Playing the hand you're dealt

Citation for published version (APA):

Magnée, C. A. J. (2020). Playing the hand you're dealt: The effects of family structure on children's personality and the effects of educational policy on educational outcomes of migrant children. [Doctoral Thesis, Maastricht University]. ROA. <https://doi.org/10.26481/dis.20201009cm>

Document status and date:

Published: 01/01/2020

DOI:

[10.26481/dis.20201009cm](https://doi.org/10.26481/dis.20201009cm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Playing the hand you're dealt

The effects of family structure on children's personality and the effects of educational policy on educational outcomes of migrant children

Cécile A.J. Magnée

© Cécile Magnée, Maastricht 2020

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing, from the author.

Published by ROA
Postbus 616
6200 MD Maastricht

ISBN: 978-90-5321-595-1

Printed in the Netherlands by Canon
Image cover: Peter Soudant

Playing the hand you're dealt

The effects of family structure on children's personality and the effects
of educational policy on educational outcomes of migrant children

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public on
Friday 9 October 2020, at 12:00 hours

by

Cécile Alice Jeanne Magnée

Supervisors

Prof. dr. Inge de Wolf

Prof. dr. Bart Golsteyn

Assessment Committee

Prof. dr. Gerard Pfann (chair)

Prof. dr. Silke Anger, Institute for Employment Research, Germany

Dr. Carla Haelermans

Prof. dr. Rolf van der Velden

Prof. dr. Dinand Webbink, Erasmus University Rotterdam

Acknowledgements

The perfect recipe for a successful PhD is the right combination of support, love, advise, constructive criticism and laughter. I was very lucky that I received all of this over the past few years, allowing me to explore my curiosities and develop my talents. I would like to take a moment to thank all of you.

Inge and Bart, thank you for letting me pick your brains, for pushing me to take a step further and for giving me the freedom to design my PhD in a way that fitted me best. Bart, you were the one who sparked my interest in research back in the days when I wrote my Bachelor thesis with you. I have really enjoyed working with you: the nice combination of hard work and fun conversations on all different types of topics (but especially music), the conferences during which we gladly indulged in the good food and drinks, and the ‘my heart skipped a beat’ moments when we received mails from journals. I honestly don’t think I would have even considered doing a PhD if it wasn’t for you. Inge, I remember our first conversation when we brainstormed about interesting PhD topics and your enthusiasm and willingness to let me design a PhD that sparked my interest. The fact that you managed to get me a working space at the Dutch Inspectorate of Education and access to their data gave me the opportunity to experience a different working environment besides academia and made doing research a lot easier. Although things didn’t always go as planned, we managed to stick together as a team and made it to the finish line!

I would also like to thank the members of the reading committee, Gerard Pfann, Silke Anger, Carla Haelermans, Rolf van der Velden and Dinand Webbink, for taking the time to read my manuscript and provide me with valuable comments. Gerard, thank you for getting in touch with Christian Dustmann which made my research visit to UCL possible. Rolf, thank you for checking in on me from time to time to make sure I had everything I needed to continue my research. Furthermore, I would like to thank the Graduate School of Business and Economics of Maastricht University for the 3-year fulltime PhD grant, which made this whole journey possible. Peter and Miranda, thank you for helping me with the design of the front page and the design of the content of this thesis, respectively.

Given the fact that I had multiple working places during my PhD, I got very lucky to meet a bunch of intelligent, sweet and funny colleagues. First of all, thanks to the members of the writing group in Maastricht for all the support, not only academically, but also mentally. You always made the four-hour travel time back and forth to Maastricht worthwhile. Inge and Maja, thank you for being very dedicated to the writing group and making serious writing days feel like a fun day out. Second of all, thanks to my colleagues at the Dutch Inspectorate for helping me with data and policy questions. Esmee, Evelien, Frans and Jaap: thank you for

all the coffee and lunch breaks, your patience when I was on a rant and your great advice. Third of all, thanks to my colleagues at University College London. You made me feel like home from the very start and provided me with the feedback I needed to improve my research and take the final steps in writing up this thesis.

Finally, thanks to my family, friends and Louwie for your unwavering support throughout this journey. Whenever I was at my highest and lowest points, you were always there to listen to me and reassure me that it would all be fine. Mum and dad, thank you for supporting all the choices I made and making sure I had all the means to pursue my dreams. Now that I've read a lot of research on the importance of your childhood environment on later in life outcomes, I'm pretty sure you did a stellar job! Mum, thanks for always reading my work and giving me feedback. Dad, you were right after all about me getting a PhD, who would have thought? Sylvia and Luca, thank you for being my paranymphs, for all the discussions we've had over the years that improved my research, and for always being there when I needed you. Iris, Jules and Lara, thank you for being my Maastricht crew, for being silly but also for helping me in tougher times. Louwie, thank you for your excellent cooking skills and making sure there was always food around when I needed it, your ability to listen to me and give advice at the right times, and for always being able to make me laugh.

Contents

Chapter 1 Introduction	1
1.1 Motivation	2
1.2 Aim	4
1.3 Outline	5
Chapter 2 Does birth spacing affect personality?	9
2.1 Introduction	10
2.2 Data	13
2.3 Empirical Strategy	17
2.4 Results	21
2.5 Conclusions and policy implications	29
2.6 Appendix	31
Chapter 3 Does sibling gender affect personality traits?	41
3.1 Introduction	42
3.2 Materials and methods	44
3.2.1 Sample	44
3.2.2 Dependent variables	45
3.2.3 Socio-economic status (SES)	47
3.3 Methodology	47
3.4 Results	50
3.5 Consequences for life outcomes	54
3.6 Conclusion and discussion	59
3.7 Appendix	61
Chapter 4 How does the inability to observe primary school exit test scores affect the teacher's assessment of immigrants?	63
4.1 Introduction	64
4.2 Dutch institutional background	66
4.3 Conceptual framework	68
4.3.1 Pre policy change	68
4.3.2 Post policy change	69
4.3.3 Hypotheses	70
4.4 Empirical analysis	71
4.4.1 Selection check	72
4.5 Data	74

4.6 Results	75
4.6.1 Descriptive analyses	75
4.6.2 Empirical results	78
4.6.3 Robustness checks	82
4.7 Discussion	85
4.8 Appendix	87
Chapter 5 The effect of specialized refugee schools on the primary school outcomes of refugee children	89
5.1 Introduction	90
5.2 The Dutch educational context and placement policy	92
5.3 Data	94
5.4 Methodology	96
5.5 Results	98
5.6 Conclusion and discussion	102
5.7 Appendix	105
Chapter 6 Conclusions	107
Valorisation addendum	115
Bibliography	119
Curriculum Vitae	129
ROA Dissertation Series	131

List of Tables

Table 2.1	Descriptive statistics	14
Table 2.2	First stage regressions of birth gap on miscarriage	19
Table 2.3	OLS regression of the controls on miscarrying	20
Table 2.4	IV regressions of birth gap on personality	22
Table 2.5	IV regressions of birth gap on personality, varying gaps	23
Table 2.6	IV regressions of birth gap on personality, narrow definition	24
Table 2.7	IV regressions of birth gap on personality, boys	25
Table 2.8	IV regressions of birth gap on personality, girls	26
Table 2.9	IV regressions of birth gap on achievement tests, education and other life outcomes	28
Table A2.1	Panel attrition	31
Table A2.2	Miscarriages between first and second child (broad definition)	31
Table A2.3	Miscarriages between first and second child (narrow definition)	31
Table A2.4	Questions about Rutter score answered by the mother when the youngest child was 10 years old	32
Table A2.5	Questions about self-esteem answered by youngest child at age 10	32
Table A2.6	Questions about locus of control answered by youngest child at age 10	33
Table A2.7	Questions about disorganized, asocial, neurotic and introverted behavior answered by the teacher when the youngest child was 10 years old	33
Table A2.8	Smoking behavior mother during pregnancy of youngest child	34
Table A2.9	Social class husband at time of birth of youngest child	35
Table A2.10	Social class mother at time of birth of youngest child	35
Table A2.11	Questions about mother's wellbeing answered by her at age 5 of youngest child	35
Table A2.12	Examples of questions about mother's attitude toward child rearing and toward other views about life answered by her at age 5 of youngest child	36
Table A2.13	Miscarriages before the first child (broad definition)	36
Table A2.14	OLS regression of gap on personality traits of child	37
Table A2.15	First stage regressions of birth gap on miscarriage with control variables displayed	38
Table 3.1	Relationship between sibling gender and attrition at age 16	45
Table 3.2	Relationship between parental characteristics and the chance of having a third child	49
Table 3.2A	Balance test 1 child vs. 2 children	49

Table 3.2B	Balance test 2 children vs. 3 children	49
Table 3.3	Relationship between sibling gender and personality traits	51
Table 3.4	Relationship between sibling gender and personality for high and low SES categories	52
Table 3.4A	Results for boys	52
Table 3.4B	Results for girls	53
Table 3.5	Relationship between sibling gender and various life outcomes for high and low SES categories	55
Table 3.5A	Results for boys	55
Table 3.5B	Results for girls	57
Table A3.1	Construction of socio-economic status (SES) via social class of father	61
Table A3.2	Division over low and high SES per gender combination	61
Table 4.1	Description of which information is available for which students	73
Table 4.2	DiD estimation results	79
Table 4.3	Event study	81
Table 4.4	Robustness check 1: DiD estimation without 2014	82
Table 4.5	Robustness check 1: Event study estimation without 2014	83
Table 4.6	Robustness check 2: DiD estimation with revised teacher assessments	84
Table 4.7	Robustness check 2: Event study estimation with revised teacher assessments	85
Table A4.1	Test scores and track assessments	87
Table 5.1	Balance table refugees	97
Table 5.2	OLS regressions of enrollment at a refugee school on test scores of refugees	101
Table 5.3	OLS regressions of enrollment at a refugee school on teacher assessment of refugees	101
Table A5.1	Top 5 countries of origin of refugees housed by COA per arrival year	105
Table A5.2	Teacher track assessments	105
Table A5.3	Test on non-random sample attrition	106

List of Figures

Figure 2.1	Cumulative distribution of the birth gap	15
Figure 2.2	Histogram birth gap for miscarriage and no miscarriage	18
Figure A2.1	Mother's age at time of delivery of the first child	34
Figure 4.1	Mean test score trend	75
Figure 4.2	Mean teacher assessment trend	76
Figure 4.3	Mean difference (ET-T) trend	77
Figure A4.1	Dutch education system	87
Figure A4.2	Timing test score and teacher assessment pre- and post-policy change	88
Figure 5.1	Boxplots of test score per country of origin of refugee children	99
Figure 5.2	Division of teacher assessments per country of origin of refugee children	100

Chapter 1

Introduction

1.1 Motivation

The expression *playing the hand you're dealt* is commonly used in card games. It means that one can only play the cards one is given, hence one has to try to make the most out of these cards. This concept can be applied to everyday life as well; where and when one is born determine the cards that are dealt. Over time, more cards are handed, and especially during childhood it is difficult to influence the cards received.

This thesis investigates the consequences of two cards that are dealt to children during their childhood and the effects of these cards on their personality and educational outcomes. The first part of this thesis analyzes the effect of *family structure* on children's personality, and the second part studies the effect of *educational policy* on the educational outcomes of migrant children.

The family one is born in largely determines the environment that one grows up in. The family structure such as the number of siblings, the age difference between siblings, birth order, and sibling gender composition, can affect the interaction with one's parents and siblings. The effects of these interactions with sibling(s) and parents have been studied in a wide variety of fields.

Single-parents families have been associated with e.g. delinquent behavior (Griffin et al., 2000), worse (mental) health outcomes (Byrne et al. 2011; Watanabe et al. 2011; Brown 2006; Bzostek & Beck 2011; Harknett 2009; Langton & Berger 2011; Biehl et al. 2014; Chen & Escarce 2014), and worse educational attainment and cognition (Gennetian, 2005; Krein & Beller, 1988; McLanahan & Sandefur 1994). Furthermore, there might be differences in outcomes between boys and girls due to gender-differentiated parental behavior (McHale et al, 2003) and different sibling interactions with respect to gender (Conley, 2000). A strand of literate studies the effects of birth order on e.g. parenting styles (McHale et al. 2012), educational attainment (Hotz & Pantano, 2015) and intelligence (Barclay, 2015; Bjerkedal et al., 2007; Black et al., 2007). In addition, others have studied the effect of the age difference between siblings – i.e. the birth gap – on various outcomes, such as personality (Conde-Agudelo et al., 2016) and achievement tests (Buckles & Munnich, 2012).

The large amount of literature on the topic demonstrates the complexity and interrelationship between various factors of family structure, outcomes and potential mechanisms. Given the fact that the interaction with siblings and parents can influence behavior and other outcomes of both the child and the other family members at the same time, causal evidence on the short and long term effects of family structure is scarce, especially in the field of economics.¹

¹ Exceptions are, e.g., Altonji et al., 2017; Bingley et al., 2017; Black et al., 2017; Joensen and Nielsen, 2018; Breining et al., 2017.

The first part of this thesis analyzes two aspects of family structure that can be considered exogenous – the age difference between siblings and the gender of the sibling – on an outcome that has increasingly become more important in the field of economics, namely personality of children. Personality traits are crucial for success in life (Borghans et al., 2016; Roberts et al., 2007) and the first part of this thesis studies both short and long term effects of the age difference between siblings and sibling gender on personality traits of children.

Besides the structure of a child's family, educational outcomes of children may be affected by the educational setting and policies of the country one attends school in. The second part of this thesis focuses on effects of the Dutch educational setting on primary education outcomes of children with a migration background (refugees and other migrant children). The Netherlands has a relatively early tracking system.² Early tracking systems have been documented to be harmful for the long run educational and human capital outcomes of students, especially of those with a migration background (see e.g. Checchi & Flabbi, 2006; Hanushek & Woessmann, 2006; Brunello & Checchi, 2007; Bol & van de Werfhorst, 2013; Ludemann & Schwerdt, 2013; Hopwood et al., 2016). The Dutch secondary education track placement depends on a standardized school exit test and an assessment from the teacher. Cognitive test scores are assumed to give objective measures of the child's ability. However, especially students with a migrant background face more challenges than native children that could negatively affect their scores. They have to cope with traumatic experiences, adjust to a new culture, learn a different language and get used to a new educational system. Teacher assessments could convey different information about the child's ability, as the teacher observes other aspects of ability, such as learning behavior and motivation (Segal, 2008). However, this subjective ability measure can be prone to teacher bias, e.g. towards gender and migration background (Dee, 2005; Burgess and Greaves, 2013; Fairlie et al., 2014).

Since education is one of the main channels for integration, characteristics of the educational system can have long term consequences for children with a migration background. The first question I study in the second part of the thesis is how teachers' inability to observe the school exit test score information affects their assessments of immigrants, by using a Dutch policy change.³ With this policy change, the assessment procedure for track assignment changed from a mainly

2 In the Netherlands, students are tracked when they are 12 years old. Only Austria, Germany, Hungary, and the Slovak Republic track at an earlier age (10 years).

3 This policy change dictated that from 2014/2015 onwards, the teacher assessment is leading in terms of secondary school track placement. After the policy change, teachers give an assessment before the students have taken the school exit test, compared to a situation where students first took a test and teachers could base their assessment on the test scores.

test based assignment to an assignment by teachers. I study the effect of this policy change on the teacher assessment for migrant children relative to native children. The second question I study is how specialized refugee schools (called newcomer facilities⁴) affect test scores and teacher assessments of refugee children. I use the random placement of refugee children in specialized or regular schools to test which type of school is better for the educational outcomes of refugee children.

1.2 Aim

This thesis aims to investigate the effect of family structure on children's personality traits and the effect of educational policy on educational outcomes of migrant children. The main research questions and sub-questions of this thesis are as follows:

Part I:

1. How does family structure affect children's personality traits?

Chapter 2. What is the effect of the age gap between siblings on personality traits?

Chapter 3. What is the effect of sibling gender on personality traits?

Part II:

2. How does the educational setting in the Netherlands affect educational outcomes of migrant and refugee children?

Chapter 4. What is the effect of the inability to observe the primary school exit test scores on the teacher's assessment of immigrant children?

Chapter 5. What is the effect of specialized refugee schools on refugees' test scores and teacher assessments?

The academic contribution of this this thesis is that I study effects of characteristics of the family and educational setting on children's outcomes using unique datasets and employing methods that plausibly test for causal effects.

The four different studies are also relevant for policy makers. This holds especially for the two studies on effects of educational settings on migrant and refugee children. These children have special consideration by policy makers as they face more challenges than native children. They may have to adjust to a new

4 These newcomer facilities are primary schools established for newly arrived immigrants. Immigrant children attend these facilities for a maximum of two years, during which the focus is on learning the Dutch language. After this period, they enroll at regular primary schools.

culture, a different language and refugee children may even have had traumatic experiences. Was the introduction of a new policy that shifts tracking decisions from a test-based assessment to a teacher-based assessment beneficial for the secondary education tracking of migrants, as was expected by policy makers? And is it a good idea to organize specialized schools for refugee children or is it better to place these children in regular schools? These are important questions for policy makers in the education field. The evidence on the causal effect of certain family structure aspects on the outcomes of children also provides new insights for policy makers. These studies focus on the effect of family structure on the personality of children. Policy makers can keep these insights in mind when introducing policies that possibly affect family structure, such as parental leave.

1.3 Outline

Part I

Chapter 2

The second chapter studies the causal effect of birth spacing (i.e., the age difference between siblings) on personality traits. The birth gap can be expected to affect personality traits for several reasons. When the birth gap is small, siblings compete for similar parental resources around the same time (e.g., help with homework). On the other hand, a small birth gap implies that when parents help one child, the other may be more able to learn from this as well, and that at young ages, the siblings may be more able to play with each other and learn from each other. The birth gap may affect personality traits such as competitiveness, neuroticism, extraversion, social skills, self-esteem, locus of control, etc. It may also affect behavioral problems in school and disorganized behavior. Because of the contrasting mechanisms, the signs of the effects of the birth gap on personality traits are difficult to predict.

I use the British Cohort Study to study the effects of family structure on personality traits of children. The British Cohort Study reports longitudinal data on personality traits from a large British cohort.⁵ The identification strategy follows earlier studies and employs miscarriages between the first and second child as

5 The British Cohort Study (BCS70) follows a cohort of approximately 17,000 babies who were born in the same week in April 1970 in England, Scotland and Wales since birth. These individuals have been followed over a time span of 50 years in 12 surveys, namely at the ages of 2, 5, 10, 16, 21, 26, 29, 34, 38, 42, 46 and 50 (forthcoming). The data contain information on personality traits, and a wide variety of life outcomes, such as achievement tests, educational outcomes, health, and criminality.

an instrument for birth spacing. The results show that a larger age gap between siblings negatively affects disorganized and neurotic behavior of the youngest child in two-child households. This result sheds a first light on the causal effects of birth spacing on personality traits.

Chapter 3

The third chapter studies whether sibling gender affects personality traits. In previous literature, boys and girls have been shown to display stereotypical behavior and the effect of sibling gender can have long term consequences regarding education, wages and family formation. The effect of sibling gender on personality traits has only recently been explored and only short term outcomes have been studied. The British cohort study provides data to explore the effect of sibling gender on personality traits at several ages and covers a wide variety of outcomes measured across the life span. The results show that for both boys and girls having a sister has positive consequences for personality traits. Girls with a sister are less disorganized and anti-social than girls with a brother, while boys with a sister have fewer behavioral problems than boys with a brother. Having a sister is also reflected in more positive life outcomes.

Part II

Chapter 4

The fourth chapter investigates the effect of the teacher's inability to observe the school exit test score on the teacher assessment of the level of secondary education for migrant children in the Netherlands by exploiting a policy change in primary education. This policy change dictates that while in earlier years teachers had to give assessments of the level of secondary school (partly) based on an exit test, from 2014/2015 onwards, the teacher had to give an assessment of the level of secondary school before knowing the exit test score of students. The idea behind this policy change is to put more weight on the assessment of the teacher in the secondary school track placement decision than on the exit test scores. A student can have a bad test day, but the teacher has observed the students for a longer period of time and knows their learning attitude and previous study results. Since the exit test score presumes a certain level of knowledge about the Dutch language, one would expect that the exit test scores of migrant children are lower than their capacities (most of these children do not speak Dutch at home). If this is true, the policy change would have a positive effect on the teacher assessments of students with a migrant background. After all, their exit test scores (which are lower than the natives' scores) are no longer observed by the teacher when assessing the child after the policy change. However, the results indicate that the policy change had the opposite of the intended effect: first generation non-Western immigrants

receive significantly lower teacher assessments compared to natives and other immigrants after the policy change.

