81	3.853	10.7	3.928
Morther zero earnings	0.109		0.115	
Mother outside labor force	0.119		0.122	
Mother working part time	0.208		0.200	
Mother working full time	0.673		0.678	
Mother's age	43.07	4.374	42.69	4.462
Parents separated	0.121		0.13	
Father's logearnings	10.97	4.259	10.74	4.421
Father zero earnings	0.126		0.140	
Father outside labor force	0.159		0.166	
Father working part time	0.017		0.016	
Father working full time	0.824		0.817	
Father's age	45.45	5.044	45.21	5.11
Mother's education				
—None/missing	0.424		0.709	
—Max. high school	0.020		0.019	
—Vocational	0.306		0.331	
—Higher	0.459		0.388	
Father's education				
—None/missing	0.035		0.038	
—Max. high school	0.177		0.215	
—Vocational	0.338		0.394	
—Higher	0.451		0.353	
Total observations	99,649		160,706	

Notes. Sample includes individuals of Danish ancestry born in 1970–1986 with at least a BA at age 28. Observations with missing information are excluded from the table unless otherwise indicated. When non-primary covariates are included in regressions, a dummy-variable adjustment for missing observations is used.

education. To establish a baseline, we report the estimated coefficients from eq. (3.1) with a full set of controls in column (1) of Table 3.3.[10]

Not surprisingly, women are more likely to enter female-dominated fields of study than men. Conditioning on the full set of controls, women select into fields with a 27-percentage points greater share of female graduates in the year of enrollment. This coefficient is surprisingly similar to Antecol and Cobb-Clark (2013), who further condition on self-reported psychosocial characteristics to estimate a gender gap of 22 percentage points. Notice that this difference in preferences is driven neither by demographic and socioeconomic characteristics nor by human capital as captured by high school GPA. Although we cannot rule out discrimination before education completion, it seems likely that this gender segregation at least in part reflects different preferences for educational (and later occupational) characteristics.

As expected, the coefficient on high school GPA in Table 3.3 is negative meaning that high-ability individuals sort into less female-dominated fields. While more females in either parents' education has the individual selecting into a more female-dominated field, we suspect that these correlations are highly heterogeneous by gender. For example, if children mainly reflect the behavior of their same-sex parent (Ruef et al. 2003, Johnston et al. 2014) or parents have a preference for same-sex children (Maccoby 1998, Lundberg 2005). Therefore, columns (2) and (3) repeat the analyses separately for sons and daughters.[11]

3.4.1.1 Correlation coefficients by gender

The results in Table 3.3, columns (2) and (3) establish significant and positive intergenerational correlations in female-dominated choice of education, although the same-sex parent correlations dominate. Roughly speaking, a 10-percentage point increase in share of females in father's education increases the share of females in the son's education by 0.9 percentage points, although, in terms of standard deviations the size of the correlations are modest. The correlation is lower along the female (mother–daughter) dimension but from a markedly higher baseline.[12] Also for fe-

[10] We do not include fixed effects for parental field of study in our regressions with the shares of female graduates in parents' education as the primary regressors. Were we to condition on parental educational field, the intergenerational correlations would be identified only by the variation in the share of females to obtain a certain degree over time, which is undesirable if educational gender stereotyping only evolve slowly over time. Including fixed effects for 11 broad educational fields does not change our findings, although to some extent affecting the magnitude of the correlation coefficients.

[11] Although the estimated intergenerational correlations in female-dominated education choice is significant, our measures of parental gender-stereotypical education choice are only weakly related to individual's labor market outcomes at age 30, see columns (1)–(4) in Table A.1. Male logearnings are negatively related to share of females in the father's education, whereas female unemployment is positively related to parents having a non-stereotypical degree.

[12] A fractional outcome raises certain specification issues. Most notably, our linear models ignore the bounded nature of the outcome variable and cannot guarantee that our predicted values lie between 0 and 1. In practice, though, only very few fitted values fall outside of the range 0–1. We choose to ignore the structure of our outcome measure, although the effects on the share of female graduates cannot realistically be constant on full range.

Table 3.3: Determinants of female-dominated field of study: fraction female (OLS regression)

	(1)			(2)			(3)		
	All, baseline			Male, baseline			Female, baseline		
VARIABLES	Coef.		SE	Coef.		SE	Coef.		SE
Female	0.270	***	(0.005)						
Frac female, mother's educ	0.028	***	(0.003)	-0.004		(0.004)	0.054	***	(0.005)
Frac female, father's educ	0.039	***	(0.002)	0.084	***	(0.004)	0.013	***	(0.003)
High school GPA	-0.047	***	(0.001)	-0.018	***	(0.002)	-0.064	***	(0.002)
Born in second quarter	-0.003	**	(0.001)	-0.003	*	(0.002)	-0.001		(0.001)
— third quarter	-0.003	**	(0.001)	-0.001		(0.002)	-0.003	**	(0.001)
— fourth quarter	-0.002		(0.002)	0.001		(0.002)	-0.004	**	(0.002)
Low birthweight (< 2500 g)	0.013	***	(0.003)	0.019	***	(0.004)	0.008	**	(0.003)
Firstborn	-0.009	***	(0.002)	-0.005		(0.003)	-0.010	***	(0.002)
Multiple born	-0.004		(0.003)	-0.005		(0.005)	0.000		(0.005)
No. of siblings (by mother)	-0.002	**	(0.001)	-0.005	***	(0.001)	0.001		(0.001)
No. of older sisters	0.009	***	(0.002)	0.012	***	(0.003)	0.006	***	(0.002)
No. of older brothers	0.012	***	(0.001)	0.019	***	(0.003)	0.007	***	(0.002)
Mother's logearnings	-0.003	***	(0.001)	-0.002		(0.001)	-0.004	***	(0.001)
Mother zero earnings	-0.036	***	(0.007)	-0.020		(0.014)	-0.043	***	(0.009)
Mother works < 30 hrs/week	-0.002		(0.001)	-0.007	***	(0.002)	0.003	**	(0.001)
Mother outside labor force	-0.007	***	(0.002)	-0.011	***	(0.003)	-0.006	**	(0.002)
Mother's age	-0.010	***	0.000	-0.007	***	0.000	-0.001		0.000
Mother's education:									
—None/missing	0.000		(0.006)	0.007		(0.011)	-0.001		(0.007)
—Vocational	-0.014	***	(0.001)	-0.006		(0.003)	-0.017	***	(0.003)
—Higher	-0.011	***	(0.002)	0.014	***	(0.004)	-0.024	***	(0.003)
Separated parents	0.010	***	(0.001)	0.015	***	(0.002)	0.006	***	(0.001)
Father's logearnings	-0.006	***	(0.001)	-0.007	***	(0.001)	-0.005	***	(0.001)
Father zero earnings	-0.071	***	(0.008)	-0.084	***	(0.012)	-0.060	***	(0.007)
Father works < 30 hrs/week	-0.006		(0.005)	0.008		(0.009)	-0.015	***	(0.003)
Father outside labor force	-0.015	***	(0.003)	-0.020	***	(0.004)	-0.012	***	(0.003)
Father's age	-0.003	***	(0.000)	-0.011	***	(0.001)	-0.002	***	(0.000)
Father's education:									
—None/missing	-0.009		(0.007)	-0.024	*	(0.012)	0.002		(0.008)
—Vocational	0.007	***	(0.001)	0.009	***	(0.002)	0.007	***	(0.002)
—Higher	-0.014	***	(0.003)	0.000		(0.003)	-0.021	***	(0.003)
Mother died before BA	0.019	***	(0.006)	0.002		(0.013)	0.026	***	(0.005)
Father died before BA	0.018	***	(0.004)	0.024	***	(0.008)	0.012	***	(0.003)
Constant	1.478	***	(0.030)	1.413	***	(0.051)	1.414	***	(0.034)
Observations	227,042			86,297			140,745		
R-squared	0.323			0.067			0.121		
Birth year & region FE	YES			YES			YES		

Notes. Samples include individuals of Danish ancestry with at least a BA at age 28. All covariates are included (see Table 3.2); excluded categories are born in the first quarter, mother/father working ≥ 30 hrs/week and mother/father having basic education level. Standard errors corrected for clustering within cohort member birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

males, there is a significant and positive correlation with share of females in father's education; although it is only one quarter the size of the correlation with the mother's educational characteristics.

Overall, the intergenerational same-sex correlations in female-dominated education choice decrease significantly once we condition on high school GPA, although markedly more for females (see Appendix Table A.2). In line with our earlier reasoning, we expect genetic factors and transfers of cognitive skills reflected in GPA to pick up some of the intergenerational correlation. Conversely, high school GPA is potentially affected by parental gender attitudes, for example, a mother with traditional gender attitudes may raise her daughter to be less ambitious which will likely be captured by high school GPA. Moreover, the negative coefficients on GPA, suggesting that individuals who are more able graduate in less female-dominated fields, is three times larger for women than for men (albeit the difference is potentially caused by selection into our sample). High-ability women thus seem to endeavor to enter less female-dominated fields.

Considering the other controls, having parents who work part-time (less than 30 hours a week) during one's adolescent years appears to be a strong signal for especially opposite-sex children; men with part-time working mothers and (i.e. a traditional division of market labor within the household) and women with part-time working fathers (untraditional) obtain degrees in programs that are significantly less female-dominated. There is no relationship between sons and part-time working fathers. Interestingly, having a mother with a higher education increases the share of females in men's education, whereas they appear unaffected by fathers' higher education. On the other hand, women graduate from BA programs that are significantly less female-dominated when either parent has a higher education or increased earnings. High-ability parents fostering high-ability (female) children who endeavor to enroll in relatively less female-dominated BA programs could explain this, however, the coefficients remain largely unchanged when excluding high school GPA (see Table A.2). It seems likely that having a mother with a higher education may inspire children to choose a less gender-stereotypical field of study, i.e. more female-dominated for sons, less for daughters.

When we exclude individuals graduating in the same or a similar education as one of their parents in an attempt to disentangle the effect of education-specific human capital and information transfers, the correlation coefficients drop but remain positive and statistically significant (see columns (3) and (6) in Appendix Table A.2).[13] We interpret this as evidence that transfers of education- (occupation-)specific human capital drive a smaller part but not all of the same-sex correlations in female-dominated educational choice across generations. In general, we note that our measure of parents' gender-stereotypical education choice is significantly related to

[13] Eriksson (2015) presents correlation estimates from specifications excluding observations where children to a varying degree have the same occupation as one of their parents for Swedish cohorts born in 1943-1952. Her findings are consistent with ours.

overall high school grades, math grades and probability of taking advanced level math courses, thus, we cannot rule out that they to some extent capture transmissions of comparative advantages in specific skills (see columns (5)–(10) in Table A.1). Interestingly, children of fathers with degrees in less stereotypical fields are less likely to take advanced level math courses, however, conditional on taking advanced math thet obtain better math grades (insignificant for daugthers).