Chapter 5

The fifth chapter investigates the effect of specialized refugee schools on the primary school outcomes of refugees in the Netherlands by exploiting the Dutch immigration allocation policy as a natural experiment. Refugees are randomly assigned to housing facilities and their children are then enrolled at the nearest school; a regular primary school, or a specialized refugee school (i.e. a school especially established for newly arrived immigrants and refugees). Specialized refugee schools are solely meant for newly arrived immigrants and provide a different educational environment than regular schools. At these specialized schools, the main focus is on learning the Dutch language, with personalized learning goals and teachers who have experience with traumatized children and teaching Dutch for non-natives. At regular schools, immigrants follow the standard curriculum with some additional Dutch classes. The effects on students' outcomes of these two different approaches and educational settings have not been studied. Some argue that specialized refugee schools are better for refugee children, because of their tailored approach and experience with language and traumas. Others argue that refugee schools are a parallel schooling system, which might have a negative impact on immigrant children's outcomes due to negative peer effects and lack of contact with native students (e.g. Baubock & Tripkovic, 2017) and thus less exposure to the Dutch language. The results show that there are no significant effects of being placed at a specialized refugee school or a regular school on exit test scores and teacher assessments of refugee children in the Netherlands.

Chapter 2

Does birth spacing affect personality?[†] [‡]

[†] Published as: Golsteyn, B.H.H. & Magnée, C.A.J. (2017). Does birth spacing affect personality? *Journal of Economic Psychology*, 60, pp. 92-108.

[‡] We thank Lex Borghans, the editor Martin Kocher, an anonymous referee, participants at the Fourth Lisbon Research Workshop on Economics, Statistics and Econometrics of Education, and seminar participants at Maastricht University for their valuable comments.

2.1 Introduction

Family structures have important effects on individuals' lives. It has been documented in economics, psychology and medicine that the number of children and the order of the child within the family affect vital outcomes in life such as educational attainment, intelligence and health (see, e.g., Rohrer et al., 2015; Bjerkedal et al., 2007; Hotz & Pantano, 2015; Black et al., 2007; and Barclay, 2015).

This chapter investigates whether the age difference between siblings affects personality traits. Personality traits are crucial for success in life (see, e.g., Borghans et al., 2016). The birth gap can be expected to affect personality traits for several reasons. When the birth gap is small, siblings compete for similar parental resources around the same time (e.g., help with learning how to ride a bike, payment of school fees, help with homework). On the other hand, a small birth gap implies that when parents help one child, the other may be more able to learn from this as well, and that at young ages, the siblings may be more able to play with each other and learn from each other. The birth gap may affect personality traits such as competitiveness, neuroticism, extraversion, social skills, self-esteem, locus of control, etc. It may also affect behavioral problems in school and disorganized behavior. Because of the contrasting mechanisms, the signs of the effects of the birth gap on personality traits are difficult to predict.

We use the 1970 British Cohort Study, a longitudinal data base consisting of approximately 17,000 children born in the UK in one week of April 1970. These children have been followed in 10 surveys from birth (parental survey) up to an age of 42. The data contain the following personality traits of the children at age 10: Rutter behavioral problems, self-esteem, locus of control, disorganized behavior, anti-social behavior, neuroticism, and introversion. The data also contain information about the number of siblings and their year of birth, and about the pregnancy and health of the child's mother. Because we only have information about miscarriages (i.e., our instrumental variable: see below) before the birth of the child that is followed in the survey, we can only estimate the effects of birth spacing on the personality of the youngest child. We furthermore focus on two-child households to abstract from analytical complexities of several age differences between siblings within the family (e.g., in three child families: youngest versus the oldest child, youngest versus the middle child), family size and birth order effects. In addition, we study the effects on male and female children separately, since parental gender preferences and differences in sibling interactions with respect to gender can affect the outcomes of boys and girls differently (e.g. Conley, 2000).

An analytical challenge is that the age gap between children within a household is endogenous. Various confounding factors may relate both to birth spacing and to the personality of the youngest child. For instance, the personality of parents

may be related to the choice to wait longer to have a second child, and to the personality of the second child. This implies that in order to study the causal relationship between birth spacing and personality, we need exogenous variation in birth spacing. In line with Buckles and Munnich (2012) who study the effect of birth spacing on achievement test scores, we use miscarriages between the first and second child as an instrument for birth spacing. The assumptions underlying this method are that (1) women who miscarry between their first and second child on average have a larger age difference between the children, and (2) that miscarriage occurs at random.

We carefully inspect the assumptions underlying our method. With respect to the first assumption, we show that miscarriages indeed highly correlate with birth spacing. The F-statistics in the first stage regressions also show that miscarriage is a strong instrument for birth spacing. Concerning the second assumption, we show that miscarriages do not correlate with several observables: smoking behavior of the mother during pregnancy, age of the mother at the birth of her first child, feelings of depression, and social class. However, we do find that mothers who miscarried between their first and second child are under closer surveillance by the hospital in their pregnancy of the youngest child. They more often receive antenatal care. We control for the use of antenatal care in our regressions. More importantly, we also find that women who miscarry between their first and second child were more likely to also have miscarried before their first child. This indicates that there may be a genetic or behavioral component to miscarriage. We therefore control for the number of miscarriages that occurred before the first child in our regressions. Conditional on these controls, miscarriages arguably occur at random, implying that we can use them as exogenous shocks. We also show that the results remain similar when we do and do not control for these variables.

An important additional challenge is that our instrument may be related to children's outcomes via different variables than the birth gap. The most obvious candidates are maternal mental and physical wellbeing. However, the literature overview given by Buckles and Munnich (2012) reveals that it is unlikely that our instrument is related to maternal mental or physical wellbeing. More details are given below.

The main result of our analysis is that a larger age gap between siblings negatively affects personality traits. Specifically, a larger birth gap leads to more disorganized behavior, more neuroticism, and more introversion. For small gap ranges (gaps of less than 4 years or a gap of 2 or 3 years)⁶, we find that a larger gap leads to less self-esteem, more introversion and more anti-social behavior.

⁶ In the robustness checks, we show estimates for the full sample, a restriction on a gap smaller than 7 years (the 90th percentile), a gap smaller than 5 years, smaller than 4 years, and a gap of 2 or 3 years.

Separating the results for boys and girls, girls become more neurotic due to a larger birth gap, while for smaller gap ranges, they become more anti-social and more introverted. Boys become more neurotic for large gap ranges and more disorganized for small gap ranges.

Our study contributes to a large literature on the effects of family structure on important life outcomes. One part of this literature focuses on the effects of birth order on personality and intelligence. Rohrer et al. (2015) and Bleske-Rechek and Kelley (2014) find no effect of birth order on personality. However, Rohrer et al. (2015) do find that intelligence decreases with a higher birth order. This last result has been confirmed by Bjerkedal et al. (2007), Hotz and Pantano (2015), Black et al. (2007) and Barclay (2015), and challenged by Kanazawa (2012). Salmon et al. (2016) find that birth order has a moderate (positive) effect on pro-social behavior. The findings on the effects of birth order on personality thus remain inconclusive.

In a recent meta-study in pediatrics, the effects of birth spacing on one facet of personality were taken into consideration. Conde-Agudelo et al. (2016) analyze the non-causal relationships of the birth gap and autism. Their conclusion is that short birth intervals are associated with an increased risk of autism spectrum disorder. Our contribution to this literature is that we study the effects on several facets of personality, and that we do take the endogeneity of the birth gap into consideration.

There have been few papers which have studied causal effects of birth spacing on other outcomes and to our knowledge the causal relationship between birth spacing and personality has not been studied before. Buckles and Munnich (2012) use miscarriages as an instrument for birth spacing, and find no effect of the birth gap on the achievement test scores for the youngest child in a sample of US children. Our study shows negative effects of a larger birth gap on personality. Because personality is positively related to educational outcomes, this appears to be inconsistent with Buckles and Munnich's research. However, a negative effect on personality does not necessarily imply that educational outcomes are also negatively affected by the birth gap. The reason for this is that other factors related to education may also be affected by the birth gap. In an extension of the analysis, we investigate the causal relationship between the birth gap and a large vector of achievement tests and educational outcomes. We find no robust significant effects of the birth gap on achievement test scores which is in line with the findings in Buckles and Munnich's article. In some specifications, there is a significant negative effect on educational outcomes but for most educational attainment variables there are no significant effects. The results are therefore inconclusive on the relationship between birth gaps and educational outcomes.

The chapter proceeds as follows. Section 2.2 describes the data. Section 2.3 discusses the empirical strategy. Section 2.4 reports the results. Section 2.5 gives the conclusions.

2.2 Data

We use the British Cohort Study (BCS70) which follows a cohort of approximately 17,000 babies who were born in the same week in April, 1970 in England, Scotland and Wales since birth. These individuals have been followed over a time span of 42 years in 10 surveys, namely at the ages of 2, 5, 10, 16, 21, 26, 29, 34, 38, and 42.

Important for this chapter is that the dataset contains information about whether the mother had a miscarriage after the birth of the first child and before conceiving the second child, and that at the age of 10 several questions about personality traits were asked to the child itself and to its mother and its teacher.

There is considerable panel attrition over the years (see Table A2.1). At age 10, still around 87% of the sample remains. The attrition is not related to the main variables of our analyses (personality traits, birth spacing, miscarriages), so we conclude that it is unlikely that it will bias our results.

Sample

We restrict our sample to families with two children, whose first surviving child was born before 1970 and the second surviving child in April 1970. We only have data on personality for the child born in 1970 and on miscarriages before 1970 so our analysis focuses on the effects for the youngest child in the household. We exclude families with siblings born after the second child. We also exclude all twins (120 subjects) from the sample, since in this case there are no miscarriages possible between the first and second child. As a result, the sample reduces from 17,196 to 4,114 children. We have information on personality outcomes for approximately 2,600-2,850 out of these 4,114 children. Table 2.1 gives an overview of the descriptive statistics of all the variables.

Birth spacing

We define birth spacing as the difference in years between the two siblings in the household. The mean birth gap is 3.5 years (see Table 2.1), which is remarkably similar to the reported gap in the sample of Buckles and Munnich (2012) (3.4 years) and to the 1988 Natality Detail files (3.4 years).

Table 2.1 Descriptive statistics

	Obs.	Mean	St. Dev.
Rutter behavioral difficulties	2859	0	1
Self-esteem	2630	0	1
Locus of control	2589	0	1
Disorganized	2553	0	1
Asocial	2722	0	1
Neuroticism	2740	0	1
Introversion	2740	0	1
Birth gap (in years)	4113	3.508	2.393
Number of miscarriages between first and second child (broad)	4114	0.120	0.420
Number of miscarriages before first child (broad)	4114	0.130	0.422
Number of miscarriages between first and second child (narrow)	4114	0.095	0.378
Number of miscarriages before first child (narrow)	4114	0.088	0.351
Gender (1=male)	4114	0.523	0.500
Mother's age at delivery of first child (in years)	4097	22.987	3.848
Married at time of birth of second child (1=yes)	4114	0.972	0.166
Social class father	4114	3.829	1.498
Social class mother	4114	5.128	2.457
Smoking behavior mother during pregnancy second child (1=smokes)	4114	0.428	0.495
Antenatal care during pregnancy of second child (1=yes)	4114	0.163	0.369
Mother's wellbeing at age 5 of second child	2681	0	1
Second child lives with same parents since birth (1=yes)	3017	0.889	0.314
Mother's attitude to maternal employment	2913	0	1
Mother's attitude to sex equality	2913	0	1
Mother's attitude to better life for women	2913	0	1
Mother's attitude to tv viewing	2913	0	1
Mother's attitude to hospital visiting	2913	0	1
Mother's authoritarian world view	2913	0	1
Mother's authoritarian child rearing	2913	0	1
Mother's attitude to child independence	2913	0	1

Note: The sample is restricted to two-child families. All personality traits are measured at age 10 of the second (i.e. youngest) child and standardized to mean zero and standard deviation one. Questions related to the Rutter test are answered by the mother. Self-esteem and locus of control questions are answered by the child. Questions related to disorganized, asocial, neurotic and introverted traits are answered by the teacher. A high score on the Rutter test implies more behavioral difficulties. A high score on the locus of control scale implies an internal locus of control. Miscarriages are broadly defined to include the following categories: 'died under 7 days,' 'died 7 days and over,' 'stillbirth,' 'miscarriage,' 'ectopic,' and 'hydatidiform mole.' We define a miscarriage with a narrow definition if the respondents indicated the 'miscarriage' category. In a robustness check we include all other categories (excluding 'alive in April 1970'). Social class mother/father contains 6 categories. See the appendix for more information. Mother's wellbeing at age 5 of the second child is standardized to mean zero and standard deviation one. All variables concerning mother's attitudes are standardized to mean zero and standard deviation one as well.

An important caveat in our study is that we only have information on birth spacing measured in years which is crude relative to the measure used for instance by Buckles and Munnich (2012) who report the spacing in months. On the positive side, our data contain more observations of miscarriages than theirs: 424 relative to 291 in their sample.

We study the linear relationship between birth spacing and personality. However, it may be that there are important non-linearities in the relationship, e.g. that there is an optimal amount of birth spacing. We show separate regressions with varying restrictions on the range of the birth gap. Figure 2.1 shows the cumulative distribution of the birth gap. We restrict the sample to a gap of 15 years (the 99th percentile) in our baseline estimate in order to exclude outliers. In the robustness checks, we show estimates for the full sample, a restriction on a gap smaller than 7 years (the 90th percentile), a gap smaller than 5 years, smaller than 4 years, and a gap of 2 or 3 years.

Figure 2.1

Cumulative distribution of the birth gap

Note: the birth gap is measured in years which explains the stepwise progression.

Miscarriage

Information about the survival of children born before 1970 was provided in different categories, namely ‘alive in April 1970,’ ‘died under 7 days,’ ‘died 7 days

and over,' 'stillbirth,' 'miscarriage,' 'ectopic,' and 'hydatidiform mole.' We use two definitions of miscarriages. In our broad definition (the baseline analyses), we use the number of occurrences in all categories of this variable as an instrument (excluding 'alive in April 1970'). In our narrow definition, we only use the number of miscarriages between the oldest and youngest child as an instrument. The results are comparable under both definitions. In our sample, 7.8% of the women reported that they experienced a miscarriage between the first and second born child. This likelihood of miscarriage is similar to that found in other studies. For instance, Buckles and Munnich (2012) report that 5.8% of the women miscarried. Tables A2.2 and A2.3 shows the frequency table of miscarrying in the broad and narrow definition respectively.

Personality traits

The following personality traits are measured in the data: Rutter behavioral problems, self-esteem, locus of control, disorganized, anti-social behavior, neuroticism and introversion.

Rutter scores are answered by the mother. A full list of questions underlying the Rutter score is displayed in Table A2.4. A higher Rutter score indicates more negative overall behavior of the child. We use the principle component of the items standardized to a mean of zero and standard deviation of one in order to be able to compare the coefficients across the variables.

Children were asked questions about their self-esteem and about their locus of control. An overview of the questions is provided in Tables A2.5 (self-esteem) and A2.6 (locus of control). A higher score on the self-esteem scale implies a higher self-esteem of the child. A higher score on the locus of control scale implies a more internal locus of control.

Questions about disorganized, anti-social, neurotic and introverted behavior are answered by the teacher of the child. The questions about personality traits of the child were part of a bigger survey on the child's behavior. Table A2.7 shows the questions per personality trait. A higher score on the variables indicates respectively more disorganized, asocial, neurotic and introverted behavior of the child. Note that the items used to measure disorganized behavior are similar to items used nowadays to measure conscientiousness.

Control variables

We control for several important variables in our regressions. By comparing the estimates before and after controlling, we can get a sense of the extent to which our instrument indeed provides exogenous variation. Next to this, controlling for

important characteristics which are related to the personality of children can help to increase the efficiency of the estimates.

Most of the controls are characteristics of the mother: e.g., the mother's age at the delivery of the first born child,⁷ her smoking behavior during pregnancy, whether she is married or not, and her partner's socio-economic class. The age of the mother and her smoking behavior are important controls as these might be related to our instrument. We exclude mothers who were younger than 16 when they had their first child (seventeen in total). On average, women in this sample had their first child at age 23 (see Table 2.1). Figure A2.1 shows that there is a large variation in the distribution of the mothers' age when they had her first child. Around 43% of the women smoked during pregnancy. For current standards, this is a very high percentage. Table A2.8 gives more information about the number of cigarettes smoked by these women. Around 97% of the women were married in 1970 (see Table 2.1). Tables A2.9 and A2.10 show statistics on the social class of the father and the mother. Marriage and social class are important controls as they may influence the upbringing of the child.

In our most elaborate estimations, we also control for various other factors. We control for these variables in a separate estimation as the number of observations related to these variables is lower than for the other controls (see Table 2.1). We control for the stability of the marriage: Table 2.1 shows that 89% of the children lived with the same parents since birth. We also control for the mother's psychological health at age 5 of the child, for a number of variables related to the mother's attitude toward child rearing and toward other views about life (see Tables A2.11 and A2.12 for the lists of questions). These attitudes may be related to personality traits. Controlling for these variables may therefore reduce standard errors of our regressions. We control also for antenatal care (Table 2.1 shows that around 16% of the women used antenatal care), and the amount of miscarriages before the oldest child because these variables are related to our instrument as explained below (see Table A2.13 for the frequency of miscarrying before the oldest child).

2.3 Empirical Strategy

Birth spacing is endogenous which implies that we cannot rely on OLS regressions of personality traits on the birth gap.⁸ In order to study the causal relationship between birth spacing and personality, we need exogenous variation in birth spacing. Following Buckles and Munnich (2012), we employ an instrumental

⁷ It is important to control for the age at delivery of the first born and not of the youngest child since the age of the youngest child is arguably related to the instrument.

⁸ The appendix table A2.14 reports the OLS estimations.

variables approach exploiting miscarriages between the first and second child as an exogenous source of variation in the birth gap. The conditions underlying this method are that (1) women who miscarry between their first and second child on average have a larger age difference between the children, and (2) that miscarrying occurs at random. Below we discuss these conditions.

First stage estimations

In order to investigate whether miscarriages can be used as an instrument for the birth gap, we first show histograms of the birth gaps for mothers who did and who did not miscarry. Figure 2.2 clearly shows that the distribution of the birth gap of mothers who miscarried lies more to the right than for the non-miscarrying mothers.

Figure 2.2
Histogram birth gap for miscarriage and no miscarriage

Note: the figure shows two Kernel plots of the birth gap – one for the children for whom the mothers did not have a miscarriage between the siblings, and one for the children for whom the mothers did have a miscarriage in between.

First stage regressions in Table 2.2 confirm that on average, miscarriages lead to a larger birth gap. The table reports regressions of the relationship for three variants of the analysis, including and excluding controls. It is evident that miscarriage is a strong predictor of the birth gap ($F\text{-test} > 10$) in all variants of the regressions.

The regressions indicate that one more miscarriage increases the birth gap with around 1.3 years.

Table 2.2 First stage regressions of birth gap on miscarriage

	OLS gap without controls	OLS gap with a selection of controls	OLS gap with all controls
Number of miscarriages between first and second child (broad)	1.294*** (0.075)	1.253*** (0.075)	1.294*** (0.109)
F-statistic (miscarriage)	296.434***	280.341***	141.957***
Observations	4,079	4,063	2,337
R-squared	0.068	0.097	0.105

Note: Each column shows a regression with birth gap as the dependent variable and miscarriage as the independent variable. The first column shows the results when no controls are added to the regressions. The second column adds controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child. The third columns additionally add controls for mother's wellbeing at age 5 of the second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables and appendix Table A2.15 for the full regression results including all controls. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Interestingly, the table also shows that the birth gap appears to be endogenous as it is related to various controls, e.g. mother's age at delivery of the first child, smoking, social class, etc. This supports our choice to show IV regressions instead of focusing on correlations.

Exogeneity assumption

The second condition for our instrument to be valid is that miscarriages occur at random. Buckles and Munnich (2012) mention that chromosomal abnormality in the fetus is the most common reason for miscarriages. This abnormality is usually random and not associated with a higher risk of miscarrying in a next pregnancy. However, there may be other unobserved reasons for miscarrying. In order to get a sense of the validity of the assumption that miscarrying occurs at random, we show in Table 2.3 that miscarriages do not correlate with the observables in our data: e.g. smoking behavior of the mother during pregnancy, age of the mother, feelings of depression, and social class. The table shows that mothers who miscarried between their first and second child more often make use of antenatal care. This is logical because after a miscarriage, both women and their physicians may want to monitor the pregnancy more closely. We will control for the use of antenatal care (and for the other observables) in our regressions.

Table 2.3 OLS regression of the controls on miscarrying

	Miscarriage between first and second child
Gender (1=male)	-0.010
	(0.013)
Mother's age at delivery of first child (in years)	-0.002
	(0.002)
Married at time of birth of second child (1=yes)	-0.016
	(0.044)
Social class father II	-0.002
	(0.032)
Social class father III NM	-0.035
	(0.032)
Social class father III M	-0.000
	(0.029)
Social class father IV	-0.009
	(0.033)
Social class father V	0.022
	(0.041)
Social class father unsupported	-0.044
	(0.043)
Social class mother III NM	-0.023
	(0.029)
Social class mother III M	-0.024
	(0.042)
Social class mother IV	0.004
	(0.032)
Social class mother V	-0.112
	(0.079)
Housewives	-0.019
	(0.027)
Social class mother unsupported	-0.005
	(0.031)
Smoking behavior mother during pregnancy second child (1=smokes)	0.012
	(0.013)
Antenatal care during pregnancy of second child (1=yes)	0.085***
	(0.018)
Nr. Miscarriages before the first child (broad)	0.070***
	(0.016)
Constant	0.181**
	(0.071)
Observations	4,063
R-squared	0.014

Note: Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Although the main determinant of miscarrying (chromosomal abnormality in the fetus) may occur at random and not give a higher risk in the next pregnancy, the table reveals that women who miscarry between their first and second child were more likely to also have miscarried before their first child. This implies that there may be a genetic or behavioral component to miscarrying. We therefore control for miscarriages that happened before the first child in our regressions. Conditional on these controls, miscarriages arguably occur at random, implying that we can use them as exogenous shocks.

A related issue is that miscarriages may affect the outcomes we study through other channels than the birth gap. The most obvious channel may be that mothers suffer mentally or physically after a miscarriage. The literature discussed by Buckles and Munnich (2012) has concluded however that symptoms of depression or anxiety typically disappear within one year after a miscarriage. Moreover, women who have a healthy pregnancy after the miscarriage (as is the case in our analysis) appear to be less likely to suffer from depression. Miscarriages furthermore do not appear to affect attachment to the next born child, and mothers who miscarry after the first born child are at lower risk of delivering prematurely than those who had not previously given birth. We conclude from these earlier studies that it is unlikely that reduced maternal mental or physical well-being can explain our findings.

Buckles and Munnich (2012) present other interesting points with respect to the identification strategy. One issue is that a miscarriage is both related to a change in birth spacing and to a change in parental age. Therefore, the effect of spacing cannot be identified independent of parental age. However, from a policy perspective, the combined effect of spacing and parental age is of interest since any policy which increases spacing will also increase parental age. A second point is that miscarriages may be underreported. Assuming this underreporting is random, our estimates are attenuated, and hence present lower bounds.

2.4 Results

Table 2.4 shows the main result of our analysis: a larger birth gap has a significant negative effect on personality traits: disorganized behavior and neuroticism. Importantly, the relationships are robust to the inclusion of the controls described above. The results show that if the age gap between siblings increases with one year, disorganized behavior increases with approximately 0.11 standard deviations and neuroticism increases with 0.15 standard deviations.

Table 2.4 IV regressions of birth gap on personality

	IV without controls		IV with a selection of controls		IV with all controls	
	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Rutter behavioral difficulties	0.044	2841	0.022	2829	0.023	2220
	(0.037)		(0.037)		(0.044)	
Self-esteem	-0.036	2631	-0.031	2605	-0.059	1889
	(0.038)		(0.039)		(0.054)	
Locus of control	-0.025	2547	-0.017	2566	0.006	1860
	(0.038)		(0.039)		(0.050)	
Disorganized	0.109**	2537	0.101**	2528	0.107*	1827
	(0.044)		(0.045)		(0.059)	
Asocial	0.033	2704	0.023	2694	0.014	1948
	(0.037)		(0.038)		(0.050)	
Neuroticism	0.089**	2723	0.082**	2714	0.145***	1955
	(0.037)		(0.038)		(0.053)	
Introversion	0.043	2721	0.039	2711	0.068	1960
	(0.038)		(0.039)		(0.053)	

*Note: Each cell shows the coefficient of a separate IV regression with the variable indicated in the row as the dependent variable and birth gap as the independent variable. The instrument is the number of miscarriages between the siblings. This baseline estimation shows results when the gap range is restricted to 15 years. The first column shows the results when no controls are added to the regressions. The second column adds controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child. The third columns additionally add controls for mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Table 2.5 reveals that when we reduce the gap range, the coefficients increase in size. For very small gap ranges, the size of the effects on disorganized behavior and on neuroticism increase to around 0.46 and 0.57 standard deviations respectively. Note, however, that these effects are no longer significant, plausibly due to smaller sample sizes. At very small ranges, we find significant and very large effects on almost all other traits: self-esteem (0.80), asocial behavior (0.79), and introversion (1.08). The only trait which does not seem to be affected are Rutter behavioral problems. Table 2.6 shows that if we use the narrow definition of miscarriages, results remain similar although less significant at small ranges.

Table 2.5 IV regressions of birth gap on personality, varying gaps

	Full sample		Gap < 7		Gap < 5		Gap < 4		Gap = 2 or Gap = 3	
	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Rutter behavioral difficulties	0.019 (0.044)	2234	0.069 (0.090)	2063	-0.061 (0.142)	1836	0.119 (0.269)	1485	0.188 (0.354)	1355
Self-esteem	-0.055 (0.052)	1900	-0.053 (0.110)	1757	-0.248 (0.174)	1564	-0.607* (0.347)	1255	-0.799* (0.449)	1150
Locus of control	0.000 (0.048)	1870	0.001 (0.108)	1728	-0.097 (0.177)	1531	0.038 (0.321)	1237	0.027 (0.425)	1134
Disorganized	0.091* (0.055)	1836	0.273** (0.116)	1704	0.413** (0.183)	1517	0.334 (0.311)	1217	0.459 (0.419)	1114
Asocial	0.005 (0.048)	1960	0.040 (0.105)	1809	0.075 (0.163)	1608	0.563 (0.343)	1291	0.789* (0.454)	1179
Neuroticism	0.154*** (0.055)	1965	0.231** (0.111)	1817	0.352** (0.174)	1616	0.444 (0.329)	1301	0.573 (0.440)	1192
Introversion	0.067 (0.050)	1972	0.208* (0.108)	1823	0.412** (0.176)	1620	0.840** (0.363)	1303	1.083** (0.490)	1190

Note: Each cell shows the coefficient of a separate IV regression with the variable indicated in the row as the dependent variable and birth gap as the independent variable. The instrument is the number of miscarriages between the siblings with the broad definition. All columns add controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child, mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.6 IV regressions of birth gap on personality, narrow definition

	Gap<15 (Baseline)		Full sample		Gap<7		Gap<5		Gap<4		Gap=2 or Gap=3	
	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Rutter behavioral difficulties	0.033 (0.057)	2220	0.030 (0.057)	2234	0.089 (0.105)	2063	-0.053 (0.158)	1836	0.131 (0.303)	1485	0.208 (0.392)	1355
Self-esteem	-0.094 (0.071)	1889	-0.092 (0.068)	1900	-0.106 (0.137)	1757	-0.254 (0.197)	1564	-0.548 (0.379)	1255	-0.710 (0.480)	1150
Locus of control	-0.055 (0.070)	1860	-0.056 (0.066)	1870	-0.074 (0.141)	1728	-0.071 (0.201)	1531	-0.049 (0.352)	1237	-0.081 (0.459)	1134
Disorganized	0.175** (0.079)	1827	0.157** (0.074)	1836	0.384** (0.153)	1704	0.430** (0.203)	1517	0.459 (0.360)	1217	0.592 (0.473)	1114
Asocial	0.029 (0.066)	1948	0.021 (0.064)	1960	0.068 (0.133)	1809	0.035 (0.183)	1608	0.467 (0.376)	1291	0.614 (0.485)	1179
Neuroticism	0.192*** (0.070)	1955	0.207*** (0.072)	1965	0.260* (0.140)	1817	0.326* (0.192)	1616	0.363 (0.363)	1301	0.444 (0.473)	1192
Introversion	0.107 (0.071)	1960	0.104 (0.069)	1972	0.251* (0.139)	1823	0.355* (0.198)	1620	0.864** (0.410)	1303	1.087** (0.541)	1190

Note: Each cell shows the coefficient of a separate IV regression with the variable indicated in the row as the dependent variable and birth gap as the independent variable. The instrument is the number of miscarriages between the siblings with the narrow definition. All columns add controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child, mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables. Standard errors are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.7 IV regressions of birth gap on personality, boys

	Gap < 15 (Baseline)		Full sample		Gap < 7		Gap < 5		Gap < 4		Gap = 2 or Gap = 3	
	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.
Rutter behavioral difficulties	-0.012 (0.049)	1164	-0.016 (0.047)	1173	0.005 (0.105)	1087	-0.159 (0.174)	968	0.185 (0.465)	766	0.243 (0.595)	689
Self-esteem	-0.011 (0.055)	989	-0.009 (0.049)	995	0.060 (0.124)	924	-0.083 (0.185)	831	-0.597 (0.466)	655	-0.756 (0.654)	589
Locus of control	0.020 (0.053)	957	0.011 (0.048)	963	0.044 (0.136)	892	-0.054 (0.204)	801	0.288 (0.461)	634	0.411 (0.624)	572
Disorganized	0.073 (0.066)	951	0.053 (0.058)	956	0.263* (0.153)	889	0.411* (0.222)	799	0.380 (0.453)	626	0.487 (0.605)	562
Asocial	-0.034 (0.060)	1017	-0.042 (0.055)	1024	-0.074 (0.144)	948	-0.163 (0.207)	852	0.350 (0.533)	667	0.452 (0.698)	600
Neuroticism	0.116** (0.056)	1018	0.121** (0.055)	1024	0.118 (0.135)	951	0.232 (0.200)	857	0.106 (0.456)	675	0.131 (0.590)	610
Introversion	0.035 (0.059)	1019	0.035 (0.054)	1026	0.124 (0.135)	952	0.280 (0.201)	857	0.594 (0.496)	674	0.709 (0.652)	606

Note: Each cell shows the coefficient of a separate IV regression with the variable indicated in the row as the dependent variable and birth gap as the independent variable. The instrument is the number of miscarriages between the siblings with the broad definition. All columns add controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child, mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables. Standard errors are shown in parentheses. **** $p < 0.01$, *** $p < 0.05$, * $p < 0.1$.

Table 2.8 IV regressions of birth gap on personality, girls

	Gap < 15 (Baseline)		Full sample		Gap < 7		Gap < 5		Gap < 4		Gap = 2 or Gap = 3	
	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Rutter behavioral difficulties	0.099 (0.101)	1056	0.108 (0.113)	1061	0.147 (0.166)	976	0.021 (0.246)	868	0.078 (0.325)	719	0.092 (0.438)	666
Self-esteem	-0.175 (0.131)	900	-0.200 (0.155)	905	-0.190 (0.193)	833	-0.507 (0.342)	733	-0.689 (0.508)	600	-0.892 (0.640)	561
Locus of control	-0.022 (0.124)	903	-0.024 (0.140)	907	-0.032 (0.171)	836	-0.162 (0.330)	730	-0.134 (0.451)	603	-0.247 (0.607)	562
Disorganized	0.199 (0.138)	876	0.224 (0.159)	880	0.267 (0.167)	815	0.441 (0.310)	718	0.317 (0.404)	591	0.418 (0.544)	552
Asocial	0.125 (0.100)	931	0.144 (0.116)	936	0.168 (0.147)	861	0.455* (0.272)	756	0.794 (0.430)	624	1.098* (0.589)	579
Neuroticism	0.219* (0.132)	937	0.252 (0.154)	941	0.360* (0.185)	866	0.526 (0.324)	759	0.604* (0.480)	626	0.792 (0.653)	582
Introversion	0.149 (0.109)	941	0.162 (0.124)	946	0.317* (0.173)	871	0.653* (0.334)	763	1.014* (0.518)	629	1.361* (0.725)	584

Note: Each cell shows the coefficient of a separate IV regression with the variable indicated in the row as the dependent variable and birth gap as the independent variable. The instrument is the number of miscarriages between the siblings with the broad definition. All columns add controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child, mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tables 2.7 and 2.8 separate the results for boys and girls. Sample sizes reduce with around 50% so many significant results are no longer significant. Nevertheless, it appears that both boys and girls become more neurotic when faced with a larger birth gap. Boys additionally become more disorganized while girls become more anti-social and introverted.

Birth gap and educational outcomes

Our study shows negative effects of a larger birth gap on personality, which in turn suggests negative effects on outcomes related to personality such as education. In light of the earlier findings of zero effects of the birth gap on achievement test scores (Buckles and Munnich 2012), an important further question is how to interpret our results relative to their results. Are larger birth gaps good or bad for children's educational attainment?

Theoretically, it is important to note that our findings are not necessarily inconsistent with the findings of these authors because birth gaps may affect a large array of variables which in turn affect educational outcomes. So even if the birth gap negatively affects personality and personality is positively related to education, it may still be that the birth gap does not affect education.

We investigate the effects of the birth gap on a large vector of achievement tests and educational outcomes: see Table 2.9. We find no robust significant effects of a larger birth gap on any achievement test which is in line with Buckles and Munnich's result. The results show no effects or in certain specifications significant negative effects on other education outcomes. The results are therefore inconclusive on the relationship between birth gaps and educational outcomes.

Table 2.9 IV regressions of birth gap on achievement tests, education and other life outcomes

		IV regression of birth gap without controls		IV regression of birth gap with a selection of controls		IV regression of birth gap with all controls	
		Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Achievement tests	EPVT cognitive test (age 5)	0.004	2713	0.012	2704	-0.013	2199
		(0.035)		(0.035)		(0.041)	
	Human figure drawing cognitive test (age 5)	-0.006	2838	-0.007	2829	-0.035	2299
		(0.036)		(0.037)		(0.046)	
	Copy designs cognitive test (age 5)	-0.009	2874	0.004	2865	-0.036	2327
		(0.035)		(0.035)		(0.044)	
	Edinburgh Reading Test (age 10)	-0.016	955	-0.009	952	0.096	710
		(0.064)		(0.062)		(0.082)	
	Friendly Math Test (age 10)	-0.130**	891	-0.106*	887	-0.052	653
		(0.057)		(0.055)		(0.072)	
	Vocabulary Test (age 10)	-0.009	2807	-0.007	2797	0.004	2022
		(0.036)		(0.037)		(0.048)	
	Reading Test (age 10)	0.064*	2807	0.062*	2797	0.057	2021
		(0.036)		(0.037)		(0.050)	
	Spelling Test (age 10)	-0.039	2125	-0.027	2118	0.001	1525
		(0.048)		(0.049)		(0.069)	
	Sequence Test (age 10)	0.007	2521	0.015	2514	0.040	1824
		(0.038)		(0.037)		(0.050)	
	Sentence Test (age 10)	0.032	2641	0.019	2634	0.015	1900
		(0.036)		(0.037)		(0.050)	
Education	Math Score (age 16)	0.055	849	0.016	846	0.115	614
		(0.078)		(0.073)		(0.093)	
	Spelling Test (age 16)	0.096	1376	0.116*	1373	0.126	1004
		(0.065)		(0.062)		(0.079)	
	Vocabulary Test (age 16)	0.004	1376	0.022	1373	0.002	1004
		(0.064)		(0.063)		(0.081)	
	Grade Point Average age 16	0.010	1463	0.001	1459	0.026	1094
		(0.054)		(0.054)		(0.063)	
	Age left school (measured at age 26)	-0.069	2071	-0.066	2066	0.001	1452
		(0.046)		(0.045)		(0.060)	
	Highest qualification at age 26	-0.047	1995	-0.031	1990	-0.006	1401
		(0.047)		(0.044)		(0.060)	
	Age left education (graduation age corrected for highest qualification)	-0.078*	2204	-0.067*	2193	-0.020	1568
		(0.041)		(0.040)		(0.051)	

Other life outcomes	Body Mass Index	0.062	2142	0.060	2131	0.101*	1523
		(0.044)		(0.045)		(0.061)	
	Crime	0.032	2195	0.021	2184	-0.038	1562
		(0.041)		(0.042)		(0.055)	
	Life Satisfaction	-0.052	2195	-0.040	2184	0.034	1562
		(0.041)		(0.042)		(0.054)	
	Wages	-0.043	1159	-0.045	1156	-0.063	834
		(0.059)		(0.063)		(0.079)	

Note: Each cell shows the coefficient of a separate IV regression with the variable indicated in the row as the dependent variable and birth gap as the independent variable. The instrument is the number of miscarriages between the siblings with the broad definition. All columns add controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child, mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See Table 2.1 for definitions of these variables. For definitions of the dependent variables, see e.g. Borghans et al. 2016. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.5 Conclusions and policy implications

This chapter shows that larger age gaps between siblings negatively affect personality traits of the youngest child in a two-child household. Specifically, a larger birth gap leads to more disorganized behavior, more neuroticism, and more introversion. For small gap ranges (gaps of less than 4 years or a gap of 2 or 3 years), we find that a larger gap leads to less self-esteem, more introversion and more anti-social behavior. Separating the results for boys and girls, girls become more neurotic when the birth gap is larger, while for smaller gap ranges, they become more anti-social and more introverted. Boys become more neurotic for large gap ranges and more disorganized for small gap ranges.

Although personality traits appear to be negatively affected by the birth gap, it remains unclear at this point whether the birth gap affects future success of children in education and the labor market. More research is needed on this relationship. If the government targets to improve personality traits as such, it can be interesting for policy makers to consider interventions to shorten birth gaps. Policy makers can intervene in many ways, from providing information or services to economic incentives or regulations. One type of intervention can for instance be to provide information about the negative effects of a larger birth space on personality of the second child through online campaigns or leaflets at gynecologists' waiting rooms. Another implementation of such an intervention was done in Sweden, where women receive more parental leave benefits when the gap between children was less than two years (see Petterson-Lidbom & Skogman Thoursie 2009). In this way, the maternal leave period is condensed, which is

financially attractive for the government, as well as for the mother. Not only will she receive more money when she shortens the birth gap between two siblings, she may also be more able to re-enter the labor market sooner. This will result in lower depreciation of her skills and a higher likeliness to find a job that matches her skills, and thus receiving a higher income.

2.6 Appendix

Table A2.1 Panel attrition

Year of Sample	1970	1975	1980	1986	1996	2000	2004	2008	2012
Age of respondents	0	5	10	16	26	30	34	38	42
Obs.	17,196	13,135	14,875	11,615	9,003	11,261	9,665	8,874	9,841

Table A2.2 Miscarriages between first and second child (broad definition)

Nr. miscarriages	Frequencies
0	3,712
1	337
2	46
3	15
4	3
9	1

Table A2.3 Miscarriages between first and second child (narrow definition)

Nr. miscarriages	Frequencies
0	3,794
1	275
2	30
3	11
4	3
9	1

Table A2.4 Questions about Rutter score answered by the mother when the youngest child was 10 years old

• Very restless. Often running about or jumping up and down. Hardly ever still.
• Squirmy or fidgety
• Often destroys own or others' belongings
• Frequently fights with other children
• Not much liked by other children
• Often worried, worries about many things
• Tends to do things on his own - rather solitary
• Irritable. Is quick to "fly off the handle"
• Often appears miserable, unhappy, tearful or distressed
• Sometimes takes things belonging to others
• Has twitches, mannerisms or tics of the face or body
• Frequently sucks thumb or finger
• Frequently bites nails or fingers
• Is often disobedient
• Cannot settle to anything for more than a few moments
• Tends to be fearful or afraid of new things or new situations
• Is fussy or over particular
• Often tells lies
• Bullies other children

Table A2.5 Questions about self-esteem answered by youngest child at age 10

• Do you think that your parents usually like to hear about your ideas?
• Do you often feel lonely at school?
• Do other children often break friendships or fall out with you?
• Do you think that other children often say nasty things about you?
• When you have to say things in front of the teacher, do you usually feel shy?
• Do you often feel sad because you have nobody to play with at school?
• Are there lots of things about yourself you would like to change?
• When you have to say things in front of other children, do you usually feel foolish?
• When you want to tell a teacher something, do you usually feel foolish?
• Do you often have to find new friends because your old friends are playing with somebody else?
• Do you usually feel foolish when you talk to your parents?
• Do other people often think that you tell lies?