The estimated relationships between the degree to which parents and their same-sex children choose a female-dominated (gender-stereotypical) education warrant the question if these relationships indeed are linear. For example, one might hypothesize that the correlations in gender compositions are particularly strong in the tails due to stronger transmission of gender stereotypes, i.e. for fathers and mothers who have chosen an education that is either very gender stereotypical or not at all. Our findings suggest that same-sex correlations are largest when fathers and mothers have a very male-dominated education (summarized in Fig. A.3).[14] Specifically, increasing the share of women from a very low baseline, for example from 0.05 to 0.15, seem to matter most for educational choices. In comparison, a corresponding change in the share of women in parents' education from a high baseline, e.g. of 0.85, supposedly have little effect on children's educational choices.

Summarizing our insights from Table 3.3, we note that men and women seem to mirror the education choice of particularly their same-sex parent when choosing field of study.[15] However, more able women are less likely to choose a female-dominated (gender-stereotypical) education and less (more) influenced by their mother (father). Men, on the other hand, are much less sensitive to ability level and reflect the behavior of their father only. We note that although the estimated correlations are significant and positive, the correlations sizes are modest and the explanatory power of female-dominated education choice for parents is modest compared to for example high school GPA (for girls). However, the share of females in parents' education is an imperfect measure and likely to pick up only parts of and not all transmissions associated with educational preferences.

In particular, the presence of gender-specific transfers within families is interesting in light of Grönqvist et al. (*forthcoming*), who demonstrate that for Swedish youths born around 1980 the cognitive and non-cognitive skill transmissions from

[14] In particular, this pattern differs slightly for males and females. Mother's education choice influences daughters more when the mother exhibits untraditional labor market behavior, i.e. has a male-dominated education or is the household breadwinner. The father–son correlations exhibit a less clear pattern; fathers with a very male-dominated or a slightly female-dominated education have the bigger influence on the son's educational choice. These samples are the largest and have relatively traditional divisions of work hours—more than 21% of mothers work part time.

[15] Eriksson (2015) documents similar correlations for share of females in occupation. Johnston et al. (2014) present mother-daughter correlations in gender role attitudes of 0.09 SD (mother-son correlations are similar though labor market outcomes for sons appear unaffected, and gender role attitudes are only measured for mothers). Black and Devereux (2011) survey intergenerational correlations for other labor market outcomes, in particular noting stronger father-son correlations in earnings than the corresponding father-daughter associations.

mothers and fathers are equally strong for children of both sex. Taken together, these results are in line with the findings of Johnston et al. (2014): although children inherit skills from both parents their labor market behavior mainly reflect that of the same-sex parent. This evidence is consistent with gender stereotyping affecting the children's educational preferences, although it does not explicitly rule out alternative mechanisms.

3.4.2 Addressing potential channels

In this section, we make use of the varying family structures and sibling compositions across families to address the potential channels trough which stereotypical education choice is transferred across generations and reduce the influence of confounding factors.

3.4.2.1 Transmissions across family structures

If transmissions from parents other than genetic endowments take part in shaping the educational preferences of their children, we would expect that the intergenerational correlations in educational characteristics are larger when the parent takes part in upbringing the child (although the direction may be more ambiguous for sons if gender norm transmissions are dominant as child rearing is considered a traditional female task).[16]

Information on where the child lives if parents are divorced along with information on new spouses are available in detail from 1990 and onwards. Thus, to study intergenerational correlations in different family settings Table 3.4 include cohorts born only in 1975–1986 for whom this family information is available at age 15, columns (1)–(5) presents the male results and columns (6)–(10) the female. Columns (1) and (6) present the correlation coefficients for individuals living with both parents at age 15, comprising the majority of the male and female cohort members. These correlations are very similar to the baseline (Table 3.3). Notice that the correlation coefficients remain unchanged for children living with their same-sex parent alone or with a new partner (columns (4)–(5) for sons and (7)–(8) for daughters).[17] Interestingly, when sons live with their mothers alone the coefficient on the share of females in their mother's education increases and becomes marginally significantly positive

[16] We further attempted to separate out (some) genetic determinants by exploiting information of adopted individuals to establish whether the intergenerational correlations are lower for adoptive children. Unfortunately, adoptive registers for this sample do not include year of adoption, which make it difficult to distinguish biological/adoptive parents as for example stepparents may adopt individuals well into adulthood. Nonetheless, the adoption sample is relatively small (around 1,500 individuals) resulting in very imprecisely estimated intergenerational correlations of share of females in education choice with the parents of the individuals at age 15.

[17] Individuals living with their mother because their father died and vice versa are excluded from the samples. Due to small sample sizes, estimations on subsamples in which parents have died from external causes do not add much information to our analysis and are therefore omitted here.

for the first time. The influence of the father remains unchanged (column (2)). When the mother finds a new partner (column (3)), her influence disappears and, interestingly, the father's decrease as well, though the correlation with the education choice of the new partner is not significant. Although less convincing due to severely limited sample sizes, the intergenerational correlations for daughters living with their fathers exihit the same pattern.

These differential patterns are in line with our expectations if educational preferences are indeed affected by parental transmissions. In the absence of the same-sex parent the opposite-sex correlations increase, however, this could be prescribed to both transfers of skills or attitudes. The pattern blurs somewhat when the parent finds a new partner; for sons, the influence of both parents are reduced, for daughters they may even increase. However, it is important to stress that child custody decisions are not random as further indicated by the large discrepancies in the numbers of children living with their mother versus father in case of a divorce. One may therefore easily construct selection-based explanations with the same hypothesized outcomes; for example, families in which fathers gain physical custody of the children likely have untraditional family norms.

3.4.2.2 Birth order differences

Having established interesting correlation patterns across family structures in Section 3.4.2.1, we turn our attention to birth-order differences next. If the parental transmissions captured by gender-stereotypical education choice indeed contributes to forming children's educational preferences we might expect that the associations are stronger for firstborns. For example, if parents invest more in their firstborns (Averett et al. 2011, Lehmann et al. 2014, Haan 2010), they might transmit stronger signals to them, or if firstborns are more fierce in choosing their parents as role models.

To quantify the influence of birth order on the parental transmissions for children's educational choice, we estimate the following model (again separately for men and women):

$$\begin{aligned} FF_i = \beta_0 + \beta_1 FFmom_i + \beta_2 FFdad_i + \beta_3 Firstborn_i + \beta_4 FFmom_i \times Firstborn_i \\ + \beta_5 FFdad_i \times Firstborn_i + \beta_6 X_i + \nu_i, \end{aligned} \tag{3.2}$$

where β_4 and β_5 depict the differential responsiveness of firstborns compared to laterborns of the same sex to transfers from their mothers and fathers, respectively. For example, when estimating eq. (3.2) for women β_4 denotes the differential influence of mother's transmissions on daughters born as the first child compared to daughters born as the second or third child. Initially eq. (3.2) does not include family fixed effects so we do not compare daughters within families.

Columns (1) and (5) in Table 3.5 present the intergenerational correlation coefficients by birth order as estimated by eq. (3.2) for men and women from families with two or more children (all children need not be in the sample), respectively. The

Table 3.4: Intergenerational correlation coefficient in female-dominated choice of education by family structure, cohorts 1975–1986

	(1) Male, both parents	(2) Male, only mom	(3) Male, mom w. new partner	(4) Male, only dad	(5) Male, dad w. new partner	(6) Female, both parents	(7) Female, only mom	(8) Female, mom w. new partner	(9) Female, only dad	(10) Female, dad w. n new partner
Frac female, mother's educ	−0.002	0.032*	0.008	−0.012	0.014	0.050***	0.049***	0.053***	0.040*	0.073
	(0.003)	(0.016)	(0.022)	(0.035)	(0.059)	(0.005)	(0.015)	(0.013)	(0.020)	(0.062)
Frac female, father's educ	0.089***	0.087***	0.035***	0.071**	0.068	0.014***	0.013	0.007	0.031	0.046
	(0.005)	(0.013)	(0.010)	(0.025)	(0.039)	(0.004)	(0.008)	(0.009)	(0.024)	(0.026)
Frac female, new partner			0.021		−0.017			−0.003		−0.056
			(0.016)		(0.072)			(0.006)		(0.037)
Mean dependent variable	0.433	0.458	0.466	0.456	0.454	0.706	0.715	0.730	0.712	0.713
Observations	52,801	5,657	3,142	1,185	740	86,209	10,667	6,822	1,441	864
R-squared	0.065	0.101	0.143	0.286	0.412	0.121	0.148	0.171	0.307	0.466
Birth year & region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes. Samples include cohorts 1975–1986 only. Selected variables are shown, see Table 3.2 for a full list of included covariates. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

intergenerational correlations are slightly higher in these samples compared to the baseline, 0.090 ($p < 0.01$) for father–sons, 0.082 ($p < 0.01$) for mother–daughters and 0.008 ($p < 0.05$) for father–daughters. The mother–son correlation is as always insignificant.

In line with increased parental investment in firstborns (Lehmann et al. 2014) and our previous documentation of a negative relationship between ability and female-dominated education choice, firstborns of both sex generally graduate in less female-dominated fields. Importantly, though, the intergenerational correlation in female-dominated education choice is higher in the same-sex dimension only, i.e. mother–firstborn daughter and father–firstborn son. The mother's (father's) educational choice on average influences women (men) who are firstborns relatively more than women (men) who are born second or later.

In an attempt to address family-specific factors that directly affect children's latent outcomes, for example shared genetics, columns (3) and (6) consider the differential responsiveness by birth-order to parental transmissions within families.[18] Note that the coefficients to the baseline fraction female variable in parents' education are identified only by variation in these measures between siblings at age 15. Because there is little variation in attained parental education between adolescent siblings these are both imprecisely determined and difficult to interpret. The coefficients to the interaction terms are, however, determined as the differential influence of e.g. mother's educational characteristics on firstborns *relative* to later born siblings of the same sex. Here, the father's differential influence on firstborns that are boys relative to boys born later disappears, although estimating eq. (3.2) on the subsample of matched brothers suggests that the decrease is driven by selection into this sample (column (2)). The point estimate of the differential influence of mothers' education on girls who are firstborns roughly doubles. For completeness, columns (4) and (8) check the sensitivity to the firstborn of each sex, respectively, with very similar results. We find evidence that in particular mothers' educational preferences are transmitted to the firstborn daughter. We note that in Table A.6 the birth-order gap in birthweight[19] is insignificantly related to our measures of parents' gender stereotypical education choice, suggesting that differential in utero investments across birth order is unrelated to parents stereotypical education choice. The same-sex differential pattern between first- and later-borns are therefore likely to arise from post-natal transmissions.

Results are consistent with younger sisters identifying with older sisters rather than their mother (role modeling) or, alternatively, that maternal investments are particularly strong for the firstborn daughters.

---

[18]This approach is inspired by Autor et al. (2015) who analyse the SES gradient in the gender gap between siblings.

[19]Autor et al. (2015) argue that gaps in neonatal health may act as proxies for the gaps in latent outcomes between children. For example, Black et al. (2007) and Lesner (2016) show that birthweight is a strong predictor for later labor market outcomes.

Table 3.5: Birth-order differences in the intergenerational correlation in female-dominated choice of education.