Table A2.6 Questions about locus of control answered by youngest child at age 10

• Do you feel that most of the time It's not worth trying hard because things never turn out right anyway?
• Do you feel that wishing can make good things happen?
• Are people good to you no matter how you act towards them?
• Do you usually feel that it's almost useless to try in school because most children are cleverer than you?
• Is a high mark just a matter of "luck" for you?
• Are tests just a lot of guesswork for you?
• Are you often blamed for things which just aren't your fault?
• Are you the kind of person who believes that planning ahead makes things turn out better?
• Do you find it easy to get up in the morning?
• When bad things happen to you, it's usually someone else's fault?
• When someone is very angry with you, is it impossible to make him your friend again?
• When nice things happen to you is it only good luck?
• Do you feel sad when it's time to leave school each day?
• When you get into an argument it is usually the other person's fault?
• Are you surprised when your teacher says you've done well?
• Do you usually get low marks, even when you study hard?
• Do you think studying for a test is a waste of time?

Table A2.7 Questions about disorganized, asocial, neurotic and introverted behavior answered by the teacher when the youngest child was 10 years old

Disorganized	Asocial	Neurotic	Introverted
• Daydreaming	• Complains about things	• Afraid of new things	• Excitable, impulsive
• Cannot concentrate	• Displays temper	• Behaves nervously	• Solitary child
• Bored in class	• Teases to excess	• Fussy	• Lethargic, listless
• Perseverance	• Interferes with others	• Worried, anxious	• Sullen/ sulky
• Confused with difficult tasks	• Changes mood quickly		
• Easily distracted	• Quarrels with others		
• Pays attention	• Destroys belongings		
• Forgetful with complex task	• Bullies		
• Lethargic, listless	• Sullen/Sulky		
• Completes tasks	• Easily frustrated		
• Falls to finish tasks			

Figure A2.1 Mother’s age at time of delivery of the first child

Table A2.8 Smoking behavior mother during pregnancy of youngest child

Non-smoker	44.9%
Stopped pre-pregnancy	12.2%
Stopped during pregnancy	3.8%
Smoked 1-4 cigarettes a day during pregnancy	6.9%
Smoked 5-14 cigarettes a day during pregnancy	19.5%
Smoked >15 cigarettes a day during pregnancy	12.7%
Total	100%

Table A2.9 Social class husband at time of birth of youngest child

Social class number	Description of social class	Frequencies
Social class I	Professional occupations, including doctors, lawyers, ministers of religion, university teachers, professional engineers, etc.	243
Social class II	Managerial and other processionals, including nurses, school teachers, company directors, etc.	543
Social class III NM	Non-manual skilled occupations, including ship assistants, company representatives, clerical workers, draughtsman, etc.	568
Social class III M	Skilled manual workers, including mechanics, craftsmen of all types, skilled engineers, etc.	1,862
Social class IV	Semi-skilled workers, including machine operators, postmen, storekeepers, porters, caretakers, etc.	511
Social V	Unskilled workers, including laborers, cleaners, dustmen, etc.	190
Social class unsupported	Unsupported social class	197

Table A2.10 Social class mother at time of birth of youngest child

Social class number	Description of social class	Frequencies
Social class I&II	Professional occupations, including doctors, lawyers, ministers of religion, university teachers, professional engineers, managerial and other processionals, including nurses, school teachers, company directors, etc.	297
Social class III NM	Non-manual skilled occupations, including ship assistants, company representatives, clerical workers, draughtsman, etc.	820
Social class III M	Skilled manual workers, including mechanics, craftsmen of all types, skilled engineers, etc.	158
Social class IV	Semi-skilled workers, including machine operators, postmen, storekeepers, porters, caretakers, etc.	543
Social class V	Unskilled workers, including laborers, cleaners, dustmen, etc.	34
Housewives	Housewife	1,754
Social class unsupported	Unsupported social class	508

Table A2.11 Questions about mother's wellbeing answered by her at age 5 of youngest child

• Do you often have back-ache?
• Do you feel tired most of the time?
• Do you often feel miserable or depressed?
• Do you often have bad headaches?
• Do you often get worried about things?
• Do you usually have great difficulty in falling asleep or staying asleep?
• Do you usually wake unnecessarily early in the morning?
• Do you wear yourself out worrying about your health?
• Do you often get into a violent rage?
• Do people often annoy and irritate you?
• Have you at times had a twitching of the face, head or shoulders?

• Do you often suddenly become scared for no good reason?
• Are you scared to be alone when there are no friends near you?
• Are you easily upset or irritated?
• Are you frightened of going out alone or of meeting people?
• Are you constantly keyed up and jittery?
• Do you suffer from indigestion?
• Do you often suffer from an upset stomach?
• Is your appetite poor?
• Does every little thing get on your nerves and wear you out?
• Does your heart often race like mad?
• Do you often have bad pains in your eyes?
• Are you troubled with rheumatism or fibrositis?
• Have you ever had a nervous breakdown?
• Do you have any other health problems worrying you?

Note: The mother could answer each question with either ‘yes’ or ‘no.’

Table A2.12 Examples of questions about mother’s attitude toward child rearing and toward other views about life answered by her at age 5 of youngest child

Attitude to maternal employment	Attitude to sex equality	Attitude to better life for women	Attitude to TV viewing
A wife must sacrifice her right to go out to work once she has children.	Girls are just as capable of boys to be engineers.	Women need something more from life than they can get by just looking after the home and children.	Young children who never see children’s TV miss a lot which is of value.
Attitude to hospital visiting	Authoritarian world view	Authoritarian child rearing	Attitude to child independence
It’s best not to visit children under five in hospital because it is too upsetting for the child.	People should be satisfied with their lot in this world and not struggle to get more.	A child should not be allowed to talk back to his parents.	A young child must be allowed to be himself even if this means going against his parents’ wishes.

Note: The mother could answer each statement with ‘strongly agree’, ‘mildly agree’, ‘cannot say’, ‘mildly disagree’ and ‘strongly disagree’. In total 43 statement were given. Statements were positively and negatively framed, resulting in a z-score calculated by the researchers of the British Cohort Study.

Table A2.13 Miscarriages before the first child (broad definition)

Nr. miscarriages	Frequencies
0	3,683
1	352
2	63
3	11
4	3
5	2

Table A2.14 OLS regression of gap on personality traits of child

	OLS without controls		OLS with a selection of controls		OLS with all controls	
	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Rutter behavioral difficulties	0.002 (0.009)	2841	-0.003 (0.009)	2829	-0.005 (0.011)	2220
Self-esteem	0.018* (0.010)	2613	0.022** (0.010)	2605	0.025** (0.012)	1889
Locus of control	0.021** (0.010)	2574	0.028*** (0.010)	2566	0.034*** (0.011)	1860
Disorganized	0.015 (0.010)	2537	0.005 (0.010)	2528	0.011 (0.012)	1827
Asocial	-0.013 (0.009)	2704	-0.018** (0.009)	2694	-0.020* (0.011)	1948
Neuroticism	0.024** (0.009)	2723	0.015 (0.010)	2714	0.027** (0.012)	1955
Introversion	0.008 (0.009)	2721	0.003 (0.009)	2711	0.021* (0.011)	1960

Note: Each cell shows the coefficient of a separate OLS regression with the variable indicated in the rows as the dependent variable and birth gap as the independent variable. All dependent variables are standardized to mean zero and standard deviation one. The birth gap is measured in years. The first column shows the results when no controls are added to the regressions. The second column adds controls for number of miscarriages between first and second child (broad), gender, mother's age at delivery of first child, married at time of birth of second child, social class father, social class mother, smoking behavior mother during pregnancy second child, antenatal care during pregnancy of second child. The third column additionally adds controls for mother's wellbeing at age 5 of second child, second child lives with same parents since birth, and mother's attitudes. See table 1 for definitions of these variables. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2.15 First stage regressions of birth gap on miscarriage with control variables displayed

	First stage regression of gap on miscarriages between the first and second child	First stage regression of gap on miscarriages between the first and second child with a selection of controls	First stage regression of gap on miscarriages between the first and second child with all controls
Nr. Miscarriages between first and second child (broad)	1.294***	1.253***	1.294***
	(0.075)	(0.075)	(0.109)
Nr. Miscarriages before the first child (broad)		-0.121	-0.245**
		(0.075)	(0.102)
Gender (1= male)		0.013	-0.065
		(0.062)	(0.078)
Mother's age at delivery of first child (in years)		0.026***	0.014
		(0.009)	(0.011)
Married at time of birth of second child (1=yes)		0.028	-0.035
		(0.211)	(0.421)
Social class father II		0.356**	0.327*
		(0.153)	(0.188)
Social class father III NM		0.220	0.099
		(0.153)	(0.186)
Social class father III M		0.330**	0.205
		(0.138)	(0.172)
Social class father IV		0.193	0.031
		(0.159)	(0.200)
Social class father V		0.027	-0.322
		(0.197)	(0.274)
Social class father unsupported		-0.014	-0.066
		(0.205)	(0.318)
Social class mother III NM		0.047	0.014
		(0.137)	(0.177)
Social class mother III M		0.320	0.040
		(0.198)	(0.249)
Social class mother IV		0.525***	0.320
		(0.150)	(0.197)
Social class mother V		0.218	-0.077
		(0.376)	(0.499)
Housewives		-0.196	-0.404**
		(0.127)	(0.165)
Social class mother unsupported		-0.317**	-0.405**
		(0.148)	(0.191)
Smoking behavior mother during pregnancy second child (1=smokes)		0.142**	0.113

	(0.064)	(0.081)	
Antenatal care during pregnancy of second child (1=yes)	0.509***	0.386***	
	(0.085)	(0.110)	
Mother's wellbeing at age 5 of second child		-0.009	
		(0.040)	
Second child lives with same parents since birth (1=yes)		0.179	
		(0.134)	
Mother's attitude to maternal employment		0.067	
		(0.044)	
Mother's attitude to sex equality		0.068	
		(0.048)	
Mother's attitude to better life for women		0.047	
		(0.043)	
Mother's attitude to tv viewing		-0.070*	
		(0.040)	
Mother's attitude to hospital visiting		-0.083*	
		(0.045)	
Mother's authoritarian world view		0.000	
		(0.053)	
Mother's authoritarian child rearing		-0.087*	
		(0.052)	
Mother's attitude to child independence		-0.062	
		(0.044)	
Constant	3.244***	2.283***	2.847***
	(0.032)	(0.339)	(0.535)
F-statistic of instrument	296.434***	280.341***	141.957***
Observations	4,079	4,063	2,337
R-squared	0.068	0.097	0.105

Note: Each column shows a regression with birth gap as the dependent variable and miscarriage as the independent variable. See table 1 for definitions of these variables. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Chapter 3

Does sibling gender affect personality traits? † ‡

† Golsteyn, B.H.H. & Magnée, C.A.J. (2020). Does sibling gender affect personality traits? *Economics of Education Review*, 70, 102016

‡ We thank Chris van Klaveren, Arjan Non, Tom Stolp, Inge de Wolf and seminar participants at Maastricht University and Free University of Amsterdam for their valuable comments.

3.1 Introduction

The environment in which children grow up has vital implications for later in life outcomes. Parents, teachers, and peers determine choices children make. Interactions between siblings are also important. Studies have shown that the number of siblings, their birth order, and the age difference between siblings affect outcomes such as personality traits, human capital accumulation, and wages.⁹

This chapter investigates whether sibling gender affects personality traits and other life outcomes. We use the 1970 British Cohort Study, a longitudinal database consisting of approximately 17,000 children born in the UK in one week of April 1970. These children have been followed in 12 surveys from birth (parental survey) up to the age of 50. The data contain personality traits of the children, and also later in life outcomes such as achievement test scores, health and other life outcomes. We focus on two-child households to abstract from analytical complexities of several siblings within the family (having more siblings implies that the child potentially has siblings with a different gender) and family size effects.

We investigate the effect of sibling gender for both oldest and youngest children in two-child families. Sibling gender is random,¹⁰ implying that simple correlations may reveal causal effects. One important issue, however, is that while parents cannot choose the sex of their child, they *can* choose the *number* of children they have. We find that parents with a low Socio-Economic Status (SES) are more likely to have a third child than parents with a high SES in case the first two children are of the same gender. We correct for potential selection bias with respect to SES by running our analyses for low and high SES categories separately. A second analytical issue is that we are analyzing many hypotheses and hence face the risk that our results are statistically significant by chance (i.e., a false positive or type I error). We therefore apply a Bonferroni correction for multiple hypotheses testing (i.e., the strictest possible correction for such errors).

Our main results are that the gender of the sibling has important implications for personality traits. Both boys and girls generally benefit from having a sister. The relationships between sibling gender and personality traits are stronger for

9 See, e.g., Barclay, 2015; Bjerkedal et al., 2007; Black et al., 2007; Buckles & Munnich, 2012; Golsteyn & Magnée, 2017; Hotz & Pantano, 2015; Kalmijn & van de Werfhorst, 2016; Rohrer et al., 2015.

10 Only from 1978 onwards, trustworthy ultrasound machines were available to determine the gender of a child during the pregnancy. Thus, parents deciding to terminate a pregnancy in 1970 could not have based the decision on the gender of the fetus. Furthermore, Wilcox et al. (1995) find no relation between time of intercourse with respect to ovulation and sex of the resulting child and Gray et al. (1998) find no relation between maternal hormones and sex of the child.

low SES children than for high SES children. More specifically, low SES boys have fewer behavioral problems when having a younger sister compared to boys with a younger brother. Girls are less disorganized and less asocial when they have a sister instead of a brother. This applies both for low and high SES girls, and both for youngest and oldest girls. We also find a negative effect of having a sister regarding self-esteem at age 10: Boys with low SES and an older sister have a lower self-esteem than boys with an older brother. For low SES girls, having a younger sister results in a lower self-esteem.

The positive effects on personality traits are reflected in more positive life outcomes. At a later age, low SES boys with a younger sister are more likely to have a healthier BMI and high SES boys score better on some achievement tests. None of the results for oldest girls remain significant when applying the Bonferroni correction. The effect of having an older sister for low SES girls is mixed. Both high and low SES girls perform better on several achievement tests when they have an older sister. A negative effect of having a sister remains significant when applying the Bonferroni correction for youngest boys on the BAS Word test at age 10 and for youngest high SES girls on the Sentence test at age 10.

In previous literature, boys and girls have been shown to display stereotypical behavior and thus possibly have a different influence on their sibling(s). These differences in behavior can stem from different parental treatment conditional on the gender of the child (McHale et al., 2003; Danielsbacka et al., 2011; Pollet et al., 2009), innate differences between boys and girls, and differences in the way siblings interact. Conley (2000) finds that boys are more competitive when they have a brother, and Peter et al. (2018) argue that competition between two brothers is a possible mechanism of why men choose higher paying occupations.

The limited amount of papers on the causal effect of sibling gender on outcomes in life stands in contrast with the larger literature on other sibling influences, such as birth order, age difference and number of siblings. Regarding the effects of sibling gender, most previous research has focused on the effects on educational attainment or wages. With respect to the effects on educational attainment, Butcher and Case (1994) show that women raised only with brothers receive on average significantly more education than women raised with sisters, controlling for household size. These results, however, could not be replicated by later studies. Kaestner (1997) finds that sibling sex composition had little effect on educational achievement. Hauser and Kuo (1998) use the Occupational Changes in a Generation Survey, the Survey of Income and Program Participation, and the National Survey of Families and Households and find no relationship between the gender composition of sibships and women's educational attainment. Regarding the effects of sibling gender on wages, Brenoe (2018) finds that women with a brother acquire more traditional gender norms with negative consequences for their labor earnings. Rao and Chatterjee (2018) show that the wages of men

are increasing in the proportion of male siblings, while the wages of women are insensitive to sibling gender. Peter et al. (2018) find that having a same-sex sibling increases men's earnings and family formation outcomes (marriage and fertility), as compared to an opposite-sex sibling. The results for women are similar but the effects are smaller in magnitude and less robust.

Cyron et al. (2017) is one of the few studies that investigate the effects of sibling gender on cognitive and non-cognitive skills. They find that boys benefit from having a sister in terms of cognitive skills, learning skills, and self-control measured in kindergarten. They also show that these effects have faded out by the time boys reach first grade. Girls' cognitive and non-cognitive skills are not affected by sibling gender. We contribute to this line of research by following a cohort of individuals for a longer period of time (measuring personality at ages 5, 10 and 16), analyzing a greater variety of personality traits, and by investigating consequences for a vector of outcomes across a large part of their lives. This chapter thereby provides a first overview of the breadth of the effects of sibling gender on personality traits and outcomes in multiple dimensions of life.

The chapter proceeds as follows. Section 3.2 describes the material and methods. Section 3.3 reports the results of the basic OLS regressions, addresses sample selectivity issues and shows results for robustness checks. Section 3.4 shows long term outcomes, and Section 3.5 concludes.

3.2 Materials and methods

We use data from the British Cohort Study (BCS70). This cohort follows children who were born in the same week in April 1970 up until the age of 50. In total, twelve surveys have been conducted over this timespan.¹¹ The data contain information on personality traits, and a wide variety of life outcomes, such as achievement tests, educational outcomes, health, and criminality. The size of the cohort and the broad scope of life outcomes make this dataset unique and allow us to investigate short- and long-term effects.

3.2.1 Sample

Our estimation sample consists of two-child families. We have data on the outcomes later in life of the children born in April 1970 and of the gender of the sibling. The latter information is available for siblings born before and after 1970, so the child we follow can either be the older or the younger of the two siblings. We select siblings with an age gap of maximum five years, because the data only

¹¹ These surveys are held at the ages of 2, 5, 10, 16, 21, 26, 29, 34, 38, 42, 46 and 50 (forthcoming).

report the date of a sibling's birth up until 1975.¹² About 85% of the oldest children of two-child families born in 1970 have a brother or sister by 1975. As a result of these selections, our remaining estimation sample contains 4771 children.

Since we follow children across a long time span, it is important to check whether there is selective attrition with respect to our main independent variable. Table 3.1 shows the OLS regression of the chance of still being in the sample at age 16 as a function of the gender of the child and the gender of the sibling. The results show that boys are less likely to remain in the sample, but most importantly, attrition does not correlate with the gender of the sibling and the sibling gender composition (indicated by the interaction term).

Table 3.1 Relationship between sibling gender and attrition at age 16

	Still in sample at age 16
Male	-0.086*** (0.016)
Same sex sibling	-0.003 (0.017)
Male * Same sex sibling	-0.008 (0.023)
Intercept	0.224*** (0.011)
N	4,738

Note: The table reports the OLS regression coefficients with a dummy for still being in the sample at age 16 as the dependent variable and a male dummy, a same sex sibling dummy, and an interaction between these two as the independent variables. Asterisks indicate significance levels where ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Standard errors are displayed in brackets underneath each coefficient. The regression is run on the two-child family sample.

3.2.2 Dependent variables

Personality traits

The data contain measures of the following personality traits: Rutter (anti-social behavior), self-esteem, locus of control, being disorganized, anti-social behavior, neuroticism and introversion. Rutter scores have been tested at three different ages of the child (5, 10, and 16 years old) and are answered by the mother. A full list of questions is displayed in Table A2.4. A higher Rutter score indicates more

¹² Besides the data restriction of a five year gap, the interaction between siblings with a large age gap is most likely very different than the interaction between siblings with a small age gap. When the age difference between siblings is large, the oldest sibling is more likely to take on the role of carer, while when the age gap is smaller, siblings may behave more as peers.

negative overall behavior of the child. The answering scale differed between the years. We standardize the principle components of the items to a mean of zero and standard deviation of one in order to be able to compare the estimates across the years.