	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	2+ families	Matched	Matched	Matched	2+ families	Matched	Matched	Matched
VARIABLES		brothers OLS	brothers FE	brothers FE		sisters OLS	sisters FE	sisters FE
Frac female, mother's educ	−0.006	0.005	0.066	0.064	0.047***	0.040***	−0.003	−0.004
	(0.006)	(0.008)	(0.090)	(0.090)	(0.005)	(0.009)	(0.038)	(0.037)
Frac female, father's educ	0.077***	0.082***	0.045	0.048	0.012***	0.017**	−0.063	−0.060
	(0.005)	(0.009)	(0.107)	(0.107)	(0.004)	(0.006)	(0.050)	(0.050)
Firstborn	−0.019**	−0.012	−0.021		−0.021***	−0.036***	−0.033***	
	(0.007)	(0.010)	(0.014)		(0.005)	(0.010)	(0.009)	
First of each sex				−0.026*				−0.031***
				(0.012)				(0.007)
Birth-order interactions								
Firstborn $\times FFMom$	0.007	0.001	0.014		0.014**	0.030**	0.027**	
	(0.008)	(0.014)	(0.018)		(0.005)	(0.012)	(0.011)	
Firstborn $\times FFDad$	0.018***	0.008	0.006		0.001	0.001	−0.004	
	(0.006)	(0.011)	(0.009)		(0.004)	(0.008)	(0.009)	
First of each sex $\times FFMom$				0.021				0.029***
				(0.016)				(0.008)
First of each sex $\times FFDad$				0.001				−0.007
				(0.009)				(0.008)
Observations	79,139	17,792	17,792	17,792	129,218	32,911	32,911	32,911
R-squared	0.069	0.092	0.624	0.624	0.122	0.133	0.613	0.613
Birth year & region FE	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES
Parental FE	NO	NO	YES	YES	NO	NO	YES	YES

*Notes.*Selected variables are shown, see Table 3.2 for a full list of included covariates. Matched brothers and sisters, respectively, denote the subsamples of matched same-sex siblings. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.4.2.3 Sibling sex composition

Next, we attempt to pin down the channel of the differential birth-order correlations for daughters in particular. To the extent that parental differential investments, for example due to same-sex preferences, are responsible for our results in Table 3.5, it follows that the competition for attention and, hence, investments increases with the number of same-sex siblings (Eriksson 2015). More generally, sibling sex composition has been associated with parental educational investments of particularly females. For example, Oguzoglu and Ozbeklik (2016) propose that in the absence of a son, fathers may choose to invest in (one of) their daughters; that the presence of siblings of the opposite sex may reinforce gender roles (alternatively, reduce these if siblings mirror each other); or lastly, that daughters being more adverse to competition are discouraged from paternal investments in the presence of sons. Empirically, Oguzoglu and Ozbeklik (2016) find that having a brother significantly lowers the probability of women choosing STEM majors when the father is in a STEM occupation relative to when he is not.

To explore differential influence of parents' transmissions across sibling sex composition, Table 3.6 presents the results of eq. (3.2) where the firstborn indicator and interaction terms are substituted with an indicator for having siblings and older siblings of the same sex (columns (1) and (3)). The results in Table 3.6 do not support the hypothesis of educational transmissions through parental investments to the extent that they are a rival good (results are unchanged when the same-sex sibling indicators are included separately). Having any brother(s) (sisters) does not significantly affect the intergenerational correlations in female-dominated education for men (women) compared to when only opposite-sex children are in the family, in accordance with the birth-order correlations presented in Table 3.5, only the presence of older same-sex sibling(s) significantly decreases the correlation in female-dominated education choice for mothers-daughters and fathers-sons. These findings are consistent with a role modeling interpretation of the correlation coefficients, i.e. that children reflect the educational choice of the same-sex parents rather than the parent explicitly transmitting skills and information to the same-sex child.

Columns (2) and (4) in Table 3.6 examines whether opposite sex siblings reinforce stereotypes by including indicators for having siblings of the opposite sex. Column (2) presents marginal evidence that the presence of any sister(s) decreases the identity transmissions from mothers to sons, which is consistent, although not overwhelmingly so, with mothers preferring daughters to sons. However, unlike Oguzoglu and Ozbeklik (2016), we do not find evidence that the presence of sons lowers the educational transmission from fathers to daughters.[20]

[20] Differential fertility patterns for parents on the range of female-dominated education programs potentially explains these results. For example, the difference in father–daughter correlation estimates when a brother is present compared to when not will be attenuated if fathers with strong male identities are more likely to continue having children until they father a boy. Evidence from supplemental analyses

Table 3.6: Sibling composition differences in the intergenerational correlation in female-dominated choice of education

	Male		Female		All	
	(1) 2+ families	(2) 2+ families	(3) 2+ families	(4) 2+ families	(5) Matched siblings, OLS	(6) Matched siblings, FE
Female					0.229***	0.205***
					(0.005)	(0.007)
Frac female, mother's educ	−0.006	0.006	0.057***	0.050***	−0.013**	−0.069***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.022)
Frac female, father's educ	0.090***	0.081***	0.014***	0.009*	0.095***	0.039
	(0.006)	(0.005)	(0.003)	(0.005)	(0.004)	(0.055)
Female × FFmom					0.069***	0.084***
					(0.008)	(0.009)
Female × FFdad					−0.077***	−0.073***
					(0.004)	(0.008)
Have brother(s)	−0.007			−0.001		
	(0.007)			(0.005)		
Have older brother(s)	0.033***					
	(0.008)					
Have sister(s)		0.004	−0.003			
		(0.007)	(0.006)			
Have older sister(s)			0.019***			
			(0.006)			
Interactions for sibling presence						
Have brother(s) × $FFmom$	0.012			0.005		
	(0.008)			(0.006)		
Have brother(s) × $FFdad$	0.005			0.005		
	(0.006)			(0.005)		
Have older brother(s) × $FFmom$	−0.011					
	(0.008)					
Have older brother(s) × $FFdad$	−0.025***					
	(0.008)					
Have sister(s) × $FFmom$		−0.014*	0.003			
		(0.007)	(0.008)			
Have sister(s) × $FFdad$		0.007	−0.002			
		(0.007)	(0.004)			
Have older sister(s) × $FFmom$			−0.017**			
			(0.006)			
Have older sister(s) × $FFdad$			−0.002			
			(0.005)			
Observations	79,139	79,139	129,218	129,218	87,897	87,897
R-squared	0.069	0.069	0.122	0.122	0.309	0.682
Birth year & region FE	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES
Parental FE	NO	NO	NO	NO	NO	YES

Notes. Selected variables are shown, see Table 3.2 for a full list of included covariates. Regressions in columns (1)–(4) are carried out on the subsamples of men and women with at least one sibling, columns (5)–(6) include individuals with at least one matched sibling in the BA estimation sample only. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In particular, estimating the intergenerational transmission of female-dominated education choice for sisters relative to brothers in the spirit of Autor et al. (2015) allows us once again to reduce the influence from family-specific confounding factors. Columns (5) and (6) in Table 3.6 present results from linear regression and including parental fixedd effeccts, respectively, on the sample of matched siblings. In both specifications, we add a female indicater as well as interactions between the female indicator and the share of female graduates in both parents' education to the full set of covariates in Table 3.2. Again one should not pay too much attention to the estimated baseline correlation coefficients. However, even when including parental fixed effects in column (6), female cohort members on average choose educational fields that are 20-some percentage points more female-dominated. Further, the within-family daughter–son gap in female-dominated education choice is positively affected by mothers' gender-stereotypical choice of education and negatively affected by fathers'. I.e. mothers with a more gender-stereotypical education influence their daughters' field choice relatively more than they do their sons', and so they increase the gap in female-dominated education choice between daughters and sons. [21]

3.4.2.4 Sensitivity checks

Twenty-six percent of the individuals completing at least a BA at age 28 have previously been enrolled in (or completed) another program. The reasons for not completing the first choice are many, including student skill level being unaligned with the actual requirements of the program and the program not meeting ex-ante expectations. Thus, we suspect that our measure of parents' stereotypical education, in addition to parental transfers, to some degree also picks up parental influence on completing a BA program. We therefore propose that 'first choice of BA enrollment' is likely to reflect educational preferences that are more immediate and less related to skill set. Our results using completed education are similar when substituting with first choice of BA enrollment, however (Table A.4, columns (1) and (4)).

The intergenerational correlation in educational characteristics clearly operates through different channels for men and women, i.e. mainly of the same-sex parent. However, differential sorting into BA programs for men and women, as strongly suggested by Table 3.2, complicates a formal comparison of the correlations. In particular, men obtaining a BA are more able students on average than the corresponding women. By restricting the analysis to top students of both sexes based on their high

on parents of 2+-children families with at least one child born in 1970–1986 suggests that once controlling for parental education level, logearnings and birth year fixed effects the probability of having boys are on average not significantly related to share of females in either parents' education.

[21] As the gender gap in birthweight is generally unrelated to parental gender-stereotypical labor market behavior (Table A.6), these differential effects likely arise from post-natal influences: Differential sensitivity of boys versus girls to transmissions from mothers and fathers and/or differential parental investments in boys versus girls. Still, our analyses have not addressed that neighborhoods and school environments may vary with parental gender attitudes.

school GPA, we aim at creating samples that are more comparable.[22] The results underline our earlier findings that the intergenerational correlation in share of females in education choice decreases in magnitude for individuals that are more able, although stagnating around 0.04–0.05 for both men and women (Table A.4, columns (2)–(3) and (4)–(6)). Interestingly, women with a GPA above 10 are equally influenced by the share of females in the mother's *and* father's education and the difference in fathers' influence between sons and daughters is no longer significant.[23] This may equally be caused by fathers investing more in daughters that are highly (mathematically) skilled or alternatively that they are more skilled because of greater paternal investments. For sons, the results confirm earlier findings (e.g. Johnston et al. 2014) that even *if* males receive identity inputs from both mother and father, they seem to act mainly by mirroring the behavior of their fathers. This is consistent with the gender identity transmission hypothesis that men with a very gender-stereotypical educational background (low share of females) father boys that experience relatively larger losses in identity by not adhering to manly prescriptions (choosing a male-type education). Alternatively, these males may have relatively larger comparative advantages in skills that are not sought after in typically female-dominated fields, however, this does not rule out transfers of stereotypes. Women's outcomes seem to a greater extent affected by other factors such as skill level.

3.4.3 Alternative measures of household gender roles

Having established that educational characteristics are predominantly correlated across generations through gender-specific channels, we now focus on supplementary measures of parental gender attitudes in the labor market in the attempt to pick up transmissions that are less related to skill-specific transfers. The shares of females in parents' field of study are likely to pick up a range of factors besides the intergenerational transmission of gender norms. As mentioned, there is a relatively high degree of persistence in female to male graduates over time, and therefore, a positive same-sex intergenerational correlation may simply reflect Polachek-type education choices or transmissions of human capital and genetic factors that shape the educational preferences of children through comparative advantages. Consistent with this, Section 3.4.1 presents evidence that excluding children in the same educational field as at least one of their parents does decrease the magnitude of the

[22]For example for the individuals with high school GPAs above 10 (equivalent to US grade A) the difference in mother's earnings between sons and daughters in the sample drops from DKK 37,621 to just DKK 3,669, with sons still having the higher-earning mothers. The corresponding difference in father's earnings drops from DKK 25,027 to DKK 9,104.