Children were asked questions about their self-esteem at ages 10 and 16, and about their locus of control at age 10. An overview of the questions is provided in Tables A2.5 (self-esteem) and A2.6 (locus of control). A higher score on the self-esteem scale implies a higher self-esteem of the child. A higher score on the locus of control scale implies a more internal locus of control.

Questions about disorganized behavior, anti-social behavior, neuroticism and introversion are answered by the teacher of the child when the child was 10 years old. The questions about personality traits of the child were part of a bigger survey on the child's behavior. Table A2.7 shows the questions per personality trait. A higher score on the variables indicates respectively more disorganized, asocial, neurotic or introverted behavior by the child.

Achievement tests

Several achievement tests are taken at ages 5, 10 and 16. At the age of 5, three cognitive tests are conducted; the English Picture Vocabulary Test (EPVT), the Human Figure Drawing Test¹³, and the Copy Designs Test.⁵ At age 10, the child participates in nine tests, two of which are related to math skills, and the other seven are related to spelling and reading. At age 16, three more tests are conducted; a math, a spelling and a vocabulary test. We standardize all the achievement test scores to a mean of zero and a standard deviation of one.

Educational outcomes

The data include three measures of educational outcomes. At age 16, we have information on the Grade Point Average (GPA) of the child. At age 26, the age the child left school, and the highest obtained qualification are measured.

Other life outcomes

There are several later in life outcomes in the dataset, but we focus on four main ones, covering outcomes regarding health, criminality, happiness, and labor market participation. Three of these measures (health, criminality, and happiness)

13 The English Picture Vocabulary Test measures verbal vocabulary, the Human Figure Drawing Test reflects conceptual maturity, and the Copy Design Test assesses visual-motor coordination.

are taken from the survey at age 34. Health is measured by a dummy variable indicating whether or not the respondent has a healthy BMI.¹⁴ Criminal behavior is measured by whether the participant has ever been in contact with the police. Furthermore, happiness is measured by a series of questions about life satisfaction regarding several aspects of life, which are standardized to a mean of zero and standard deviation of one. Next to these three aspects, labor market outcomes are measured by log wages at age 38.

3.2.3 Socio-economic status (SES)

We have information about the social class of the father¹⁵ of the children we follow, which is an important indicator for the environment in which the child is raised. Table A3.1 displays a list of occupations related to each social class. The highest social class indicates professional occupations, while the lowest social class indicates unskilled workers. We use these social classes to create two categories for socio-economic status (SES) of the parents. We define the high SES category by the two highest social classes and the low SES category by the other social classes. Table A3.2 shows the number of children in each social class for every outcome group. The percentage of children per gender combination is comparable between the low and high SES groups.

3.3 Methodology

We use a sample of two-child families in order to keep family size constant and to avoid complicating the analysis by including multiple gender exposures (e.g. two sisters, or a sister and a brother, or a brother and a sister, and so on). The gender of a child is random, implying that we can use Ordinary Least Squares regressions to analyze the effect of the sibling's gender on personality traits and several life outcomes of the child born in 1970.

However, note that while parents cannot choose the sex of their child, they can choose the number of children they have. To see how such a selective sample can bias our results, consider the following example. Suppose that poor parents are more likely to have a third child if their first two children are of the same

¹⁴ BMI is measured as weight in kilograms divided by height in meters squared. A healthy BMI lies between 18.5 and 25 BMI points.

¹⁵ The data reports the SES of the husband of the mother of the child. We assume that this is either the biological father of the child or the most present father figure in the child's life. We include the social class of the father, since a very high proportion of the mothers were housewives.

sex. This implies that of the two-child households in our sample, households with same sex siblings will predominantly be of higher SES. If educational achievement is lower in poor families, it appears as if having a sibling of the same sex has a positive effect on educational achievement, while the reason instead is that poor families are underrepresented in two-child households with children of the same sex. Therefore, our two-child family sample may be selective because the existence of more than two children depends on the gender combination of the first two children and on parental characteristics. Following this rationale, we need to check whether parents who decide to have a second child after having a first-born boy (girl) are different from parents who decide not to have a second child after having a boy (girl).

In order to determine whether parents who decide to have a second child are different from parents who only have one child depending on the gender of the first born, we perform a balancing test on the SES of the father. Additionally, we check whether parental characteristics (i.e. SES) are different for parents who decide to have an additional child when the first two children are both boys (girls/mixed gender) compared to parents who decide not to have a third child after having two boys (girls/mixed gender). The results for these balancing tests are displayed in Table 3.2. It is clear from these tables that parents with different SES background make different decisions regarding the number of children. The table reveals that parents with a low SES are more likely to have a third child than parents with high SES in case the first two children are of the same gender. We correct for potential selection bias with respect to SES by running our analyses for low and high SES categories separately. A second analytical issue is that we test many hypotheses. This can result in significant outcomes due to chance, i.e. a type I error may occur. We employ Bonferroni tests to account for such false positives.

Table 3.2 Relationship between parental characteristics and the chance of having a third child**Table 3.2A** Balance test 1 child vs. 2 children

Gender first born		1 child	Had a second child	Difference
Boy	SES (0=low, 1=high)	0.171	0.200	0.030
		(0.377)	(0.400)	(0.022)
	Observations	452	2,465	2,917
Girl	SES (0=low, 1=high)	0.165	0.204	0.039*
		(0.372)	(0.403)	(0.022)
	Observations	467	2,273	2,740

Note: This table shows the balance test results between parents who decided to stop having children after having the first child of a certain gender versus parents who decided to have a second child. The first column shows the gender of the first-born child. The columns “1 child” and “Had a second child” report the mean value of the SES of the parents with respect to the gender of the first-born and the decision whether to have a second child. The last column reports the difference in the two previous columns. Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.1$. Standard errors are displayed in brackets underneath each coefficient.

Table 3.2B Balance test 2 children vs. 3 children

Gender combination first 2 children		2 children	Had a third child	Difference
2 boys	SES	0.207	0.178	-0.029
	(0=low, 1=high)	(0.405)	(0.383)	(0.018)
	Observations	1,244	880	2,124
2 girls	SES	0.201	0.167	-0.033*
	(0=low, 1=high)	(0.401)	(0.374)	(0.020)
	Observations	1,053	716	1,769
boy-girl	SES	0.193	0.158	-0.035*
	(0=low, 1=high)	(0.395)	(0.365)	(0.019)
	Observations	1,221	694	1,915
girl-boy	SES	0.207	0.161	-0.046**
	(0=low, 1=high)	(0.405)	(0.368)	(0.019)
	Observations	1,220	698	1,918

Note: This table shows the balance test results between parents who decided to stop after having two children with a certain gender combination versus parents who decided to have a third child. The first column shows the gender combination of the first two children. The columns “2 children” and “Had a third child” report the mean value of the SES of the parents with respect to the gender of the first two children and the decision whether to have a third child. The last column reports the difference in the two previous columns. Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.1$. Standard errors are displayed in brackets underneath each coefficient.

3.4 Results

Table 3.3 shows the baseline results of the OLS regressions on personality traits for 4 groups separately: (1) boys who are the oldest, (2) boys who are the youngest, (3) girls who are the oldest, and (4) girls who are the youngest. In these analyses, we do not separate the results yet by SES.

Oldest boys have fewer behavioral problems at ages 5, 10 and 16 (about 0.11-0.15 standard deviations, from now on s.d.) and are less disorganized (0.13 s.d.) when they have a younger sister instead of a brother. Younger boys also have fewer behavioral problems at ages 5 and 10 when they have a sister (about 0.11 s.d.), but they also have lower self-esteem (0.2 s.d.). Oldest girls have fewer behavioral problems (0.13-0.15 s.d.), are less disorganized (0.32 s.d.), and show less asocial (0.20 s.d.) and introverted behavior (0.11 s.d.), but also a lower self-esteem (0.15 s.d.) when they have a sister. Youngest girls have fewer behavioral problems (0.10 s.d.), are less disorganized (0.29 s.d.), and show less asocial (0.23 s.d.) and introverted behavior (0.09 s.d.) when they have a sister.

As indicated above, our sample may be selective because the number of children parents choose to have is endogenous. To address this concern, we run the regressions separately for low and high SES groups. Table 3.4 displays the results of the OLS regressions on personality traits for the 4 groups, and separately for low and high SES groups. Table 3.4A displays the results for boys, Table 3.4B for girls. Oldest boys from the low SES group display fewer behavioral problems up until age 16 (0.11-0.19 s.d.) and less disorganized behavior (0.14 s.d.) when they have a sister. For the oldest boys from the high SES group, we observe only a small effect on behavioral problems (0.21 s.d. at age 10) when having a sister. Youngest boys from the low SES group show fewer behavioral problems up until age 10 (0.13 s.d.) but a lower self-esteem (0.19 s.d.) when they have a sister. Youngest boys from the high SES group have a more internal locus of control (0.18 s.d.) and are less introverted (0.19 s.d.) when they have a sister.

Table 3.3 Relationship between sibling gender and personality traits

	Group 1: Boys — oldest		Group 2: Boys — youngest		Group 3 — Girls — oldest		Group 4: Girls — youngest	
	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N
Rutter behavioral difficulties (5)	0.153*** (0.042)	2922	0.106** (0.049)	1816	-0.128*** (0.044)	2922	-0.053 (0.054)	1816
Rutter behavioral difficulties (10)	0.110** (0.045)	2467	0.113** (0.054)	1601	-0.052 (0.047)	2467	-0.104* (0.059)	1601
Rutter behavioral difficulties (16)	0.155*** (0.055)	1833	-0.061 (0.056)	1455	0.013 (0.055)	1833	0.071 (0.061)	1455
Self-esteem (10)	0.052 (0.043)	2176	0.200*** (0.052)	1442	-0.146*** (0.045)	2176	0.025 (0.057)	1442
Self-esteem (16)	0.064 (0.077)	1103	0.066 (0.087)	810	-0.042 (0.073)	1103	-0.113 (0.085)	810
Locus of control (10)	-0.055 (0.043)	2169	0.076 (0.052)	1417	-0.009 (0.045)	2169	0.061 (0.055)	1417
Disorganized (10)	0.129*** (0.049)	2110	0.078 (0.061)	1392	-0.317*** (0.050)	2110	-0.286*** (0.066)	1392
Asocial (10)	0.062 (0.045)	2247	0.066 (0.055)	1484	-0.203*** (0.047)	2247	-0.229*** (0.060)	1484
Neurotic (10)	-0.019 (0.047)	2269	0.003 (0.056)	1490	-0.068 (0.049)	2269	0.020 (0.061)	1490
Inverted (10)	0.035 (0.041)	2263	0.060 (0.049)	1490	-0.107** (0.043)	2263	-0.087* (0.054)	1490

Note: Each cell in a 'Same sibling sex' column represents a separate OLS regression. The table reports OLS regressions of various options of sibling sex as the independent variable (displayed in the column headings) on various personality traits as dependent variables (displayed in the first column of the table). Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.01$. Standard errors are displayed in brackets underneath each coefficient. All variables are standardized with mean zero and standard deviation of one. All results are based on data on two-child families. Group 1 reports the results for boys who are the oldest child in a two-child family and compares oldest boys who have a sister with oldest boys who have a brother. The 'Same sex sibling' is a dummy which takes on a value of 1 if the oldest boy has a younger brother. The same reasoning holds for the other three groups. Coefficients in bold are significant under the Bonferroni multiple hypothesis testing rule: $\alpha = 0.1$; $0.1/10 \approx 0.01$.

Table 3.4 Relationship between sibling gender and personality for high and low SES categories
Table 3.4A Results for boys

	Group 1a:			Group 1b:			Group 2a:			Group 2b:		
	Boys — oldest — low SES	Boys — oldest — high SES		Boys — oldest — high SES	Boys — youngest — low SES	Boys — youngest — high SES	Boys — youngest — low SES	Boys — youngest — high SES		Boys — youngest — low SES	Boys — youngest — high SES	
	Same sex sibling	Same sex sibling	N	Same sex sibling	Same sex sibling	Same sex sibling	Same sex sibling	Same sex sibling	N	Same sex sibling	Same sex sibling	N
Rutter behavioral difficulties (5)	0.181*** (0.049)	0.029 (0.088)	2202	0.029 (0.088)	0.129** (0.058)	0.039 (0.090)	0.129** (0.058)	0.039 (0.090)	1379	0.039 (0.090)	0.039 (0.090)	385
Rutter behavioral difficulties (10)	0.107** (0.051)	0.209** (0.098)	1875	0.209** (0.098)	0.125** (0.063)	0.039 (0.105)	0.125** (0.063)	0.039 (0.105)	1215	0.039 (0.105)	0.039 (0.105)	343
Rutter behavioral difficulties (16)	0.192*** (0.063)	0.030 (0.118)	1392	0.030 (0.118)	-0.053 (0.064)	-0.093 (0.121)	-0.053 (0.064)	-0.093 (0.121)	1104	-0.093 (0.121)	-0.093 (0.121)	313
Self-esteem (10)	0.068 (0.049)	0.039 (0.101)	1669	0.039 (0.101)	0.189*** (0.061)	0.171 (0.107)	0.189*** (0.061)	0.171 (0.107)	1098	0.171 (0.107)	0.171 (0.107)	306
Self-esteem (16)	0.060 (0.090)	0.056 (0.162)	819	0.056 (0.162)	0.121 (0.104)	-0.107 (0.165)	0.121 (0.104)	-0.107 (0.165)	592	-0.107 (0.165)	-0.107 (0.165)	192
Locus of control (10)	-0.070 (0.050)	-0.042 (0.092)	1661	-0.042 (0.092)	0.026 (0.060)	0.183* (0.105)	0.026 (0.060)	0.183* (0.105)	1080	0.183* (0.105)	0.183* (0.105)	300
Disorganized (10)	0.138** (0.056)	0.117 (0.114)	1616	0.117 (0.114)	0.069 (0.070)	0.128 (0.124)	0.069 (0.070)	0.128 (0.124)	1059	0.128 (0.124)	0.128 (0.124)	294
Asocial (10)	0.089* (0.052)	-0.018 (0.093)	1730	-0.018 (0.093)	0.056 (0.066)	0.156 (0.096)	0.056 (0.066)	0.156 (0.096)	1137	0.156 (0.096)	0.156 (0.096)	310
Neurotic (10)	-0.038 (0.054)	0.077 (0.114)	1747	0.077 (0.114)	-0.022 (0.065)	0.043 (0.119)	-0.022 (0.065)	0.043 (0.119)	1140	0.043 (0.119)	0.043 (0.119)	310
Introverted (10)	0.028 (0.047)	0.117 (0.102)	1735	0.117 (0.102)	0.038 (0.057)	0.189* (0.106)	0.038 (0.057)	0.189* (0.106)	1133	0.189* (0.106)	0.189* (0.106)	319

Note: Each cell in a 'Same sibling sex' column represents a separate OLS regression. The table reports OLS regressions of various options of sibling sex as the independent variable (displayed in the column headings) on various life outcomes as dependent variables (displayed in the first two columns on the left of the table). Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.01$. Standard errors are displayed in brackets underneath each coefficient. All variables are standardized with mean zero and standard deviation of one. All results are based on data on two-child families. Group 1a reports the results for boys who are the oldest child in a two-child family with a low SES and compares oldest boys from a low SES family who have a sister with oldest boys from a low SES family who have a brother. The 'Same sex sibling' is a dummy which takes on a value of 1 if the oldest boy has a younger brother. The same reasoning holds for the other three groups of boys. Coefficients in bold are significant under the Bonferroni multiple hypothesis testing rule: $\alpha = 0.1$; $0.1/10 \approx 0.01$.

Table 3.4B Results for girls

	Group 1a: Girls — oldest — low SES			Group 1b: Girls — oldest — high SES			Group 2a: Girls — youngest — low SES			Group 2b: Girls — youngest — high SES		
	Same sex sibling	N		Same sex sibling	N		Same sex sibling	N		Same sex sibling	N	
Rutter behavioral difficulties (5)	-0.126** (0.051)	2202		-0.152 (0.093)	522		-0.044 (0.063)	1379		-0.054 (0.100)	385	
Rutter behavioral difficulties (10)	-0.074 (0.053)	1875		0.067 (0.103)	437		-0.051 (0.068)	1215		-0.244** (0.118)	343	
Rutter behavioral difficulties (16)	0.028 (0.062)	1392		0.046 (0.120)	331		0.046 (0.069)	1104		0.153 (0.132)	313	
Self-esteem (10)	-0.185*** (0.051)	1669		0.035 (0.107)	371		0.048 (0.066)	1098		-0.025 (0.116)	306	
Self-esteem (16)	-0.029 (0.083)	819		-0.045 (0.164)	225		-0.055 (0.101)	592		-0.232 (0.167)	192	
Locus of control (10)	-0.017 (0.052)	1661		0.009 (0.096)	374		0.018 (0.064)	1080		0.191* (0.115)	300	
Disorganized (10)	-0.289*** (0.058)	1616		-0.398*** (0.117)	366		-0.316*** (0.075)	1059		-0.290*** (0.136)	294	
Asocial (10)	-0.204*** (0.054)	1730		-0.170* (0.095)	380		-0.265*** (0.071)	1137		-0.093 (0.107)	310	
Neurotic (10)	-0.051 (0.056)	1747		-0.204* (0.116)	384		0.011 (0.070)	1140		0.001 (0.135)	310	
Introverted (10)	-0.112** (0.049)	1735		-0.078 (0.104)	387		-0.155** (0.061)	1133		0.136 (0.117)	319	

Note: Each cell in a 'Same sibling sex' column represents a separate OLS regression. The table reports OLS regressions of various options of sibling sex as the independent variable (displayed in the column headings) on various life outcomes as dependent variables (displayed in the first two columns on the left of the table). Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.01$. Standard errors are displayed in brackets underneath each coefficient. All variables are standardized with mean zero and standard deviation of one. All results are based on data on two-child families. Group 3a reports the results for girls who are the oldest child in a two-child family with a low SES and compares oldest girls from a low SES family who have a brother with oldest girls from a low SES family who have a sister. The 'Same sex sibling' is a dummy which takes on a value of 1 if the oldest girl has a younger sister. The same reasoning holds for the other three groups of girls. Coefficients in bold are significant under the Bonferroni multiple hypothesis testing rule: $\alpha = 0.1$; $0.1/10 \approx 0.01$.

Oldest girls from the low SES group display fewer behavioral problems at age 5 (0.13 s.d.), and are less disorganized (0.29 s.d.), asocial (0.20 s.d.) and introverted (0.11 s.d.) when they have a sister. They also have a lower self-esteem (0.19 s.d.) compared to girls with a younger brother. Oldest girls from the high SES group are less disorganized (0.40 s.d.), asocial (0.17 s.d.) and neurotic (0.20 s.d.) when they have a sister. Youngest girls from the low SES group are less disorganized (0.40 s.d.), asocial (0.27 s.d.) and introverted (0.16 s.d.) when they have a sister. Youngest girls from the high SES group display slightly fewer behavioral problems (0.24 at age 10) and are less disorganized (0.19 s.d.) when they have a sister. The fact that the effects of the gender of the sibling remain significant conditional on SES could be seen as an indication that the correlations in our main analysis can be interpreted as causal relations.

The bold coefficients in Tables 3.3 and 3.4 indicate which results remain significant when using the Bonferroni multiple hypothesis testing correction at the $\alpha=0.1$ significance level. Having 10 personality traits outcome variables, this method implies that results with a p-value of 0.01 are significant at a 90 percent confidence level. The effects for oldest boys on behavioral problems remain significant. The results for the youngest boys remain significant with respect to self-esteem. For the oldest low SES girls, the results on self-esteem, disorganized and asocial behavior remain significant, just as the high SES girls' result on disorganized behavior. The results on disorganized and asocial behavior of the youngest girls from the low SES group remain significant as well.

3.5 Consequences for life outcomes

Table 3.5 shows the results of the OLS regressions on other life outcomes for the 4 groups, and separately for low and high SES groups. Table 3.5A shows the results for the boys, Table 3.5B for the girls. Oldest boys from the low SES group are more likely to have a healthy BMI (0.12 s.d.) and are less likely to have been in contact with the police (0.12 s.d.) when they have a sister. Oldest boys from the high SES group score lower on several achievement tests (0.2-0.6 s.d.) and are more likely to have a healthy BMI (0.13 s.d.) when they have a sister. Youngest boys from the low SES group are more likely to have a healthy BMI (0.14 s.d.) when they have a sister but have a higher GPA (0.15 s.d.) when they have a brother. Youngest boys from the high SES group perform worse on several achievement tests (0.2-0.5 s.d.) but are less likely to have been in contact with the police (0.13 s.d.) when they have a sister.

Table 3.5 Relationship between sibling gender and various life outcomes for high and low SES categories
Table 3.5A Results for boys

	Group 1a: Boys — oldest — low SES		Group 1b: Boys — oldest — high SES		Group 2a: Boys — youngest — low SES		Group 2b: Boys — youngest — high SES	
	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N
EPVT cognitive test (age 5)	0.152*** (0.051)	2063	0.133 (0.105)	498	0.114* (0.059)	1296	0.286*** (0.111)	361
Human figure drawing cognitive test (age 5)	-0.082 (0.050)	2165	-0.297*** (0.101)	518	0.003 (0.061)	1360	0.063 (0.123)	381
Copy designs cognitive test (age 5)	-0.014 (0.050)	2190	0.182* (0.097)	522	0.084 (0.063)	1375	0.291*** (0.113)	384
Edinburgh Reading Test (age 10)	-0.120 (0.096)	586	-0.025 (0.156)	177	-0.288** (0.115)	410	-0.061 (0.137)	137
Friendly Math Test (age 10)		512	0.321* (0.163)	129	0.193* (0.106)	395	0.360** (0.171)	107
Vocabulary Test (age 10)	-0.010 (0.055)	1788	-0.241** (0.114)	395	-0.069 (0.067)	1172	0.095 (0.117)	323
Reading Test (age 10)	-0.063 (0.056)	1786	-0.041 (0.116)	394	0.043 (0.073)	1170	0.045 (0.129)	322
Spelling Test (age 10)	-0.070 (0.061)	1328	-0.182* (0.108)	311	-0.105 (0.076)	888	-0.023 (0.116)	258
Sequence Test (age 10)	-0.106* (0.057)	1611	-0.045 (0.106)	373	-0.100 (0.070)	1066	0.150 (0.121)	292
Sentence Test (age 10)	-0.004 (0.058)	1668	0.164 (0.135)	372	0.021 (0.075)	1102	-0.153 (0.105)	308
BAS Math test (10)	-0.007 (0.064)	1356	-0.046 (0.118)	325	0.083 (0.072)	892	0.262* (0.135)	262
BAS Word test (10)	0.059 (0.056)	1746	-0.082 (0.138)	383	0.279*** (0.068)	1151	0.589*** (0.146)	315
Math Score (age 16)	-0.042 (0.099)	579	-0.670*** (0.181)	159	-0.107 (0.114)	425	0.021 (0.160)	133

Achievement tests

Spelling Test (age 16)	-0.013 (0.083)	955	-0.026 (0.125)	262	-0.062 (0.097)	695	-0.032 (0.124)	238
Vocabulary Test (age 16)	-0.133 (0.084)	955	-0.124 (0.120)	262	-0.217** (0.101)	695	-0.033 (0.151)	238
Grade Point Average (16)	-0.038 (0.075)	1039	0.084 (0.124)	315	0.149* (0.084)	782	0.185 (0.140)	281
Age left school (26)	-0.045 (0.068)	1309	0.052 (0.121)	352	-0.111 (0.076)	890	0.162 (0.142)	267
Highest qualification (26)	-0.025 (0.068)	1263	0.094 (0.133)	347	-0.111 (0.079)	858	0.205 (0.146)	263
Body Mass Index (34)	-0.115*** (0.023)	2202	-0.134*** (0.050)	522	-0.143*** (0.030)	1379	-0.092 (0.059)	385
Crime (34)	0.166*** (0.059)	1433	-0.135 (0.096)	363	0.101 (0.073)	968	0.134** (0.061)	277
Life Satisfaction (34)	-0.083 (0.064)	1432	-0.156 (0.103)	363	-0.082 (0.075)	968	-0.044 (0.133)	277
Wages (38)	0.073 (0.048)	613	-0.040 (0.040)	163	0.034 (0.057)	415	0.081	119
Other life outcomes								

Note: Each cell in a 'Same sibling sex' column represents a separate OLS regression. The table reports OLS regressions of various options of sibling sex and SES category as the independent variable (displayed in the column headings) on various life outcomes as dependent variables (displayed in the first two columns on the left of the table). Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.01$. Standard errors are displayed in brackets underneath each coefficient. All variables are standardized (with mean zero and standard deviation of one) except for the BMI and crime variables, which are dummies. BMI takes on a value of 1 if someone has a healthy BMI and 0 if the BMI is unhealthy. The crime dummy takes on a value of 1 if someone has ever been in contact with the police, and a value of 0 if they have not. All the results are based on data on two-child families. Group 1A reports the results for boys who are the oldest child in a two-child low SES family and compares oldest boys who have a sister with oldest boys who have a brother. The 'Same sex sibling' is a dummy which takes on a value of 1 if the oldest boy from a low SES family has a younger brother. The same reasoning holds for the other three groups. Coefficients in bold are significant under the Bonferroni multiple hypothesis testing rule: $\alpha = 0.1$; $0.1/10 \approx 0.01$.

Table 3.5B Results for girls

	Group 1a:				Group 1b:				Group 2a:				Group 2b:			
	Girls — oldest — low SES		Girls — oldest — high SES		Girls — youngest — low SES		Girls — youngest — high SES		Girls — youngest — low SES		Girls — youngest — high SES		Girls — youngest — low SES		Girls — youngest — high SES	
	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N	Same sex sibling	N
EPVT cognitive test (age 5)	-0.092* (0.054)	2063	-0.173 (0.111)	498	-0.21*** (0.064)	1296	-0.235* (0.123)	361								
Human figure drawing cognitive test (age 5)	0.019 (0.052)	2165	0.020 (0.106)	518	0.20*** (0.066)	1360	0.324** (0.136)	381								
Copy designs cognitive test (age 5)	-0.066 (0.052)	2190	-0.089 (0.102)	522	0.017 (0.069)	1375	0.152 (0.125)	384								
Edinburgh Reading Test (age 10)	-0.140 (0.102)	586	-0.213 (0.157)	177	-0.073 (0.120)	410	0.283* (0.153)	137								
Friendly Math Test (age 10)	-0.175* (0.105)	512	0.082 (0.188)	129	-0.092 (0.123)	395	-0.011 (0.186)	107								
Vocabulary Test (age 10)	0.018 (0.057)	1788	-0.074 (0.117)	395	-0.083 (0.071)	1172	0.117 (0.130)	323								
Reading Test (age 10)	0.002 (0.059)	1786	-0.098 (0.119)	394	0.017 (0.078)	1170	-0.020 (0.142)	322								
Spelling Test (age 10)	0.157** (0.063)	1328	0.277*** (0.108)	311	0.114 (0.079)	888	0.201 (0.130)	258								
Sequence Test (age 10)	0.112* (0.060)	1611	-0.078 (0.111)	373	0.112 (0.076)	1066	0.286** (0.134)	292								
Sentence Test (age 10)	-0.028 (0.060)	1668	-0.002 (0.137)	372	-0.029 (0.079)	1102	-0.345*** (0.119)	308								
BAS Math test (10)	0.045 (0.065)	1356	0.034 (0.121)	325	0.011 (0.077)	892	0.407*** (0.151)	262								
BAS Word test (10)	-0.165*** (0.058)	1746	-0.206 (0.142)	383	-0.192*** (0.073)	1151	0.141 (0.162)	315								
Math Score (age 16)	-0.008 (0.097)	579	-0.075 (0.183)	159	-0.062 (0.115)	425	-0.107 (0.167)	133								

Achievement tests

Spelling Test (age 16)	0.117 (0.077)	955	-0.027 (0.117)	262	0.106 (0.096)	695	0.084 (0.132)	238
Vocabulary Test (age 16)	0.081 (0.078)	955	-0.071 (0.112)	262	-0.032 (0.098)	695	0.277* (0.161)	238
Grade Point Average (16)	0.079 (0.070)	1039	0.073 (0.123)	315	0.113 (0.089)	782	0.284* (0.147)	281
Age left school (26)	0.014 (0.065)	1309	0.088 (0.121)	352	0.077 (0.076)	890	0.241* (0.142)	267
Highest qualification (26)	0.061 (0.065)	1263	0.011 (0.133)	347	0.006 (0.078)	858	0.246* (0.145)	263
Body Mass Index (34)	0.064*** (0.024)	2202	0.064 (0.053)	522	0.067** (0.033)	1379	0.107 (0.065)	385
Crime (34)	-0.116** (0.058)	1433	-0.192** (0.095)	363	-0.040 (0.076)	968	-0.000 (0.066)	277
Life Satisfaction (34)	0.080 (0.063)	1432	-0.005 (0.101)	363	-0.069 (0.078)	968	-0.063 (0.143)	277
Wages (38)	-0.078 (0.050)	613	-0.023 (0.096)	163	0.012 (0.062)	415	0.090 (0.172)	119
Other life outcomes								

Note: Each cell in a 'Same sibling sex' column represents a separate OLS regression. The table reports OLS regressions of various options of sibling sex and SES category as the independent variable (displayed in the column headings) on various life outcomes as dependent variables (displayed in the first two columns on the left of the table). Asterisks indicate significance levels where *: $p < 0.01$, **: $p < 0.05$, ***: $p < 0.01$. Standard errors are displayed in brackets underneath each coefficient. All variables are standardized (with mean zero and standard deviation of one) except for the BMI and crime variables, which are dummies. BMI takes on a value of 1 if someone has a healthy BMI and 0 if the BMI is unhealthy. The crime dummy takes on a value of 1 if someone has ever been in contact with the police, and a value of 0 if they have not. All the results are based on data on two-child families. Group 3A reports the results for girls who are the oldest child in a two-child low SES family and compares oldest girls who have a sister with oldest girls who have a brother. The 'Same sex sibling' is a dummy which takes on a value of 1 if the oldest girl from a low SES family has a younger sister. The same reasoning holds for the other three groups. Coefficients in bold are significant under the Bonferroni multiple hypothesis testing rule: $\alpha = 0.1$; $0.1/10 \approx 0.01$.

Oldest girls from the low SES group are more likely to have a healthy BMI (0.06 s.d.) and are less likely to have been in contact with the police (0.12 s.d.) when they have a sister. The effect of sibling gender on their achievement tests are mixed. Oldest girls from the high SES group are less likely to have been in contact with the police (0.19 s.d.) when they have a sister and score better on the Spelling test at age 10 (0.28 s.d.). Youngest girls from the low SES group are more likely to have a healthy BMI (0.07 s.d.) when they have a sister and display mixed results on achievement tests. Youngest girls from the high SES group are more likely to have a higher GPA (0.28 s.d.) and a higher qualification at age 26 (0.25 s.d.) but are also more likely to be older when they finish their education (0.24 s.d.) when they have a sister. Youngest girls from the high SES group display mixed results of the effect of sibling gender on achievement tests as well.

We also correct these results for multiple hypothesis testing (p-value of 0.004 applies at a significance level of 0.10). The coefficients that are significant under this Bonferroni hypothesis test are displayed in bold in Table 3.5. The results which remain significant are as follows: Oldest boys from the low SES have a healthier BMI if they have a younger sister but score lower on the EPVT cognitive test at age 5. Oldest boys from the high SES group perform better on two achievement tests when they have a younger sister. Both low and high SES youngest boys perform worse at the BAS Word test at age 10 when they have a sister. For oldest girls, none of the results remain significant under the Bonferroni tests. For youngest girls from the low SES group, the two results on achievement tests at age 5 that remain significant are ambiguous. Girls from the high SES group with an older sister score worse on the Sentence test at age 10.

3.6 Conclusion and discussion

This chapter studies the effects of the gender of the sibling on personality traits and a large vector of life outcomes. We employ a rich British data set, allowing us to follow children from birth until age 38. The results indicate that boys and girls generally benefit from having a sister in terms of having personality traits typically linked to positive life outcomes. We find that oldest boys from both low and high SES groups display fewer behavioral problems, and oldest boys from the low SES group also show less disorganized behavior when they have a sister. Youngest boys from the low SES group display fewer behavioral problems but also have a lower self-esteem when they have a sister. Youngest boys from the high SES group are less introverted and have a more internal locus of control when they have a sister. When we look at long run outcomes, we find that boys are overall less likely to have been in contact with the police and more likely to have a healthy BMI when they have a sister. For girls we find that overall, they display less

disorganized behavior when they have a sister. Furthermore, oldest girls from the low SES group are less asocial and introverted, but also have a lower self-esteem when they have a sister. Oldest girls from the high SES group display less asocial and neurotic behavior when they have a sister. Youngest girls from the low SES group show less asocial and introverted behavior and youngest girls from the high SES group display fewer behavioral problems when they have a sister. We find a similar patterns for long run outcomes for girls as we found for boys: oldest girls from both SES groups and youngest girls from the low SES group are less likely to have been in contact with the police and are more likely to have a healthy BMI when they have a sister.

These results have implications for parents and policy makers. Children in two-child families who have a brother are at risk with respect to having personality traits which have been shown to negatively relate to outcomes in certain domains. Parents and policy makers may use this information to target the children in need for help in these domains. The results can be used by policy makers to better inform teachers and parents about the possible long-term risks for children with a brother. Extra attention from parents and teachers, for instance with respect to advice on a healthy diet and negative consequences of criminal behavior, could mitigate the negative effects of having a brother. Future research could explore possible mechanisms, such as gender norms, to further explain these results.

3.7 Appendix

Table A3.1 Construction of socio-economic status (SES) via social class of father

SES high	Social class I	Professional occupations, including doctors, lawyers, ministers of religion, university teachers, professional engineers, etc.
	Social class II	Managerial and other processionals, including nurses, school teachers, company directors, etc.
SES low	Social class III NM	Non-manual skilled occupations, including ship assistants, company representatives, clerical workers, draughtsman, etc.
	Social class III M	Skilled manual workers, including mechanics, craftsmen of all types, skilled engineers, etc.
	Social class IV	Semi-skilled workers, including machine operators, postmen, storekeepers, porters, caretakers, etc.
	Social class V	Unskilled workers, including laborers, cleaners, dustmen, etc.

Table A3.2 Division over low and high SES per gender combination

SES	Boy-boy		Girl-girl		Boy-girl		Girl-boy		Total	
	N	%	N	%	N	%	N	%	N	%
Low	938	79%	792	80%	944	81%	907	79%	3,581	80%
High	245	21%	199	20%	226	19%	237	21%	907	20%
Total	1,183	100%	991	100%	1,17	100%	1,144	100%	4,488	100%

Chapter 4

How does the inability to observe primary school exit test scores affect the teacher's assessment of immigrants? † ‡

† Joint work with Bart Golsteyn and Inge de Wolf.

‡ We thank Ilja Cornelisz and Chris van Klaveren for their valuable comments.

4.1 Introduction

Over the past years, the number of students with a migratory background enrolled in Dutch primary schools has increased to 25% of the total Dutch primary school population. This number includes all second and first generation immigrants. Especially the recent inflow of refugees into Europe has increased the number of first generation non-Western immigrants (Inspectie van het Onderwijs, 2019). This shift in the average number of primary school students with a migration background has refueled the debate on the integration of immigrants. For young immigrants, the educational system of the host country is especially important, since this is one of the main channels of integration.

The aim of this chapter is to investigate how the teachers' assessment of the educational level of immigrants is affected by the inability to observe the results of the school exit test at the end of primary education. In the Netherlands, the teachers' assessment is a crucial element in the educational careers of students, because the teacher assessment functions as an entry ticket to a specific secondary education level. The tracking takes place at the end of primary school at a relatively early age (about 12 years old)¹⁶ and the teachers assign students to one of 8 educational tracks that make up the three main Dutch secondary education levels (varying from 4 to 6 years of secondary schooling).¹⁷ The Dutch secondary school track assignments depend on a nationwide school exit test score (objective ability measure) and an assessment from the teacher about which secondary education track would be the best fit (subjective ability measure).¹⁸

A recent policy implementation changed the track assignment procedure, as well as the timing of the school exit test and teacher assessment. Before the policy change, students first took the school exit test and then the teacher would give an

16 Only Austria, Germany, Hungary, and the Slovak Republic track at an earlier age (10 years).

Early tracking systems have been documented to be harmful for the long run educational and human capital outcomes of students, especially those with a migration background (see e.g. Checchi & Flabbi, 2006; Hanushek & Woessmann, 2006; Brunello & Checchi, 2007; Bol & van de Werfhorst, 2013; Ludemann & Schwerdt, 2013; Hopwood et al., 2016).

17 The three main Dutch secondary education school levels are pre-vocational (including vocational preparatory and theoretical preparatory education), secondary general and pre-university education. More details on the Dutch education system are given in section 4.2 and Figures A4.2 and A4.3.

18 The difference between objective and subjective ability measures has been related to several background characteristics of students and to discrimination (Dee, 2005; Lavy, 2008; Gibbons & Chevalier, 2008; De Boer et al., 2010; Burgess & Greaves, 2013; Glock et al., 2015; Timmermans et al., 2015; Pit-ten Cate et al., 2016; Feron, 2018).

assessment (after having observed the test score). After the policy implementation, teachers had to give an assessment before the students had taken the school exit test. The identification strategy of this chapter exploits this policy change and uses a difference-in-differences framework to compare the difference in teacher assessments and test scores pre- and post-policy change for different migrant groups relative to natives. We test two conflicting hypotheses. On the one hand, moving from a track placement that is partly based on a test score to a teacher assessment-based track placement in the post-policy period, might result in *higher* teacher assessments for migrants. This is due to the fact that migrants may score lower on the exit test than their true ability because of language barriers. When a teacher cannot observe this biased exit test score anymore, he/she will turn to the learning capabilities of the migrant students. Since these migrant students are still catching up, their learning curve is steep, and a teacher may recognize this catching up process and give an assessment based on the future potential of the child. On the other hand, when teachers cannot observe the school exit test score (after the policy implementation), they might use information on the mean test scores of students with comparable characteristics. This could lead to teacher bias and statistical discrimination, which could potentially result in *lower* teacher assessments for migrants after the policy change. Furthermore, since first generation immigrants have not spent their entire primary school career in the Netherlands as opposed to natives and second generation immigrants, we expect the policy effect to be larger for first generation immigrants than for second generation immigrants compared to natives.

To test the effect of the inability to observe the exit test score on the teacher assessment of migrants compared to natives, we use unique administrative data. These Dutch administrative educational data provide information on education outcomes and background characteristics of the full primary school student population of the Netherlands in the period 2008-2016. This allows us to identify different migrant groups and educational outcomes before (6 years) and after (3 years) the policy change.

The main results of the analyses are that the policy change decreased the relative advantage migrant children had compared to natives in terms of the discrepancy between teacher assessments and exit test scores. Before the policy change, students with a migrant background were more likely to receive a higher teacher assessment than their test score would suggest compared to native students. After the policy change, the chance to receive a higher teacher assessment than exit test score increased more for native students, thus lowering the relative advantage of migrant students. These effects hold as well for a subsample of schools which participated in the same school exit test throughout the sample period, decreasing the chance that the results are driven by selectivity. Robustness checks that exclude the first year after the policy implementation and that use revised

teacher assessments display similar results as well, indicating that the policy implementation negatively affected the relative advantage of migrant students compared to natives in terms of receiving a higher teacher assessment compared to the exit test score, especially for first generation non-Western migrants.

This chapter contributes to studies on the effects of this policy change by focusing on migrant groups. Earlier literature focused on the effect of the policy change on secondary school track changing. Swart et al. (2019) show that students are more likely to change secondary education tracks when the teacher no longer uses the test score while giving an assessment, i.e. the situation after the policy change. Feron (2018) shows that teachers gave more favorable assessments to girls and students with highly educated parents after the policy change. It contributes to the international literature on the discrepancy between objective and subjective ability measures (e.g. Burgess & Greaves, 2013; Dee, 2005) by exploiting a unique policy change.

The remainder of this chapter is organized as follows. Section 4.2 provides an overview of the Dutch education system and explains the policy change. Section 4.3 describes the conceptual framework of the school exit test scores and teacher assessments pre- and post-policy change. The empirical strategy is discussed in Section 4.4, and Section 4.5 describes the data. Section 4.6 presents the findings. Section 4.7 concludes and discusses the results.