[23]Supplemental analyses on an alternative sample for which detailed information about subject-specific grades is available to us, suggest that only fathers influence girls with very high math skills. Further, better high school grades in Danish actually increase the fraction of females in men's education, while decreasing it for women. For both genders, better grades in high-level math reduce the probability of choosing a female-dominated education.

intergenerational correlations but does not eliminate them.[24]

3.4.3.1 Measures based on labor market behavior

In the following, we model household gender attitudes by parental labor market behavior to supplement the preceding intergenerational education analyses. First, inspired by the behavioral prescription examined in Bertrand et al. (2015): "a man should earn more than his wife", we include mother's share of household earnings in our analyses. We define mother's share of household earnings as $HHshareMom_i = \frac{(EarningsMom_i)}{(EarningsMom_i + EarningsDad_i)}$. Traditional gender norms would prescribe the father's status as breadwinner, and thus we would expect sons (daughters) to select into relatively more male-(female-) dominated fields the lower $hhshareMom_i$.[25] We use the term household casually, as we do not condition on parents living together. Where both parents have zero earnings, we set $hhshareMom_i = 0$ and define a dummy for zero total household earnings. However, a large ratio of maternal to paternal earnings may both reflect that the mother is highly career-oriented (or at least successful in generating earnings), and that the father is not a high-income earner. These two explanations may have different implications if children predominantly reflect the behavior of the same-sex parent as hinted previously.

Second, we attempt to capture parental career ambitions individually as driven by earnings deviations from their respective demographic groups (see Bertrand et al. 2015). As such, the variable $EarningsGapParent$ captures parental labor market behavior, for example in response to family size, partner's earnings profile and own career orientation. For each individual i, we construct the potential earnings of both parents ($PotentialMom_i$ and $PotentialDad_i$) as the mean of the earnings in the mother's or father's demographic group. We assign demographic groups based on gender, a five-year age interval[26] and field of education. The potential (deflated)

---

[24] Reminiscent of Fernández and Fogli (2009) and Blau et al. (2012), we make one additional attempt to capture parental gender norms by source country cultural proxies for second-generation immigrants. In particular, we focus on country aggregate measures of gender role attitudes in parents source countries measured by fertility rate (the World Bank Indicators) and the ratio of male to female labor force participation rates (International Labour Organization, ILO), relating an increase in either measure to originating from countries with more traditional gender roles. These cultural proxies ideally reflect the aggregate preference and attitude distributions of parents' source countries. Therefore, if they hold explanatory power over the share of females in the educational fields of second-generation immigrants even after controlling for individual attributes, the societal gender norms contained in these proxies are responsible for the correlation. Since all individuals in this sample are born and grow up in Denmark, they are exposed to the same institutional setting. However, the gender attitudes as embodied in the cultural proxies may still matter if parents transmit their gender identity to their children. Contrary to Fernández and Fogli (2009) and Blau et al. (2012), we fail to demonstrate significant and robust correlations in cultural proxies, see Table A.5. For daughters, there is marginal evidence that a larger male relative to female labor force participation rate in the father's source country increases the share of females in the completed education. Education choices of second-generation immigrant are less related to high school GPA than in the Danish sample.

[25] In the estimation sample, mothers on average earn 42% of household earnings (the distribution is clearly skewed with 27% at the 25th percentile and 50% at the 75th, see Figure A.4).

[26] Due to incomplete information of graduation year, we are unable to use years of experience intervals.

earnings within demographic groups are calculated annually and are then merged to our data set at child age 15. We then define the deviations from potential earnings for the mother and father as $EarningsGapParent_i = \frac{(EarningsParent_i - PotentialParent_i)}{(PotentialParent_i)}$, for $Parent = \{mom, dad\}$.

Unlike Bertrand et al. (2015), we do not restrict the demographic groups to working individuals only as we prefer unemployment to be considered, at least in part, a consequence of educational choice. Thus, the earnings gap summarizes the relative difference between the parent's actual earnings and the average earnings for a person of the same gender, in the same age range and with the same educational degree as him or her. Table 3.7 presents the sample means of these alternative gender attitude measures.

Table 3.7: BA estimation sample descriptives, alternative household gender role measures

	Males		Females	
Variables	Mean	SD	Mean	SD
Mother's share of household income	0.416	0.270	0.423	0.277
Total household earnings = 0	0.032		0.036	
Log household earnings	12.58	2.424	12.45	2.565
Mother's potential earnings	190,829	61,055	181,631	58,291
Mother's earnings gap	0.036	0.609	0.036	0.602
Father's potential earnings	279,369	99,414	260,613	94,272
Father's earnings gap	0.118	0.753	0.099	0.760
Total observations	99,649		160,706	

Notes. Samples include individuals born of Danish ancestry in 1970–1986 with at least a BA at age 28. Observations with missing information are excluded from the table unless otherwise indicated.

Columns (1) and (6) of Table 3.8 present the selected point estimates in regressions of share of females in the education of men and women, respectively, on all control variables in Table 3.2. Further, because we are no longer interested in the variation of parents' educational characteristics, we include education fixed effects for both mother and father. The results in columns (1) and (6) underline that within parental educational background increased earnings for both parents are associated with obtaining a degree in a less female-dominated education. These results are in line with out previous findings. Consistent across specifications in Table 3.8, women alone (columns (6)–(10)) seem influenced by the household gender roles that are potentially reflected in parents' division of work hours. Having a mother (father) working part-time compared to full-time, thus, exhibiting more (less) gender-stereotypical household attitudes, is associated with a more (less) female-dominated choice of education, although the signal of a part-time working father is almost twice as large as that of the mother. Columns (2) and (7) introduce mother's share of household earnings as well as log-household earnings. Although of modest size, the coefficient to $HHshareMom_i$ is of the expected sign indicating that the larger share of household earnings brought in by the mother, the less of a male-dominated education her sons graduate from. For daughters, the coefficient is only marginally significant. As discussed in Bertrand et al. (2015), gender identity in relative earnings is more plausibly related to the prescription “a husband should earn more than his wife” rather than the actual earnings difference, although children in our setting may not observe the precise earnings difference. Correspondingly, columns (3) and (8) substitute $HHshareMom$ with an indicator for mothers earning more than fathers do ($HHshareMom > 0.5$). The coefficients to this indicator reflect our ex-ante expectations for women. Daughters of breadwinning mothers obtain a less female-dominated degree (less gender stereotyped) while sons are largely unaffected by this margin.

Specifications (4) and (9) of Table 3.8 model parental income gaps instead. A positive earnings gap indicates that the parent earns more than the potential (mean) earnings in his or her demographic group based on gender, field of study and age range, for example, because of higher career ambition or better skills. If we use the interpretation proposed by Bertrand et al. (2015), a large negative (positive) value of the mother’s (father’s) variable for income gap may reflect gender-stereotyped labor market behavior of parents and, thus, we should expect negative coefficients of the same-sex parent’s variable and—possibly—a positive for the parent of opposite sex. The coefficient on father’s income gap in column (4) is significantly negative for sons and the mother–daughter association in column (9) likewise, though slightly larger numerically. Consistent with expectations, there is no significant relationship between mother’s earnings gap and son’s education choice, but daughters’ education choices seem influenced by their fathers income gap as well: within education, fathers earning 10% more than their potential earnings, tend to have sons and daughters who obtain degrees in programs with 9 percentage points lower share of female graduates. Although, the earnings gap may to some extent pick up the same as the logearnings measures. Columns (5) and (10) include all alternative gender norm measures with unchanged results. Thus, conditional on a wide range of covariates, including individuals’ high school abilities, the more successful one’s parents are compared to their equals, the less female-dominated is the son’s or daughter’s choice of education. The observed pattern only partly confirms our a priori expectations concerning gender identity in the family and children’s educational choices. However, our results may point to the same mechanisms as the results concerning children’s GPAs, parental educational level and household income level: The more success in the educational system or in the labor market, the lower is the female fraction in the education chosen by the individual; and, men predominantly reflect paternal behavior while women are influenced by both parents although more strongly by the mother.

Table 3.8: Alternative measures of household gender roles

	(1) Male, baseline	(2) Male	(3) Male	(4) Male	(5) Male	(6) Female, baseline	(7) Female	(8) Female	(9) Female	(10) Female
Mother's logearnings	−0.002*	−0.003*	−0.001	−0.002	−0.001	−0.003***	−0.003***	−0.002*	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mother works < 30 hrs/week	−0.001	0.000	−0.001	0.000	0.000	0.005***	0.006***	0.005***	0.003**	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Father's logearnings	−0.003**	0.002	0.000	0.002	0.003*	−0.003***	0.000	−0.002*	0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Father works < 30 hrs/week	0.011	0.010	0.010	0.011	0.011	−0.009**	−0.010***	−0.009**	−0.011***	−0.010***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$HHshareMom$		0.024**					0.010*			
		(0.008)					(0.006)			
$HHshareMom > 0.5$			0.003		0.001			−0.004**		−0.004**
			(0.002)		(0.003)			(0.001)		(0.002)
Log household earnings		−0.006***	−0.006***		−0.003		−0.004**	−0.003**		0.000
		(0.002)	(0.002)		(0.002)		(0.001)	(0.002)		(0.002)
Income gap, mother				−0.001	−0.001				−0.014***	−0.013***
				(0.002)	(0.002)				(0.002)	(0.002)
Income gap, father				−0.009***	−0.009***				−0.008***	−0.009***
				(0.002)	(0.002)				(0.001)	(0.001)
Observations	92,951	92,947	92,947	92,708	92,704	150,622	150,617	150,617	150,237	150,232
R-squared	0.087	0.087	0.087	0.087	0.087	0.130	0.130	0.130	0.131	0.131
Birth year & region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Parent education FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes. Selected variables are shown, see Table 3.2 for a full list of included covariates; excluded categories are mother/father working ≥ 30 hrs/week. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.5 Concluding remarks

In a number of recent papers explaining the remaining observed gender gaps in the labor market, the role of gender-stereotypical preferences have been emphasized and documented. In this paper, we investigate the intergenerational correlation in gender-stereotypical choice of educational field. We find a positive and significant correlation between parents and their children with respect to the share of females in the educational fields of both generations. Same-sex correlations dominate, indicating that daughters' choices are more correlated with mothers' educational behavior, and sons' choices are more correlated with fathers' compared to the mother–son and father–daughter correlations, which are small and in most cases insignficant. Interestingly, the father–son correlations seem strongest, less related to ability level and less sensitive across specifications, suggesting that male group affiliation is potentially stronger. Additionally, sons only rarely reflect the behavior of the mother.

Gender-stereotypical norms in families are extremely difficult to measure and we are only able to use proxies based on administrative register data. Therefore, we check our results using several alternative measures of gender-stereotypical behavior such as the share of earnings in the household earned by the mother, and mother or father working part-time. These results generally confirm our findings.

What do these correlations reflect? Nurture or nature? Although, we cannot unambiguously attribute our same-sex correlations to within-family transmissions of gender stereotypes, our estimation results are consistent with the hypothesis that gender-stereotypical norms and gender identity are transferred from parents to children. The regressions based on families with more children where firstborn children's choices are more correlated with same-sex parental choices, the role of same-sex older siblings indicates that nurture plays a significant role. The same tendency is observed in the results across family constellations. It is the present parent whose choices are mainly correlated with son's or daughter's choices with respect to gender-stereotypical educational field, though there are differences between sons and daughters with respect to this pattern. We find no evidence to support that the correlations are generated by parental skill transmissions to the extent that they are a rival good. Parts of the correlations seem driven by education-specific skill transfers, although the associations remain significant when excluding children obtaing a degree in the same field as their parents. We suggest that role modeling may contribute to our results, however, skill transfers through role modeling are difficult to disentagle from transmissions of gender stereotypes both conceptually and empirically.

Even if the demonstrated intergenerational correlations are driven by skill rather than gender norm transmissions, the dominant same-sex correlations in gender-stereotypical educational choice offer an explanation of why large gender gaps are still observed in labor markets where women have been members of the labor force for decades and where they have outperformed their male peers with respect to quantity of education. If men and women reflect the choice behavior of their same-sex parent, horizontal occupational segregation will remain. Changes in job flexibility and other job attributes may facilitate women's entrance into more male-dominated occupations and may also play a role for occupational segregation but slow-moving gender norms will possibly postpone the final chapter of the remaining gender gaps.

Acknowledgements

The authors are grateful for valuable comments from Helena Skyt Nielsen, Marc Gurgand, and Peter Fredriksson as well as participants at the 2016 ESPE conferences and seminar participants at Aarhus University.

3.6 References

Akerlof, G.A. and R.E. Kranton. 2000. Economics and Identity. *Quarterly Journal of Economics* 115: 715–753.

Akerlof, G.A. and R.E. Kranton. 2002. Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature* 40 (4): 1167–1201.

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

Altonji, J.G. and R.M. Blank. 1999. Race and gender in the labor market. In: Ashenfelter, O.C. and D. Card (Eds.) *Handbook of Labor Economics* 3: 3143–3259.

Altonji, J., P. Arcidiacono and A. Maurel. 2015. The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. *NBER Working Paper* No. 21655.

Antecol, H., and D. Cobb-Clark. 2013. Do psychosocial traits help explain gender segregation in young people's occupation? *Labour Economics* 21: 59–73.

Autor, D., D. Figlio, K. Karbownik, J. Roth and M. Wasserman. 2015. Family disadvantage and the gender gap in behavioral and educational outcomes. *Northwestern Working Paper Series* WP-15-16.

Averett, S., L. Argys and D. Rees. 2011. Older siblings and adolescent risky behavior: does parenting play a role? *Journal of Population Economics* 24 (3): 957–978.

Bertrand, M. 2011. New perspectives on gender. In: Ashenfelter, O.C. and D. Card (eds.) *Handbook of Labor Economics* 4: 1543-1590.

Bertrand, M., E. Kamenica, and J. Pan. 2015. Gender identity and relative income within households. *Quarterly Journal of Economics* 130 (2): 571-614.

Björklund, A. and K.G. Salvanes. 2011. Education and Family background: Mechanisms and Policies, Ch. 3 in Hanushek, E.A., S. Machin and L. Woessmann, eds., *Handbook of the Economics of Education*, Volume 3: 201–247.

Black , S.E., P.J. Devereux and K.G. Salvanes. 2007. From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *Quarterly Journal of Economics* 122 (1): 409–439.

Black, S.E. and P.J. Devereux. 2011. Recent developments in intergenerational mobility, Ch. 16 in Ashenfelter, O. and D. Card, eds., *Handbook of Labor Economics*, Volume 4b: 1487–1541.

Blau, F.D. and L.M. Kahn. 2006. The U.S. Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review* 60 (1): 45-66.

Blau, F.D., L.M. Kahn, A.Y. Liu and K.L. Papps. 2013. The transmission of women's fertility, human capital and work orientation across immigrant generations. *Journal of Population Economics* 26: 405–435.

Bonin, H., T. Dohmen, A. Falk, D. Huffman and U. Sunde. 2007. Cross sectional earnings risk and occupational sorting: the role of risk attitudes. *Labour Economics* 14 (6), 926-937.

Booth, A., L. Cardona-Sosa and P. Nolen. 2014. Gender Differences in Risk Aversion: Do Single-Sex Environments Affect their Development. *Journal of Economic Behavior and Organization* 99: 126 – 154.

Booth, A. and P. Nolen. 2012a. Gender Differences in Risk Behaviour: Does Nurture Matter? *The Economic Journa*l 122: F56-F78.

Booth, A. and P. Nolen. 2012b. Choosing to compete: How different are girls and boys? *Journal of Economic Behaviour and Organisation* 81: 542 -555.

Bordalo, P., K. Coffman, N. Gennaioli and A. Shleifer. 2016. Stereotypes. Forthcoming in *Quarterly Journal of Economics.*

Burt, K.B. and J. Scott. 2002. Parent and Adolescent Gender Role Attitudes in 1990s Great Britain. *Sex Roles* 46(7): 239-45.

Coen-Pirani, D., A. León and S. Lugauer. 2010. The effect of household appliances on female labor force participation: Evidence from microdata. *Labour Economics* 17 (3): 503–513.

Corak, M. and P. Piraino. 2011. The intergenerational transmission of employers. *Journal of Labor Economics* 29 (1): 37-68.

Datta Gupta, N., N. Smith and M. Verner. 2008. The impact of Nordic countries' family friendly policies on employment, wages, and children. *Review of Economics of the Household* 6 (1): 65-89.

Dolado, J., F. Felfueroso and J. Jimeno. 2002. Recent Trends in Occupational Segregation by Gender: A Look Across the Atlantic. *IZA working paper* No. 524.

Dunn, T. and D. Holtz-Eakin. 2000. Financial Capital, Human Capital and the Transition to Self-Employment: Evidence from Intergenerational Links. *Journal of Labor Economics* 18 (2): 287–305.

Eriksson, K.H. 2015. Occupational segregation by sex: The role of intergenerational transmission. *Working paper*, Stockholm.

Eriksson, T., N. Smith and V. Smith. 2016. Gender stereotyping and self-stereotyping attitudes among managers. Prevalence and consequences. *EALE 2016 Conference paper.*

Fernández, R. 2013. Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review* 103(1): 472–500.

Fernández, R. and A. Fogli. 2009. Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics* 1(1): 147–177.

Fernández, R., A. Fogli and C. Olivetti. 2004. Mothers and sons: Preference formation and female labor force dynamics. The *Quarterly Journal of Economics* 119(4): 1249–1299.

Fogli, A., and L. Veldkamp. 2011. Nature or nurture? Learning and the geography of female labor force participation. *Econometrica* 79 (4): 1103–1138.

Fortin, N. 2005. Gender role attitudes and the labour-market outcomes of women across the OECD countries. *Oxford Review of Economic Policy* 21(3): 416-438.

Fryer, R.G. and S.D. Levitt. 2010. An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics* 2: 210–240.

Goldin, C. 2014a. A Grand Gender Convergence: Its Last Chapter. *American Economic Review* 104 (4): 1091–1119.

Goldin, C. 2014b. A Pollution Theory of Discrimination. Chapter 9 in L. Platt Boustan, C. Frydman and R.A. Margo (eds.), *Human Capital in History: The American Record.* University of Chicago Press, Chicago.

Goldin, C. and L.M. Katz. 2002. The power of the pill: Oral contraception and women's career and marriage decisions. *Journal of Political Economy* 100: 730–770.

Guiso, L, F. Monte, P Sapienza, and L. Zingales. 2008. Culture, gender and math. *Science* 320 (5880): 1164–1165.

Haan, D. 2010. Birth order, family size and educational attainment. *Economics of Education Review* 29: 576–588.

Hellerstein, J.K. and M.S. Morrill. 2011. Dads and daughters: The changing impact of fathers on women's occupational choices: *Journal of Human Resources* 46(2): 333–372.

Humlum, M.K., K.J. Kleinjans and H.S. Nielsen. 2012. An economic analysis of identity and career choice. *Economic Inquiry* 50 (1): 39–61.

Johnston, D.W., S. Schurer, and M.A. Shields. 2014. Maternal gender role attitudes, human capital investment, and labour supply of sons and daughters. *Oxford Economic Papers* 66: 631–659.

Kirkeboen, L., E. Leuven and M. Mogstad. 2016. Field of study, earnings and self-selection. Forthcoming in *Quarterly Journal of Economics.*

Laband, D.N. and B.F. Lentz. 1992. Self-recruitment in the legal profession. *Journal of Labor Economics* 10(2): 182–201.

Lehmann, J., A. Nuevo-Chiquero and M. Vidal-Fernandez. 2014. The early origins of birth order differences in children's outcomes and parental behavior. *Working paper* (previous version: *IZA Discussion Paper* No. 6755).

Lentz, B.F. and D.N. Laband. 1989. Why so many children of doctors become doctors: Nepotism vs. human capital transfers. *Journal of Human Resources* 24 (3): 396–413.

Lesner, R.V. 2016. Testing for Statistical Discrimination based on Gender. *Economics Working Papers* 2016–7, Aarhus University.

Lindquist, M.J., J. Sol and M. Van Praag. 2015. Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics* 33(2): 269-296.

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

Maccoby, E.E. 1998. The two sexes: growing up apart, coming together. *Havard University Press*, Cambridge.

Maestripieri, D., P. Sapienza and L. Zingales. 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Science* 106(36): 15268–15273.

Oguzoglu, U., and S. Ozbeklik. 2016. Like father, like daughter (unless there is a son): sibling sex composition and women's STEM major choice in college. *IZA Discussion Paper* No. 10052.

Olivetti, C. and B. Petrongolo. 2016. The evolution of gender gaps in industrialized countries. Forthcoming in *Annual Review of Economics*.

Pan, J. 2015. Gender segregation in occupations: The role of tipping and social interactions. *Journal of Labor Economics* 33 (2): 365–408.

Polachek, S. 1981. Occupational self-selection: A human capital approach to sex differences in occupational structure. The *Review of Economics and Statistics* 63(1): 60–69.

Reuben, E., M. Wiswall and B. Zafar. 2015. Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. Forthcoming in *The Economic Journal*.

Ruef, M., H. Aldrich and N. Carter. 2003. The structure of founding teams: homophily, strong ties, and isolation among U.S. entrepeneurs. *American Sociological Review* 68: 195–222.

SFI. 2016. Et kønsopdelt arbejdsmarked (Larsen, M., H. Holt and M.R. Larsen). *Det Nationale Forskningscenter for Velfærd (SFI)*, Copenhagen.

Thomas, D. 1994. Like father, like son; like mother, like daughter: Parental resources and child height. *The Journal of Human Resources* 29 (4): 950–988.