4.2 Dutch institutional background

The compulsory school age in the Netherlands ranges from ages 5 to 16,¹⁹ but usually students enter the first grade when they are 4 years old. Primary school consists of 8 grades, including 2 years of kindergarten. The Netherlands has a relatively early tracking system; students are on average 11 or 12 years old when they take a school exit test and receive an assessment by their teacher about the secondary school track placement. There are 8 secondary education tracks that make up the three main secondary education levels: pre-university education (6 years), secondary general education (5 years), and theoretical and vocational preparatory education (4 years).²⁰ The secondary education track placement

¹⁹ When students do not have a secondary vocational, a secondary general or a pre-university degree before the age of 16, they have to stay in school until the age of 18.

²⁰ Besides the three main secondary education levels, the Netherlands also has a special education and a practical secondary education level. These two types of education levels are not included in this chapter, since they are designed for students with low IQ's and learning and behavioral difficulties. Due to the distinct learning abilities of the students that enroll at special and practical education, often they do not partake in the school exit test and/or have been enrolled at a special primary school.

determines the tertiary education enrollment options level. Figure A4.1 displays an overview of the Dutch education system.

Track placement is based on two signals about the ability of the child: a test score and an assessment from the teacher about which track would be the best fit. At the end of primary school, students participate in a *school exit test*, a 2-day test which consists of multiple-choice questions about mathematics and language. Optionally, study skills and science can be tested as well, but the results on these topics are not taken into account in the calculation of the final test score. The average test scores are comparable across years, since the test scores are calibrated every year on a scale from 500 – 550. It is important to note that the answers are not checked by teachers, but by a machine. Therefore, the test score can be seen as a fully objective measure of a student's ability (blind testing). Table A4.1 displays how test scores can be converted to secondary track suggestions.²¹ The *teacher assessment*, on the other hand, is a subjective measure (non-blind testing) for student ability. Teachers can use several sources of information in their decision process, such as previous test scores, experiences of other teachers, learning attitude and motivation of the student, gender, socio-economic status (SES), or ethnicity. There is no strong incentive for teachers to give as high as possible assessments, since their compensation scheme does not depend on the level of their assessments. Additionally, Feron (2018) points out that there is a repeated interaction between the school population and the secondary schools they enroll at, which aligns the incentives and increases the understanding of the ability signals between primary school teachers and secondary schools over time.

Policy change

In 2014, there was a change in the Netherlands in the procedure for track assignment of students at the end of primary education. The basic notion behind this policy change was that the teacher assessment should be leading in the enrollment in secondary education instead of the exit test score of students. This policy change was a combination of three changes in the assessment procedure of students at age 12. First of all, this resulted in a change in the timing of test scores and teacher assessments. Before the policy change, students would first take the school exit test, so that when the teacher carried out their assessment,

²¹ The organization that creates the most used school exit test, Cito, suggests a secondary education track placement for each Cito test score (on a scale of 500-550). The Dutch Education Inspectorate uses a similar method to transfer school exit test scores into suggested track placements. Table A4.1 shows the method of the Dutch Education Inspectorate. We use this method to scale teacher assessments and test scores on the same ordinal 8-point scale, since this method is the most widely used method by teachers throughout the Netherlands.

he/she could take the test score into account. From 2014 onwards, the teacher instead had to give the assessment before the students took the test. A detailed overview of the pre- and post-policy timing of events are shown in Figure A4.2. Secondly, the new policy made the participation in a school exit test obligatory for students. Prior to 2014, schools could decide which students would participate in the exit test and which students would not be participating in the exit test. Since school quality as measured by the Dutch Education Inspectorate is partly based on the test scores of a school, schools had an incentive to only let the better students participate in the test. Thirdly, the market for test providers was opened up. Before the policy change, there was mainly one school exit test available; the Cito test. However, after the policy implementation, schools were able to choose among different tests.

The motivation behind the policy change to delay the test was threefold. First, the teacher cannot observe the test score while giving an assessment. Arguably, this results in the teacher using information on the development of the student through a longer duration, and on the student's characteristics such as the ability to concentrate, persistence, etc. instead of a one-time ability measure as reported by the test score. Second, students are forced to remain attentive in school for a longer period of time instead of spending time on non-educational activities after the exit test. The assumption here was that once students had taken the school exit test, they would not pay much attention in class anymore, since what they did after the test did not count towards the track assignment decision. Third, the time between the availability of the test score and the start of the secondary education is shorter. This could lead to less skill deterioration and a more accurate ability measure for secondary education institutions (Rijksoverheid, 2016a; Rijksoverheid, 2016b; Feron, 2018).

4.3 Conceptual framework

4.3.1 Pre policy change

The secondary school track assignment is determined by two signals about the ability of the student. If track assignment happens only conditionally on the test scores achieved on the primary school exit test, then the track assignment function can be represented by the following function:

$$(1) \quad T_i(a_i(a_i^*, \varepsilon_i))$$

The track assignment depends on a proxy of student ability (a_i), but this proxy is measured with error (ε_i) and depends on the true but non-observed student ability (a_i^*). The expectation of teachers (ET_i) about the ability of student i (and thus about which track assignment is the best fit) might not only be driven by achieved test scores on the exit exam, but might also be conditional on previous ability proxies $a_{i,t-k}$ and student background characteristics. Since we are interested in how ethnical background affects tracking assignment, we distinguish between ethnical background characteristics, Z_i , and other background characteristics, X_i . The teacher track assignment function for student i can be represented as:

$$(2) \quad ET_i(a_i(a_i^*, \varepsilon_i), a_{i,t-k}, Z_i, X_i), \text{ for } k = 1, 2, \dots, K.$$

ET_i represents the track assignment based on the expectation of the teacher and if track assignment of teachers is purely conditioned on the suggested track assignment by the primary school exit test, then ET_i is equivalent to T_i :

$$(3) \quad ET_i = T_i$$

The expectation of the teacher (ET_i) will differ from the test score signal (T_i) when T_i gives a noisy signal about the ability of the student. When a teacher can observe previous test scores ($a_{i,t-k}$), he/she can observe whether T_i is in line with these previous ability proxies. If a student has a good (bad) test day and a teacher corrects for this one-time over (under) performance, ET_i will be lower (higher) than T_i .

4.3.2 Post policy change

After the policy change, students take the school exit test once they have received their teacher assessment. This means that teachers are not able to observe the test score anymore while forming an expectation on the ability of the student. Hence, formula (2) becomes:

$$(4) \quad ET_i(a_{i,t-k}, Z_i, X_i), \text{ for } k = 1, 2, \dots, K.$$

After the policy change, the teacher expectation (ET_i) can only be based on previous ability proxies ($a_{i,t-k}$) and background characteristics such as migration background (Z_i), gender and age (X_i).

Once students have received the teacher assessment, they participate in the school exit test. Afterwards, teachers are able to observe the exit test scores and if the track assignment signal of the test score is higher than the track assignment signal of the initial teacher assessment, the teacher is allowed to adjust his/her

assessment of a student upwards. The expectation of the teacher after being able to observe the test score is based on the same information as prior to the policy change, and formula (2) can be used for the teacher expectation again.

4.3.3 Hypotheses

This chapter investigates the effect of the policy change on track assignment for migrant students in the Netherlands compared to natives. The paper exploits the effect of the teacher assessment with and without a school exit test score (before and after the policy change), particularly for migrant children. Two conflicting hypotheses are tested. The first hypothesis is that the teacher assessment will be closer to the true ability of a migrant student without a school exit test score, because the test scores are relatively low for migrant students (due to a language barrier). This hypothesis is in line with one of the motivations for introducing the policy: the teacher assessment of migrant children is higher when a teacher cannot observe the school exit test score. Therefore, the share of children with a migration background who receive a higher track placement based on their teacher assessment than the suggested track based on their test score ($ET > T$) should *increase* compared to the share of natives who receive $ET > T$ after the policy change.

The second (conflicting) hypothesis is that teacher assessments are less reflective of the migrant student's ability without a school exit test score because of teacher bias and statistical discrimination of migrant students. When a teacher has less information on ability measures of students, he/she will use information on the mean score of students with the same e.g. gender or ethnicity. Burgess & Greaves (2013) find proof of teacher bias and statistical discrimination with respect to migration background for migrants in the United Kingdom, and Alesina et al. (2018) find evidence for this relationship among Italian teachers. This conflicting hypothesis suggests that the share of children with a migration background who receive a higher track placement based on their teacher assessment than the suggested track based on their test score ($ET > T$) should *decrease* compared to the share of natives who receive $ET > T$ after the policy implementation.

Additionally, we expect larger effects of the policy change for the first generation immigrants. This is due to the fact that most of the first generation immigrants have not been enrolled at a Dutch primary school for the entire primary school age, resulting in language barriers when taking the school exit test. Second generation immigrants have been enrolled at a Dutch primary school from the first grade onwards and should not face any language barriers (or, at the very least, significantly lower barriers compared to first generation immigrants) when taking the school exit test.

4.4 Empirical analysis

The empirical strategy relies on a difference-in-differences (DiD) approach in which the discrepancy in teacher assessments and test scores of three groups of migrants are compared to the discrepancy in teacher assessments and test scores of natives before and after the policy change. The basic specification is as follows:

$$I. \quad ET_i - T_i = \alpha_0 + \alpha_1 [Migrant_i \times Policy_i] + \alpha_2 Policy_i + \alpha_3 Migrant_i + \alpha_4 X_i + \varepsilon_i$$

where the outcome variable is the difference in the track assignment as derived from the teacher assessment (ET_i) and the track assignment as derived from the test score (T_i), $Migrant_i$ is a dummy which takes on a value of 1 if the student has a migration background,²² $Policy_i$ is a dummy which takes on a value of 1 if the student received a teacher assessment after the policy change, X_i includes dummies for gender, age when taking the test and age at arrival (for first generation immigrants), and ε_i is an error term. We are interested in the DiD estimator α_1 , i.e., the interaction between the policy and migrant dummies. This estimator reports how the policy affected the difference in teacher assessment and test score for three categories of students with a migration background compared to natives. This chapter distinguishes between second generation immigrants, first generation Western immigrants, and first generation non-Western immigrants, and we compare each group with native students. It is possible that the policy change affected the native students (control group) as well. The DiD estimates the effect of the policy change on migrant children as a net effect: i.e. the policy effect on migrants minus the policy effect on natives.

The identification assumption in a DiD framework is that there is no difference in trends between the migrant groups and natives before the policy implementation. Pre-policy common trends suggest that the initial differences between migrants and natives would have remained similar if the policy had not been implemented. To check whether the common trends assumption holds, we conduct an event study analysis with the following specification:

$$II. \quad ET_i - T_i = \beta_0 + \beta_1 [Migrant_i \times Policy_i] + \beta_2 [Migrant_i \times 2012] + \beta_3 [Migrant_i \times 2011] + \beta_4 [Migrant_i \times 2010] + \beta_5 [Migrant_i \times 2009] + \beta_6 [Migrant_i \times 2008] + \beta_7 Migrant_i + \beta_8 X_i + \gamma_i + \varphi_i$$

22 The first specification compares second generation migrants to natives, the second specification compares first generation Western migrants to natives, and the third specification compares first generation non-Western migrants to natives.

where the outcome variable ($ET_i - T_i$), the $Migrant_i$ dummy, the $Policy_i$ dummy, and X_i dummies are the same as in specification I, the interaction between $Migrant_i$ and pre-policy years (2012-2008) estimates the difference between migrants and natives in the pre-policy period, γ_i are year fixed effects and includes the $Policy_i$ dummy as year fixed effects for the post-policy period, and φ_i is an error term. The baseline year is 2013 (the year prior to the policy implementation). Therefore, this year is not included in any of the pre-policy and migrant dummy interactions or year fixed effects.

The policy effect on the difference in teacher assessments and test scores of migrants as compared to natives is given by β_j , the estimator of the $Migrant_i$ and $Policy_i$ interaction. This should be almost identical to the DiD estimator α_i of specification I. The common trends assumption holds if the pre-policy differences ($\beta_2, \beta_3, \beta_4, \beta_5$, and β_6) are jointly not significantly different from zero, which we check with an F-test.

4.4.1 Selection check

As noted above, during the policy change, participation in a school exit test also became obligatory. Moreover, at the same time, the market for these school exit tests was opened up to more than just the Cito test provider, resulting in a choice set for schools of several tests. The quality of a school is partly measured by the average test score of that school, which may drive the school's decisions to participate in exit tests and in their choice of the type of test. Prior to the policy change, schools and students did not have to participate in the Cito test. It is plausible that schools decided not to let the worse performing students participate in the Cito test, in order to avoid low aggregate test scores, and thus a lower quality indication. After the policy implementation, schools which performed worse on the Cito test may have switched to a different exit test the year afterwards (Inspectie van het Onderwijs, 2018). In other words, better performing schools may typically have continued to use the Cito test. Since we only select students for whom we can observe both a teacher assessment and the Cito test score, a selection issue might occur when we compare groups of students pre- and post-policy implementation.

Table 4.1 reports the percentage of students for whom (1) both the teacher assessment and Cito test score is known, (2) only a teacher assessment is known, (3) only the Cito test score is known, or (4) both are unknown. These percentages are reported for natives and the three migration groups separately, and before and after the policy implementation. The increase in the percentage of students for whom both their teacher assessment and their Cito test score are observable is the

largest for native students.²³ After the policy change, natives were more likely to be observed in our specifications, and first generation non-Western immigrants were the least likely to be included.

In order to check whether the results of the population are driven by selectivity, we perform the DiD and Event study analysis for a subsample of schools that have participated in the Cito test throughout the whole sample period (2008-2016). This subsample does not suffer from selectivity with respect to changing to another type of test after the policy implementation.²⁴ We argue that if the subsample results do not significantly differ from the population results, selectivity does not drive the population results.

Table 4.1 Description of which information is available for which students

		Natives	Second generation	First generation Western	First generation non-Western
Pre policy	ET and T known	64.9%	64.02%	56.16%	50.87%
	Only ET known	5.63%	6.12%	6.55%	6.98%
	Only T known	17.12%	15.44%	18.52%	19.89%
	Both unknown	12.95%	14.41%	18.77%	22.26%
	N	815,610	217,204	16,391	13,903
Post policy	ET and T known	90.88%	87.75%	79.86%	68.75%
	Only ET known	2.97%	4.13%	4.45%	5.64%
	Only T known	4.39%	4.75%	9.26%	13.50%
	Both unknown	1.76%	3.37%	6.43%	12.10%
	N	333,277	97,350	6,489	5,421

Note: This table reports the percentages of students for whom (1) both teacher assessment (ET) and Cito test score (T) are known, (2) only teacher assessment is known, (3) only Cito test score is known, and (4) both are unknown, per migration group and for pre- and post-policy separately.

²³ Even though the policy stated that everyone has to participate in a school exit test, the percentages reported are not 100%. This is due to three reasons: (1) if students participated in another test, they are not included, (2) if students received an assessment for special or practical education, which are education types for students with a low IQ and/or behavioral problems, they are not included in our sample as well since, and (3) students with a migration background and who were in the Netherlands for less than 4 years at the time of the test could be exempted from taking the test on the basis of language problems (the test is completely in Dutch).

²⁴ However, this method does not solve the problem of selectivity with respect to excluding worse performing students before the policy implementation.

4.5 Data

This chapter uses Dutch administrative data that include all primary school enrolments in the period 2008-2016. The data include information on several demographic variables (such as gender, age, country of origin, arrival date), school information, exit test scores²⁵ and teacher assessments. We define four migration background groups. The first group are the natives: students for whom both parents are Dutch (77 percent of the students).²⁶ The second group are second generation immigrants: students who are born in the Netherlands, but who have at least one parent who was born abroad (21 percent of the students).²⁷ The third and fourth groups are first generation immigrants: students who were born abroad and who have at least one parent who was born abroad. A further distinction can be made according to the country of origin. The third group includes first generation Western immigrants (1 percent of the students) and the fourth group contains first generation non-Western immigrants (1 percent of the students).²⁸

The outcome variable, the difference between teacher assessments and exit test scores, is calculated as follows. Both teacher assessments and exit test scores can be scaled to an ordinal scale of 8,²⁹ as displayed in Table A4.1. The difference is calculated by subtracting the teacher assessment by the test score (i.e. $ET - T$). When a teacher assessment is equal to the test score ($ET = T$), the difference equals 0. When the teacher assessment is higher than the test score ($ET > T$), the difference equals 1, and when the teacher assessment is lower than the test score ($ET < T$), the difference equals -1.³⁰

25 We only include students for whom we know the Cito test score. After the policy change, some students participated in a different exit school test, but these results are not comparable to the Cito results. Furthermore, the vast majority of the students who took a test, participated in the Cito test.

26 The percentages reported in the text are calculated based on the selection of students for whom both a test score and teacher assessment is known.

27 We do not make a distinction between Western and non-Western second generation immigrants, since they are all born in the Netherlands. They have been enrolled at a Dutch primary school from the first grade onwards, and thus should not face any language barriers while taking the school exit test.

28 First generation Western immigrants and their parent(s) are born in Europe (excl. Turkey), North-America, Oceania, Japan or Indonesia. First generation non-Western immigrants and their parent(s) are born in Africa, South-America, Asia or Turkey.

29 In Chapter 5 of this thesis, the teacher assessments are scaled on a 10-point ordinal scale. This chapter only uses an 8-point scale, because practical education and the assessment 'vmbo-k/vmbo-g(t)' are not taken into account when rescaling the test scores into secondary education tracks.

30 Note that we group all the values of $ET - T > 0$ (ranging from 1 to 7) into the value of 1, and all the values of $ET - T < 0$ (ranging from -1 to -7) into the value -1.

4.6 Results

The results for the native students and the three migrant background groups are described separately. The estimation results compare the results of the second generation, first generation Western, and first generation non-Western migrants with native students.

4.6.1 Descriptive analyses

Figure 4.1 reports the mean test scores for each migrant group per year. The test scores are given on a scale of 1 (lowest secondary education track) to 8 (highest track). The native students receive the highest mean test scores (5.4 – 5.6), and the first generation non-Western migrants the lowest (4.4 – 4.6). The first generation Western migrants perform a bit better (mean test scores of 4.9 – 5.2) than the second generation migrants (mean test scores of 4.8 – 5.0) on the school exit test. The pattern does not dramatically change after the policy change in 2014.

Figure 4.1

Mean test score trend

Note: This figure reports the mean test scores for the years 2008–2016 for four migration background categories: natives (solid line), second generation (dotted line), first generation Western (dashed line), and first generation non-Western (dashed-dotted line). The vertical line indicates when the policy change took effect (2014 was the first year after the introduction of the policy). Test scores are reported on a scale of 1–8 (see Table A4.1).

Figure 4.2 shows the mean teacher assessment for each migrant group per year. The teacher assessments are given on the same scale as the test scores, ranging from 1 (lowest track) to 8 (highest track). The patterns of the teacher assessments are comparable to the patterns of the mean test scores; natives receive the highest assessments (mean assessments of 5.3 – 5.5) and first generation non-Western migrants the lowest (mean assessments of 4.4 – 4.6), and first generation Western immigrants receive slightly higher assessments (mean assessments of 4.7 – 5.1) compared to second generation migrants (mean assessments of 4.6 – 5.0). However, the differences between the migrant groups in terms of teacher assessments is a bit smaller than the differences in the test scores. This is due to the fact that the mean teacher assessments of natives and first generation Western immigrants are slightly lower than their mean test scores. The dashed lines in the years after the policy implementation report the revised teacher assessments. This revised teacher assessment can occur after the policy implementation when a student has a higher track assessment than suggested by the test score. In that case the teacher can correct his/her initial assessment upwards. Thus, the revised teacher assessment is always higher than the initial teacher assessments. The upwards shift of the assessment line to the revised assessment line is smallest for natives.

Figure 4.2
Mean teacher assessment trend

Note: This figure reports the mean teacher assessments for the years 2008-2016 for four migration background categories: natives (solid line), second generation (dotted line), first generation Western (dashed line), and first generation non-Western (dashed-dotted line). The vertical line indicates when the policy change took effect (2014 was the first year after the introduction of the policy). Teacher assessments are reported on a scale of 1-8 (see Table A4.2). The dashed lines indicate the revised teacher assessments: i.e. the teacher assessments after the teachers have observed the school exit test score.