Waldfogel, J. 1998. Understanding the "Family Gap" in Pay for Women with Children. *The Journal of Economics Perspectives* 12 (1): 137–156.

Weinberg, B.A. 2000. Computer use and the demand for female workers. *Industrial and Labor Relations Review* 53: 290–308.

World Economic Forum. 2016. The Global Gender Gap Report 2014. Available at: http://www.weforum.org/reports/global-gender-gap-report-2014.

Zafar, B. 2013. College major choice and the gender gap. *Journal of Human Resources* 48 (3): 545–595.

Appendix

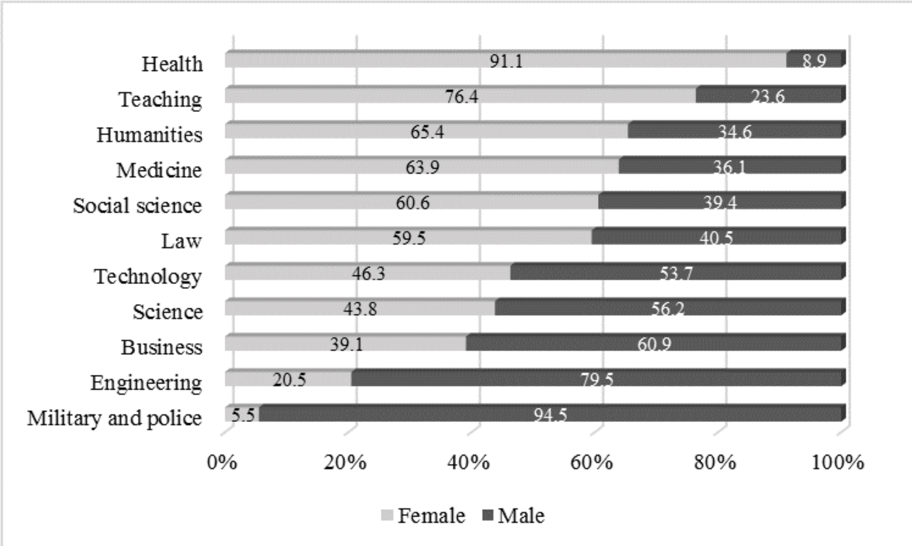

Figure A.1: Gender composition of BA graduates across fields.
This figure illustrates the gender composition of BA graduates (before or at age 28) across fields of study in the estimation sample. The sample covers birth cohorts 1970–1986.

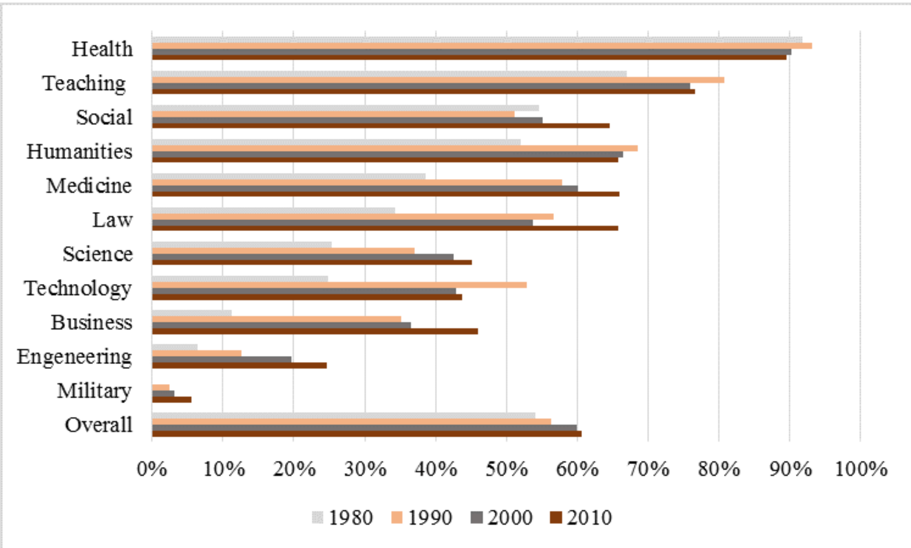

Figure A.2: Gender composition of graduates across fields, 1980–2010.
This figure illustrates the gender composition of graduates across fields of study in the period 1980–2010. University Bachelor's level was not formally introduced until the Bologna process in the 1990s. Therefore, both university Master's and Bachelor's programs are included in the figures for 2000 and 2010, while university college professionel Bachelor's and university Master's are included for 1980 and 1990.

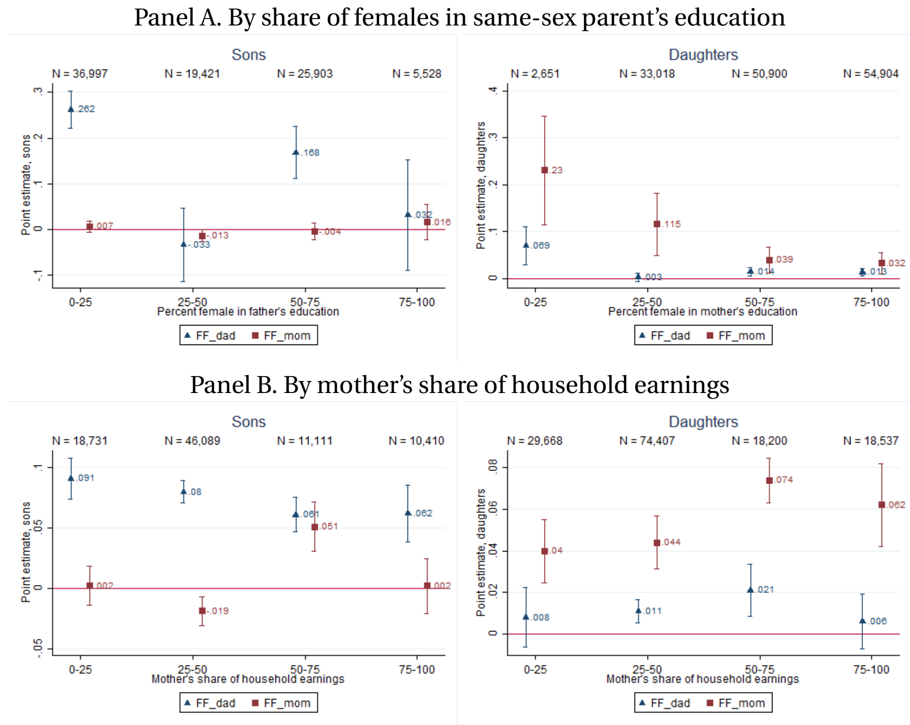

Figure A.3: The nonlinearities of the intergenerational correlation coefficients in female-dominated choice of education for males (left) and females (right).
Samples include individuals of Danish ancestry with at least a BA at age 28. X-axes denote the fraction of females in the education of the same-sex parent only (Panel A) and mother's share of household earnings (Panel B). Sample sizes are shown at the top; Table A.3 summarizes all subsample means of the dependent variable.

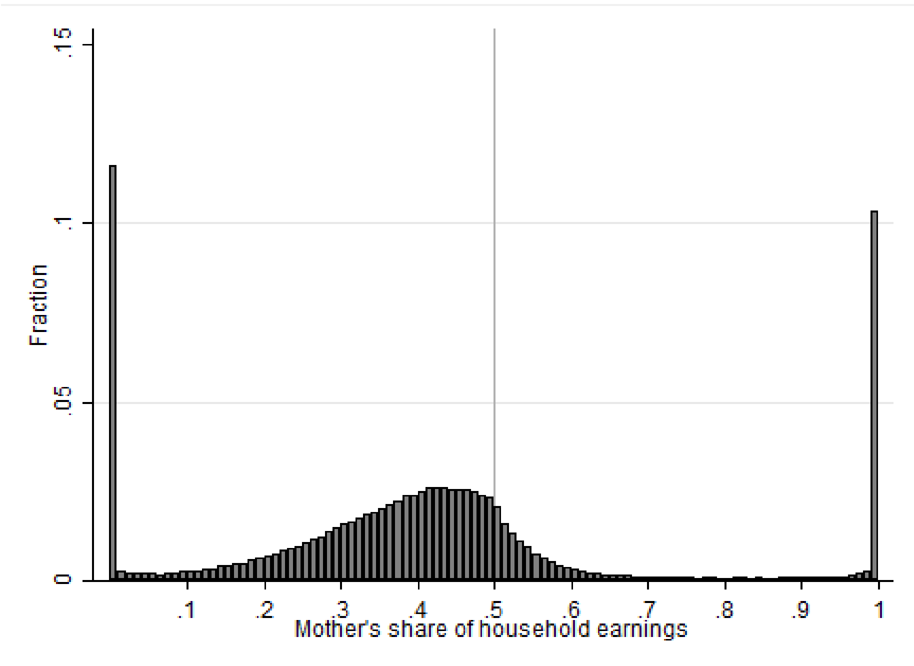

Figure A.4: Distribution of mother's share of household earnings in BA sample.
Strictly negative earnings are excluded, while observations with zero total household earnings are set to zero.

Table A.1: Determinants of age 30 labor market outcomes: Logearnings and unemployment status, and high school outcomes: High school GPA, taking advanced level math, and advanced level math grade

	Labor market outcomes, age 30				High school outcomes					
VARIABLES	(1) Male Logearn.	(2) Male Unempl. (‰)	(3) Female Logearn.	(4) Female Unempl. (‰)	(5) Male GPA	(6) Male Choose adv. math	(7) Male Grade, adv. math	(8) Female GPA	(9) Female Choose adv. math	(10) Female Grade, adv. math
Frac female, mother's educ	−0.002	−2.843	0.021	−8.852**	−0.358***	−0.072***	−0.637***	−0.458***	−0.121***	−0.602***
	(0.023)	(4.918)	(0.022)	(3.003)	(0.024)	(0.017)	(0.069)	(0.019)	(0.010)	(0.054)
Frac female, father's educ	−0.062***	1.312	−0.032	3.458**	0.066***	−0.160***	0.111**	0.076***	−0.094***	0.059
	(0.016)	(2.583)	(0.019)	(1.292)	(0.015)	(0.011)	(0.048)	(0.012)	(0.007)	(0.052)
High school GPA	0.056***	−3.859***	0.024***	4.305***						
	(0.008)	(1.063)	(0.007)	(0.754)						
Constant	11.053***	64.254***	14.242***	−75.220	10.085***	−1.283***	7.121***	5.624***	0.971***	7.693***
	(0.164)	(17.388)	(0.258)	(66.874)	(0.187)	(0.262)	(1.389)	(0.658)	(0.096)	(0.765)
Observations	48,918	51,046	78,788	82,861	69,279	37,419	21,657	123,948	71,142	21,405
R-squared	0.028	0.017	0.016	0.020	0.060	0.042	0.056	0.102	0.032	0.087
Birth year and region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes. Samples include individuals of Danish ancestry with at least a BA at age 28. Selected variables are shown, see Table 3.2 for a full list of included covariates. Columns (1)–(4) includes individuals born in 1970–1980 only. In columns (1) and (3) only individuals with positive earnings are included. Unemployment is denoted in per mille. Course-specific information and grades are available from graduation year 1997 onwards, thus, columns (6)–(7) and (9)–(10) mostly individuals born in 1978–1986. Columns Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Determinants of female-dominated field of study: fraction female, men and women separately in full specification

	(1) Male, no GPA		(2) Male, baseline		(3) Male, not same educ		(4) Female, no GPA		(5) Female, baseline		(6) Female, not same educ	
VARIABLES	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Frac female, mother's educ	-0.002	(0.004)	-0.004	(0.004)	0.000	(0.004)	0.081 ***	(0.006)	0.054 ***	(0.005)	0.028 ***	(0.005)
Frac female, father's educ	0.088 ***	(0.004)	0.084 ***	(0.004)	0.058 ***	(0.004)	0.008 ***	(0.002)	0.013 ***	(0.003)	0.009 ***	(0.003)
High school GPA			-0.018 ***	(0.002)	-0.013 ***	(0.002)			-0.064 ***	(0.002)	-0.063 ***	(0.002)
Born in second quarter	-0.004 *	(0.002)	-0.003 *	(0.002)	-0.004 *	(0.002)	-0.002	(0.001)	-0.001	(0.001)	-0.001	(0.001)
— third quarter	-0.002	(0.002)	-0.001	(0.002)	-0.001	(0.002)	-0.004 **	(0.002)	-0.003 **	(0.001)	-0.003 *	(0.002)
— fourth quarter	0.000	(0.002)	0.001	(0.002)	0.002	(0.002)	-0.005 **	(0.002)	-0.004 **	(0.002)	-0.004 **	(0.002)
Low birthweight (< 2500 g)	0.019 ***	(0.005)	0.019 ***	(0.004)	0.019 ***	(0.004)	0.010 ***	(0.003)	0.008 **	(0.003)	0.009 ***	(0.003)
Firstborn	-0.004	(0.003)	-0.005	(0.003)	-0.004	(0.003)	-0.013 ***	(0.002)	-0.010 ***	(0.002)	-0.011 ***	(0.002)
Multiple born	-0.005	(0.005)	-0.005	(0.005)	-0.002	(0.005)	-0.002	(0.005)	0.000	(0.005)	0.001	(0.006)
No. of siblings (by mother)	-0.005 ***	(0.001)	-0.005 ***	(0.001)	-0.005 ***	(0.001)	-0.001	(0.001)	0.001	(0.001)	0.000	(0.001)
No. of older sisters	0.011 ***	(0.003)	0.012 ***	(0.003)	0.011 ***	(0.002)	0.012 ***	(0.002)	0.006 ***	(0.002)	0.006 ***	(0.002)
No. of older brothers	0.019 ***	(0.002)	0.019 ***	(0.003)	0.019 ***	(0.002)	0.014 ***	(0.002)	0.007 ***	(0.002)	0.006 ***	(0.002)
Mother's logearnings	-0.001	(0.001)	-0.002	(0.001)	-0.003 *	(0.001)	-0.005 ***	(0.001)	-0.004 ***	(0.001)	-0.005 ***	(0.001)
Mother zero earnings	-0.013	(0.014)	-0.020	(0.014)	-0.027 *	(0.014)	-0.058 ***	(0.008)	-0.043 ***	(0.009)	-0.050 ***	(0.009)
Mother works < 30 hrs/week	-0.009 ***	(0.002)	-0.007 ***	(0.002)	-0.005 *	(0.002)	0.000	(0.001)	0.003 **	(0.001)	0.000	(0.001)
Mother outside labor market	-0.011 ***	(0.004)	-0.011 ***	(0.003)	-0.010 **	(0.003)	-0.010 ***	(0.002)	-0.006 **	(0.002)	-0.007 ***	(0.002)
Mother's age	-0.007 ***	(0.000)	-0.007 ***	(0.000)	-0.007 ***	(0.000)	-0.003 ***	(0.000)	-0.001	(0.000)	-0.001	(0.001)
Mother's education:												
—None/missing	0.008	(0.010)	0.007	(0.011)	0.008	(0.011)	-0.006	(0.008)	-0.001	(0.007)	-0.002	(0.007)
—Vocational	-0.005	(0.003)	-0.006	(0.003)	-0.006	(0.003)	-0.025 ***	(0.003)	-0.017 ***	(0.003)	-0.011 ***	(0.002)
—Higher	0.018 ***	(0.004)	0.014 ***	(0.004)	-0.002	(0.004)	-0.048 ***	(0.003)	-0.024 ***	(0.003)	-0.030 ***	(0.003)
Separated parents	0.014 ***	(0.002)	0.015 ***	(0.002)	0.016 ***	(0.002)	0.008 ***	(0.002)	0.006 ***	(0.001)	0.006 ***	(0.002)
Father's logearnings	-0.007 ***	(0.001)	-0.007 ***	(0.001)	-0.006 ***	(0.001)	-0.007 ***	(0.001)	-0.005 ***	(0.001)	-0.005 ***	(0.001)
Father zero earnings	-0.078 ***	(0.012)	-0.084 ***	(0.012)	-0.065 ***	(0.013)	-0.074 ***	(0.008)	-0.060 ***	(0.007)	-0.054 ***	(0.008)
Father works < 30 hrs/week	0.010	(0.009)	0.008	(0.009)	0.008	(0.010)	-0.018 ***	(0.003)	-0.015 ***	(0.003)	-0.015 ***	(0.003)
Father outside labor force	-0.019 ***	(0.004)	-0.020 ***	(0.004)	-0.019 ***	(0.004)	-0.017 ***	(0.003)	-0.012 ***	(0.003)	-0.011 ***	(0.003)
Father's age	-0.012 ***	(0.001)	-0.011 ***	(0.001)	-0.010 ***	(0.001)	-0.002 ***	(0.000)	-0.002 ***	(0.000)	-0.002 ***	(0.000)
Father's education:												
—None/missing	-0.025 *	(0.012)	-0.024 *	(0.012)	-0.026 *	(0.012)	-0.002	(0.009)	0.002	(0.008)	0.002	(0.009)
—Vocational	0.009 ***	(0.002)	0.009 ***	(0.002)	0.002	(0.002)	0.007 ***	(0.001)	0.007 ***	(0.002)	0.005 ***	(0.001)
—Higher	0.006 *	(0.003)	0.000	(0.003)	-0.006	(0.004)	-0.035 ***	(0.003)	-0.021 ***	(0.003)	-0.019 ***	(0.003)
Mother died before BA	0.002	(0.013)	0.002	(0.013)	0.002	(0.013)	0.033 ***	(0.005)	0.026 ***	(0.005)	0.025 ***	(0.004)
Father died before BA	0.021 **	(0.008)	0.024 ***	(0.008)	0.022 **	(0.008)	0.021 ***	(0.003)	0.012 ***	(0.003)	0.013 ***	(0.003)
Constant	1.257 ***	(0.048)	1.413 ***	(0.051)	1.351 ***	(0.053)	1.021 ***	(0.027)	1.414 ***	(0.034)	1.425 ***	(0.035)
Observations	86,297		86,297		82,587		140,745		140,745		132,210	
R-squared	0.04		0.067		0.063		0.054		0.121		0.119	
Birth year & region FE	YES		YES		YES		YES		YES		YES	

Notes. Samples include individuals of Danish ancestry with at least a BA at age 28. Missing indicators are omitted. Excluded categories are mother/father working ≥30 hrs/week and mother/father having basic education level. Columns (3) and (6) exclude individuals with the same education as either parent. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: Mean share of female graduates in completed BA by same-sex parent's educational characteristics and mother's share of household earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Share of female graduates in same-sex parent's education				Mother's share of household earnings			
	0–25	25–50	50–75	75–100	0–25	25–50	50–75	75–100
Males	0.403	0.443	0.437	0.463	0.403	0.428	0.448	0.418
Females	0.633	0.720	0.699	0.716	0.707	0.710	0.704	0.716

Notes. Columns (1)–(4) present the mean share of female graduates for the education of male and female cohort members by the share of females in father's education for males and in mother's education for females. Number of observations are shown in Fig. A.3.

Table A.4: Sensitivity checks

	(1) Men first choice BA	(2) Men GPA > 9	(3) Men GPA > 10	(4) Women first choice BA	(5) Women GPA > 9	(6) Women GPA > 10
Frac female, mother's educ	0.002	-0.005	0.004	0.047***	0.040***	0.044**
	(0.004)	(0.007)	(0.016)	(0.005)	(0.006)	(0.018)
Frac female, father's educ	0.073***	0.058***	0.052***	0.013***	0.024***	0.042***
	(0.004)	(0.008)	(0.014)	(0.002)	(0.005)	(0.013)
High school GPA	-0.017***	0.006	0.009	-0.063***	-0.035***	-0.011
	(0.003)	(0.003)	(0.008)	(0.002)	(0.004)	(0.011)
Mean dependent variable	0.446	0.445	0.45	0.701	0.638	0.599
Observations	106,098	24,257	4,651	163,714	36,971	5,817
R-squared	0.045	0.046	0.104	0.112	0.048	0.104
Birth year & region FE	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES

Notes. Columns (1) and (4) use the fraction of females in first choice of enrolment if that was a BA program as outcome variables. Columns (2)–(3) and (5)–(6) restrict samples based on high school GPA. Selected variables are shown, see Table 3.2 for a full list of included covariates. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Source-country characteristics as determinants for female-dominated education choice, second-generation immigrants born in 1975–1986

	(1) All	(2) Men	(3) Men	(4) Men	(5) Women	(6) Women	(7) Women
Female	0.247***						
	(0.007)						
High school GPA	-0.029***		0.006	0.010		-0.050***	-0.047***
	(0.004)		(0.009)	(0.012)		(0.003)	(0.004)
Source-country proxies, age 15 (father)							
Total fertility rate (births per woman)	-0.004	-0.004	-0.001	-0.001	0.001	-0.003	-0.003
	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)
Labor force participation rate ratio	0.006	0.006	0.004	0.005	0.010**	0.006	0.007
	(0.005)	(0.009)	(0.008)	(0.009)	(0.004)	(0.005)	(0.007)
Observations	2,955	1,143	1,143	1,143	1,812	1,812	1,812
R-squared	0.329	0.02	0.108	0.218	0.013	0.144	0.201
Birth year	YES	YES	YES	YES	YES	YES	YES
Parent birth year FE	YES	NO	YES	YES	NO	YES	YES
Region FE	YES	NO	NO	YES	NO	NO	YES
Covariates	YES	NO	YES	YES	NO	YES	YES

Notes. Source country information is available from 1990 onwards and can be matched to 201 Statistics Denmark country codes, see https://data.worldbank.org/indicator/ and http://www.ilo.org/global/statistics-and-databases. Ninety percent of the second-generation immigrants have both parents originating from the same country, thus multicollinearity prevents us from including source country measure for both mothers and fathers. Labor force participation rate ratios are defined as male labor force participation rate for 35-44-year olds divided by female labor force participation rate for the same age group. See table for included covariates. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.6: Testing for a confounding relationship between parental female-dominated education choice and the birth-order gap in birthweight

	Male				Female				All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Outcome: birthweight (g)	2+ families	2+ families	Brother OLS	Brother FE	2+ families	2+ families	Sisters OLS	Sisters FE	Siblings FE	Siblings FE
Female									-139.391***	-139.391***
									(5.474)	(5.474)
Firstborn	-116.819***	-90.685***	-90.370**	-75.670*	-109.055***	-61.855***	-51.435*	-61.348*		
	(17.694)	(20.022)	(30.385)	(40.537)	(16.483)	(13.983)	(23.899)	(29.124)		
Frac female, mother's educ	-21.822	-43.237**	-75.408**	-32.181	29.675**	-48.547***	-33.962	-1.566	-49.623	
	(18.119)	(15.975)	(34.759)	(152.213)	(12.323)	(12.383)	(19.569)	(149.481)	(92.260)	
Frac female, father's educ	-11.104	-18.786	5.483	-253.016	-10.488	-7.650	-32.208*	-76.873	-156.237*	
	(11.107)	(10.608)	(16.615)	(145.067)	(11.668)	(9.330)	(17.911)	(162.435)	(78.765)	
EarningsGapMom										21.533
										(15.037)
EarningsGapDad										-8.671
										(4.949)
Interaction terms										
Firstborn × FFMom	11.880	-13.332	26.539	16.079	19.229	-6.552	-36.683	-26.325		
	(20.591)	(21.468)	(47.857)	(29.980)	(21.017)	(18.399)	(30.746)	(33.343)		
Firstborn × FFDad	2.788	-1.367	-15.047	-18.825	3.648	-19.865*	9.982	-0.052		
	(15.693)	(14.106)	(25.575)	(40.923)	(13.819)	(9.810)	(22.185)	(18.644)		
Female × FFMom									-6.383	
									(27.142)	
Female × FFDad									-0.622	
									(16.567)	
Female × EarningsGapMom										-11.458
										(12.889)
Female × EarningsGapDad										-2.656
										(3.997)
Observations	67,216	67,216	15,446	15,446	111,228	111,228	28,849	28,849	76,818	78,335
R-squared	0.071	0.130	0.212	0.786	0.006	0.126	0.188	0.768	0.754	0.754
Birth year & region FE	NO	YES	YES	YES	NO	YES	YES	YES	YES	YES
Parental FE	NO	NO	NO	YES	NO	NO	NO	YES	YES	YES

Notes. Selected variables are shown. Columns (1) and (5) include only the shown regressors; other specifications further include indicators for quarter of birth and number of older sisters and older brothers. All regressions are carried out on the subsamples of men and women with at least one sibling. Standard errors corrected for clustering within birth year in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS

AARHUS UNIVERSITY
SCHOOL OF BUSINESS AND SOCIAL SCIENCES
www.econ.au.dk

PhD dissertations since 1 July 2011

2011-4 Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics

2011-5 Christian Bach: The Game of Risk

2011-6 Stefan Holst Bache: Quantile Regression: Three Econometric Studies

2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management

2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization

2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders

2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade

2012-1 Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting

2012-2 Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping

2012-3 Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing

2012-4 Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing

2012-5 Lasse Frisgaard Gunnersen: Income Redistribution Policies

2012-6 Miriam Wüst: Essays on early investments in child health

2012-7 Yukai Yang: Modelling Nonlinear Vector Economic Time Series

2012-8 Lene Kjærsgaard: Empirical Essays of Active Labor Market Policy on Employment

2012-9 Henrik Nørholm: Structured Retail Products and Return Predictability

2012-10 Signe Frederiksen: Empirical Essays on Placements in Outside Home Care

2012-11 Mateusz P. Dziubinski: Essays on Financial Econometrics and Derivatives Pricing

2012-12 Jens Riis Andersen: Option Games under Incomplete Information

2012-13 Margit Malmmose: The Role of Management Accounting in New Public Management Reforms: Implications in a Socio-Political Health Care Context

2012-14 Laurent Callot: Large Panels and High-dimensional VAR

2012-15 Christian Rix-Nielsen: Strategic Investment

2013-1 Kenneth Lykke Sørensen: Essays on Wage Determination

2013-2 Tue Rauff Lind Christensen: Network Design Problems with Piecewise Linear Cost Functions

2013-3 Dominyka Sakalauskaite: A Challenge for Experts: Auditors, Forensic Specialists and the Detection of Fraud

2013-4 Rune Bysted: Essays on Innovative Work Behavior

2013-5 Mikkel Nørlem Hermansen: Longer Human Lifespan and the Retirement Decision

2013-6 Jannie H.G. Kristoffersen: Empirical Essays on Economics of Education

2013-7 Mark Strøm Kristoffersen: Essays on Economic Policies over the Business Cycle

2013-8 Philipp Meinen: Essays on Firms in International Trade

2013-9 Cédric Gorinas: Essays on Marginalization and Integration of Immigrants and Young Criminals – A Labour Economics Perspective

2013-10 Ina Charlotte Jäkel: Product Quality, Trade Policy, and Voter Preferences: Essays on International Trade

2013-11 Anna Gerstrøm: World Disruption - How Bankers Reconstruct the Financial Crisis: Essays on Interpretation

2013-12 Paola Andrea Barrientos Quiroga: Essays on Development Economics

2013-13 Peter Bodnar: Essays on Warehouse Operations

2013-14 Rune Vammen Lesner: Essays on Determinants of Inequality

2013-15 Peter Arendorf Bache: Firms and International Trade

2013-16 Anders Laugesen: On Complementarities, Heterogeneous Firms, and International Trade

2013-17 Anders Bruun Jonassen: Regression Discontinuity Analyses of the Disincentive Effects of Increasing Social Assistance

2014-1 David Sloth Pedersen: A Journey into the Dark Arts of Quantitative Finance

2014-2 Martin Schultz-Nielsen: Optimal Corporate Investments and Capital Structure

2014-3 Lukas Bach: Routing and Scheduling Problems - Optimization using Exact and Heuristic Methods

2014-4 Tanja Groth: Regulatory impacts in relation to a renewable fuel CHP technology: A financial and socioeconomic analysis

2014-5 Niels Strange Hansen: Forecasting Based on Unobserved Variables

2014-6 Ritwik Banerjee: Economics of Misbehavior

2014-7 Christina Annette Gravert: Giving and Taking – Essays in Experimental Economics

2014-8 Astrid Hanghøj: Papers in purchasing and supply management: A capability-based perspective

2014-9 Nima Nonejad: Essays in Applied Bayesian Particle and Markov Chain Monte Carlo Techniques in Time Series Econometrics

2014-10 Tine L. Mundbjerg Eriksen: Essays on Bullying: an Economist's Perspective

2014-11 Sashka Dimova: Essays on Job Search Assistance

2014-12 Rasmus Tangsgaard Varneskov: Econometric Analysis of Volatility in Financial Additive Noise Models

2015-1 Anne Floor Brix: Estimation of Continuous Time Models Driven by Lévy Processes

2015-2 Kasper Vinther Olesen: Realizing Conditional Distributions and Coherence Across Financial Asset Classes

2015-3 Manuel Sebastian Lukas: Estimation and Model Specification for Econometric Forecasting

2015-4 Sofie Theilade Nyland Brodersen: Essays on Job Search Assistance and Labor Market Outcomes

2015-5 Jesper Nydam Wulff: Empirical Research in Foreign Market Entry Mode

2015-6 Sanni Nørgaard Breining: The Sibling Relationship Dynamics and Spillovers

2015-7 Marie Herly: Empirical Studies of Earnings Quality

2015-8 Stine Ludvig Bech: The Relationship between Caseworkers and Unemployed Workers

2015-9 Kaleb Girma Abreha: Empirical Essays on Heterogeneous Firms and International Trade

2015-10 Jeanne Andersen: Modelling and Optimisation of Renewable Energy Systems

2015-11 Rasmus Landersø: Essays in the Economics of Crime

2015-12 Juan Carlos Parra-Alvarez: Solution Methods and Inference in Continuous-Time Dynamic Equilibrium Economies (with Applications in Asset Pricing and Income Fluctuation Models)

2015-13 Sakshi Girdhar: The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis

2015-14 Wenjing Wang: Corporate Innovation, R&D Personnel and External Knowledge Utilization

2015-15 Lene Gilje Justesen: Empirical Banking

2015-16 Jonas Maibom: Structural and Empirical Analysis of the Labour Market

2015-17 Sylvanus Kwaku Afesorgbor: Essays on International Economics and Development

2015-18 Orimar Sauri: Lévy Semistationary Models with Applications in Energy Markets

2015-19 Kristine Vasiljeva: Essays on Immigration in a Generous Welfare State

2015-20 Jonas Nygaard Eriksen: Business Cycles and Expected Returns

2015-21 Simon Juul Hviid: Dynamic Models of the Housing Market

2016-1 Silvia Migali: Essays on International Migration: Institutions, Skill Recognition, and the Welfare State

2016-2 Lorenzo Boldrini: Essays on Forecasting with Linear State-Space Systems

2016-3 Palle Sørensen: Financial Frictions, Price Rigidities, and the Business Cycle

2016-4 Camilla Pisani: Volatility and Correlation in Financial Markets: Theoretical Developments and Numerical Analysis

2016-5 Anders Kronborg: Methods and Applications to DSGE Models

2016-6 Morten Visby Krægpøth: Empirical Studies in Economics of Education

2016-7 Anne Odile Peschel: Essays on Implicit Information Processing at the Point of Sale: Evidence from Experiments and Scanner Data Analysis

2016-8 Girum Dagnachew Abate: Essays in Spatial Econometrics

2016-9 Kai Rehwald: Essays in Public Policy Evaluation

2016-10 Reza Pourmoayed: Optimization Methods in a Stochastic Production Environment

2016-11 Sune Lauth Gadegaard: Discrete Location Problems – Theory, Algorithms, and Extensions to Multiple Objectives

2016-12 Lisbeth Palmhøj Nielsen: Empirical Essays on Child Achievement, Maternal Employment, Parental Leave, and Geographic Mobility

2016-13 Louise Voldby Beuchert-Pedersen: School Resources and Student Achievement: Evidence From Social and Natural Experiments

2016-14 Mette Trier Damgaard: Essays in Applied Behavioral Economics

2016-15 Andrea Barletta: Consistent Modeling and Efficient Pricing of Volatility Derivatives

2016-16 Thorvardur Tjörvi Ólafsson: Macrofinancial Linkages and Crises in Small Open Economies

2016-17 Carlos Vladimir Rodríguez Caballero: On Factor Analysis with Long-Range Dependence

2016-18 J. Eduardo Vera-Valdés: Essays in Long Memory

2016-19 Magnus Sander: Returns, Dividends, and Optimal Portfolios

2016-20 Ioana Daniela Neamtu: Wind Power Effects and Price Elasticity of Demand for the Nordic Electricity Markets

2016-21 Anne Brink Nandrup: Determinants of Student Achievement and Education Choice

ISBN: 9788793195509