Figure 4.3 documents our variable of interest and the main outcome in specifications I and II: the difference between teacher assessments and test scores. Before the policy change, first generation non-Western immigrants had a mean difference close to zero, whilst other migrants and natives had negative mean values. One potential explanation is that teachers are aware of the steep learning curve and current language barriers (and thus lower exit test scores) of first generation non-Western immigrants, and correct the test score with a higher assessment, based on higher expectations for the future ability of these students. The first year after the policy change, the mean difference between teacher assessments and test scores declined for all the students (in other words; more students received a teacher assessment lower than their eventual test score), with the highest drop for first generation immigrants and the lowest drop for natives. In the years after this initial drop, the differences between teacher assessments and test scores move towards zero difference. Since the shift in mean difference holds for all the migrant groups and natives, the relative advantage of the migrant children with respect to receiving a higher teacher assessment disappears after the policy change.

Figure 4.3

Mean difference (ET-T) trend

Note: This figure reports the difference in teacher assessments and test scores for the years 2008-2016 for four migration background categories: natives (solid line), second generation (dotted line), first generation Western (dashed line), and first generation non-Western (dashed-dotted line). The vertical line indicates when the policy change took effect (2014 was the first year after the introduction of the policy). Teacher assessments and test scores are reported on a scale of 1-8 (see Table A4.2).

4.6.2 Empirical results

In order to measure the size of the policy effect for each migrant group, we turn to the results of the DiD and Event study estimations. Table 4.2 displays the results of the DiD analyses for the total population and the subsample. The first row of Table 4.2 reports the coefficient of the interaction between the different migrant groups (second generation, first generation Western, and first generation non-Western migrants) and the $Policy_i$ dummy on the difference between teacher assessment and test scores. The $Migrant_i$ dummy compares a specific migrant group (column A is the second generation, column B the first generation Western, and column C the first generation non-Western migrants) to native students. The positive coefficients of $Migrant_i$ for all the migrant groups indicate that migrants receive a higher teacher assessment compared to the test score relatively to natives. After the policy change, the teacher assessment on average increases compared to the test score, as demonstrated by the positive coefficient for the $Policy_i$ dummy. To answer the question what the effect is of the inability to observe the exit test score on the teacher assessment of migrants, we turn to the DiD estimator: the interaction between $Migrant_i$ and the $Policy_i$ dummy. As Figure 4.3 already shows, the teacher assessment is lower than the test score most of the time ($ET < T$). The coefficient of the interaction term is negative and significant across all specifications. The negative sign of the interaction term indicates that more students receive a teacher assessment that is lower than their test score. Thus, the significant negative coefficient of the interaction term illustrates that the relative advantage of migrants compared to natives in terms of receiving higher teacher assessments compared to test scores decreases after the policy change. The first generation Western and non-Western immigrants display the largest negative policy effect: the share of first generation Western and non-Western immigrants who receive a teacher assessment that is lower than their test score (i.e. $ET < T$) increases by 0.08-0.09 compared to the share of natives who receive $ET < T$ after the policy change. The share of second generation immigrants who receive $ET < T$ after the policy increases by about 0.06 compared to natives.

Table 4.2 DiD estimation results

	(A) 2nd generation				(B) 1st generation Western				(C) 1st generation non-Western			
	All		Subsample		All		Subsample		All		Subsample	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Migrant*Policy	-0.062*** (0.004)	-0.063*** (0.004)	-0.063*** (0.007)	-0.064*** (0.007)	-0.067*** (0.013)	-0.068*** (0.013)	-0.085*** (0.026)	-0.087*** (0.026)	-0.075*** (0.018)	-0.077*** (0.018)	-0.075*** (0.035)	-0.078*** (0.035)
Policy	0.029*** (0.002)	0.0271*** (0.002)	0.044*** (0.004)	0.042*** (0.004)	0.029*** (0.002)	0.027*** (0.002)	0.044*** (0.003)	0.041*** (0.003)	0.029*** (0.002)	0.027*** (0.001)	0.044*** (0.003)	0.041*** (0.003)
Migrant	0.090*** (0.002)	0.093*** (0.002)	0.101*** (0.004)	0.105*** (0.004)	0.076*** (0.008)	0.087*** (0.008)	0.079*** (0.015)	0.090*** (0.015)	0.109*** (0.010)	0.124*** (0.010)	0.130*** (0.019)	0.146*** (0.019)
Gender	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Age taking test	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Age at arrival	No	No	No	No	No	No	No	No	No	No	No	No
Constant	-0.131*** (0.001)	0.243*** (0.016)	-0.155*** (0.002)	0.228*** (0.032)	-0.131*** (0.001)	0.270*** (0.018)	-0.155*** (0.002)	0.254*** (0.036)	-0.131*** (0.001)	0.270*** (0.018)	-0.155*** (0.002)	0.252*** (0.036)
Observations	1,051,776	1,051,776	256,490	256,490	841,692	841,692	202,537	202,537	835,346	835,346	201,110	201,110
R-squared	0.002	0.002	0.003	0.003	0.001	0.001	0.001	0.002	0.000	0.001	0.001	0.002

Note: The policy dummy takes on a value of 1 when the test score and teacher assessment has been given in 2014/2015, 2015/2016 or 2016/2017. The migrant dummy in columns A takes on a value of 1 if the child is a 2nd generation immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns B takes on a value of 1 if the child is a 1st generation Western immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is 1st generation non-Western immigrant and a value of 0 if the child is native Dutch. Controls include gender (male equals 1), age at arrival, and age when taking the school exit test. Standard errors are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The identification assumption of the DiD results is based on the existence of a common trend prior to the policy change. The Event study, as reported in Table 4.3, tests this assumption. The last row of Table 4.3 shows the p-values of the F-test that all the pre-policy interaction terms are jointly equal to zero. When these interactions are not jointly significant, there is a common trend. This is the case for the first generation Western and non-Western migrants when we include controls for gender, age when taking test and age at arrival. Furthermore, the coefficients of the interaction between the $Migrant_i$ and $Policy_i$ dummy of the DiD and Event study regressions are not significantly different for the first generation Western and non-Western immigrants. The common trend assumption for second generation migrants does not hold.

To check whether these results are driven by our selection of schools with respect to which students partake in the school exit test and the type of test, we compare the DiD and Event study results of the total population to the DiD and Event study results of the subsample, which are also reported in Tables 4.2 and 4.3. The coefficients of the interaction between the $Migrant_i$ and $Policy_i$ dummy of the subsample are not significantly different from the total population. Therefore, we argue that the results of the total population are not driven by selectivity. For the subsample, the policy effect is also negative and significant across all migrant groups, with the largest effect for the first generation Western migrants. The common trends assumption holds across all migrant groups and all specifications in the subsample. The fact that the policy effect is the largest for the first generation non-Western immigrants is in line with our hypothesis that the inability to observe the test score (which is assumed to be lower for first generation immigrants due to language barriers) would have a larger impact for these migrants than for second generation migrants.

Table 4.3 Event study

	(A) 2nd generation			(B) 1st generation Western			(C) 1st generation non-Western					
	All		Subsample	All		Subsample	All		Subsample		Subsample	
	(a)	(b)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(c)
Migrant*Policy	-0.044*** (0.006)	-0.044*** (0.011)	-0.051*** (0.011)	-0.0427*** (0.0202)	-0.064*** (0.026)	-0.001 (0.043)	-0.002 (0.043)	0.004 (0.056)	-0.082*** (0.028)	-0.087*** (0.028)	-0.103*** (0.032)	-0.088* (0.053)
Migrant*2012	0.028*** (0.007)	0.028*** (0.014)	0.014 (0.014)	0.0515*** (0.0244)	0.062*** (0.031)	0.104** (0.051)	0.105** (0.05)	0.105** (0.068)	0.001 (0.033)	0.002 (0.033)	0.012 (0.038)	0.032 (0.064)
Migrant*2011	0.011 (0.007)	0.011* (0.014)	0.009 (0.014)	0.0492** (0.0244)	0.0321 (0.032)	0.141*** (0.051)	0.140*** (0.051)	0.166** (0.068)	0.020 (0.032)	0.020 (0.032)	-0.000 (0.038)	0.023 (0.062)
Migrant*2010	-0.020*** (0.009)	-0.020*** (0.015)	0.005 (0.015)	-0.0168 (0.0244)	0.020 (0.033)	0.032 (0.046)	0.031 (0.046)	0.061 (0.063)	-0.105*** (0.032)	-0.102*** (0.032)	-0.066* (0.039)	-0.005 (0.054)
Migrant*2009	0.015 (0.009)	0.016* (0.015)	0.008 (0.015)	0.0409 (0.0253)	0.044* (0.025)	0.064 (0.044)	0.067 (0.044)	0.086 (0.061)	0.002 (0.033)	0.005 (0.033)	0.019 (0.041)	0.043 (0.053)
Migrant*2008	0.033*** (0.008)	0.033*** (0.014)	0.014 (0.014)	0.000579 (0.0235)	-0.001 (0.024)	0.049 (0.044)	0.047 (0.044)	0.010 (0.059)	0.033 (0.031)	0.032 (0.031)	-0.006 (0.038)	-0.061 (0.052)
Migrant	0.072*** (0.005)	0.075*** (0.010)	0.089*** (0.010)	0.092*** (0.0172)	0.062*** (0.017)	-0.006 (0.037)	0.005 (0.037)	-0.097** (0.048)	0.116*** (0.023)	0.129*** (0.023)	0.042 (0.027)	0.143*** (0.045)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Age taking test	No	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Age at arrival	No	No	No	No	No	No	No	Yes	No	No	Yes	No
Constant	-0.109*** (0.002)	0.260*** (0.016)	-0.133*** (0.005)	-0.109*** (0.00219)	0.286*** (0.018)	-0.133*** (0.005)	0.270*** (0.036)	0.410** (0.164)	-0.109*** (0.002)	0.285*** (0.018)	0.473*** (0.090)	-0.133*** (0.005)
Observations	1,051,776	1,051,776	256,490	841,692	841,692	33,846	202,537	8,251	835,346	835,346	27,540	201,110
R-squared	0.002	0.003	0.003	0.001	0.001	0.004	0.001	0.005	0.001	0.001	0.004	0.001
p-value F-test	0.000	0.000	0.725	0.030	0.023	0.117	0.135	0.194	0.008	0.009	0.252	0.499

Note: The policy dummy takes on a value of 1 when the test score and teacher assessment has been given in 2014/2015, 2015/2016 or 2016/2017. The migrant dummy in columns B takes on a value of 1 if the child is a 2nd generation immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is a 1st generation Western immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is 1st generation non-Western immigrant and a value of 0 if the child is native Dutch. Controls include gender (male equals 1), age at arrival, and age when taking the school exit test. Year fixed effects include dummies for all the years (and thus also the Policy dummy), except for 2013 (i.e. baseline year/year prior to the policy change). Standard errors are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.6.3 Robustness checks

The trends in Figures 4.1-4.3 show that the effect of the policy is larger in the first year after implementation, namely in 2014, than in the years 2015-2016. To check whether the policy results are driven by the apparent extremity of this year, we run the DiD and Event study regressions without the inclusion of 2014. Table 4.4 reports the results of the DiD specification without 2014. The interaction term is negative and significant between all specifications. Furthermore, the results for the first generation Western and non-Western immigrants are not significantly different in the analyses based upon the population and the subsample. Table 4.5 displays the results of the Event study without 2014. The interaction term is negative across all specifications and significant for the second generation and first generation non-Western migrants. Again, the results for the first generation Western and non-Western migrants of the Event study without 2014 are not significantly different in analyses on the population and subsample.

Table 4.4 Robustness check 1: DiD estimation without 2014

	(A) 2nd generation		(B) 1st generation Western			(C) 1st generation non-Western		
	No controls	Controls — gender & age test	No controls	Controls — gender & age test	All controls	No controls	Controls — gender & age test	All controls
Migrant*Policy	-0.0479*** (0.00420)	-0.0487*** (0.00420)	-0.0461*** (0.0149)	-0.0476*** (0.0149)	-0.0728*** (0.0196)	-0.0792*** (0.0212)	-0.0796*** (0.0212)	-0.103*** (0.0247)
Policy	0.0664*** (0.00198)	0.0643*** (0.00198)	0.0664*** (0.00197)	0.0641*** (0.00197)	0.0871*** (0.0129)	0.0664*** (0.00197)	0.0641*** (0.00197)	0.0872*** (0.0129)
Migrant	0.0775*** (0.00204)	0.0804*** (0.00205)	0.0638*** (0.00730)	0.0739*** (0.00731)	-0.0218** (0.00966)	0.102*** (0.00942)	0.115*** (0.00944)	0.0225* (0.0116)
Gender	No	Yes	No	Yes	Yes	No	Yes	Yes
Age taking test	No	Yes	No	Yes	Yes	No	Yes	Yes
Age at arrival	No	No	No	No	Yes	No	No	Yes
Constant	-0.135*** (0.000937)	0.218*** (0.0159)	-0.135*** (0.000934)	0.249*** (0.0179)	0.454*** (0.0812)	-0.135*** (0.000934)	0.250*** (0.0180)	0.443*** (0.0900)
Observations	1,051,212	1,051,212	840,977	840,977	33,226	834,854	834,854	27,144
R-squared	0.002	0.003	0.001	0.002	0.004	0.001	0.002	0.004

Note: The policy dummy takes on a value of 1 when the test score and teacher assessment has been given in 2014/2015, 2015/2016 or 2016/2017. The migrant dummy in columns A takes on a value of 1 if the child is a 2nd generation immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns B takes on a value of 1 if the child is a 1st generation Western immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is 1st generation non-Western immigrant and a value of 0 if the child is native Dutch. Controls include gender (male equals 1), age at arrival, and age when taking the school exit test. Standard errors are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.5 Robustness check 1: Event study estimation without 2014

	(A) 2nd generation		(B) 1st generation Western			(C) 1st generation non-Western		
	No controls	Controls — gender & age test	No controls	Controls — gender & age test	All controls	No controls	Controls — gender & age test	All controls
Migrant*Policy	-0.0186*** (0.00497)	-0.0191*** (0.00497)	-0.00686 (0.0183)	-0.00740 (0.0183)	-0.0345 (0.0237)	-0.0648** (0.0258)	-0.0633** (0.0258)	-0.0907*** (0.0298)
Migrant*2012	0.0518*** (0.00586)	0.0523*** (0.00586)	0.0783*** (0.0216)	0.0801*** (0.0216)	0.0924*** (0.0277)	0.0290 (0.0289)	0.0318 (0.0289)	0.0438 (0.0337)
Migrant*2011	0.0345*** (0.00598)	0.0349*** (0.00598)	0.0753*** (0.0216)	0.0766*** (0.0216)	0.0608** (0.0279)	0.0463* (0.0281)	0.0481* (0.0281)	0.0296 (0.0332)
Migrant*2010	-0.00968 (0.00862)	-0.00917 (0.00862)	-0.00175 (0.0239)	-0.000411 (0.0239)	0.0320 (0.0324)	-0.0918*** (0.0312)	-0.0885*** (0.0312)	-0.0558 (0.0380)
Migrant*2009	0.0251*** (0.00923)	0.0261*** (0.00922)	0.0543** (0.0249)	0.0573** (0.0249)	0.0697** (0.0343)	0.0176 (0.0325)	0.0209 (0.0325)	0.0314 (0.0400)
Migrant*2008	0.0479*** (0.00807)	0.0479*** (0.00806)	0.0160 (0.0223)	0.0148 (0.0223)	-0.0144 (0.0304)	0.0493* (0.0291)	0.0492* (0.0291)	0.0160 (0.0357)
Migrant	0.0482*** (0.00334)	0.0508*** (0.00334)	0.0246* (0.0129)	0.0337*** (0.0129)	-0.0598*** (0.0164)	0.0873*** (0.0174)	0.0987*** (0.0174)	0.00984 (0.0202)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	Yes	No	Yes	Yes	No	Yes	Yes
Age taking test	No	Yes	No	Yes	Yes	No	Yes	Yes
Age at arrival	No	No	No	No	Yes	No	No	Yes
Constant	-0.131*** (0.00155)	0.223*** (0.0159)	-0.131*** (0.00155)	0.250*** (0.0180)	0.468*** (0.0816)	-0.131*** (0.00155)	0.250*** (0.0180)	0.443*** (0.0903)
Observations	1,051,212	1,051,212	840,977	840,977	33,226	834,854	834,854	27,144
R-squared	0.003	0.003	0.002	0.002	0.005	0.002	0.002	0.004
p-value F-test	0.0000	0.0000	0.0003	0.0002	0.0089	0.0055	0.0058	0.2069

Note: The policy dummy takes on a value of 1 when the test score and teacher assessment has been given in 2015/2016 or 2016/2017. The migrant dummy in columns A takes on a value of 1 if the child is a 2nd generation immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns B takes on a value of 1 if the child is a 1st generation Western immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is 1st generation non-Western immigrant and a value of 0 if the child is native Dutch. Controls include gender (male equals 1), age at arrival, and age when taking the school exit test. Year fixed effects include dummies for all the years (and thus also the Policy dummy), except for 2013 (i.e. baseline year/ year prior to the policy change). Standard errors are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The actual track assignment decision after the policy change is based on the revised teacher assessments. When a child obtains a higher track assignment as suggested by the test score than suggested by the teacher, the teacher can change his/her assessment upwards. To check how the policy affects the difference between these final teacher assessments and test scores, we run the DiD and Event study

specifications with the revised teacher assessments. Table 4.6 reports the DiD regression results, with negative and significant policy effects across all the migrant groups. Table 4.7 shows the results of the Event study. Again, the interaction term has a negative effect on the difference in revised teacher assessments and test scores for migrants, with significant results for second generation and first generation non-Western migrants. The results of the DiD and Event study with revised teacher assessment of the first generation Western and non-Western migrants are comparable in the population and subsample.

Table 4.6 Robustness check 2: DiD estimation with revised teacher assessments

	(A) 2nd generation		(B) 1st generation Western			(C) 1st generation non-Western		
	No controls	Controls — gender & age test	No controls	Controls — gender & age test	All controls	No controls	Controls — gender & age test	All controls
Migrant*Policy	-0.0634*** (0.00730)	-0.0642*** (0.00730)	-0.0852*** (0.0257)	-0.0867*** (0.0257)	-0.0783** (0.0344)	-0.0754** (0.0346)	-0.0775** (0.0346)	-0.0671 (0.0416)
Policy	0.0441*** (0.00350)	0.0416*** (0.00350)	0.0441*** (0.00348)	0.0414*** (0.00349)	0.0348 (0.0229)	0.0441*** (0.00348)	0.0414*** (0.00349)	0.0346 (0.0229)
Migrant	0.101*** (0.00441)	0.105*** (0.00442)	0.0785*** (0.0154)	0.0903*** (0.0154)	-0.0150 (0.0210)	0.130*** (0.0186)	0.146*** (0.0186)	0.0414* (0.0239)
Gender	No	Yes	No	Yes	Yes	No	Yes	Yes
Age taking test	No	Yes	No	Yes	Yes	No	Yes	Yes
Age at arrival	No	No	No	No	Yes	No	No	Yes
Constant	-0.155*** (0.00206)	0.228*** (0.0319)	-0.155*** (0.00205)	0.254*** (0.0363)	0.388** (0.161)	-0.155*** (0.00205)	0.252*** (0.0364)	0.309* (0.179)
Observations								
R-squared	256,490 0.003	256,490 0.003	202,537 0.001	202,537 0.002	8,251 0.004	201,110 0.001	201,110 0.002	6,832 0.002

Note: The policy dummy takes on a value of 1 when the test score and teacher assessment has been given in 2014/2015, 2015/2016 or 2016/2017. The migrant dummy in columns A takes on a value of 1 if the child is a 2nd generation immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns B takes on a value of 1 if the child is a 1st generation Western immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is 1st generation non-Western immigrant and a value of 0 if the child is native Dutch. Controls include gender (male equals 1), age at arrival, and age when taking the school exit test. Standard errors are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.7 Robustness check 2: Event study estimation with revised teacher assessments

	(A) 2nd generation		(B) 1st generation Western			(C) 1st generation non-Western		
	No controls	Controls — gender & age test	No controls	Controls — gender & age test	All controls	No controls	Controls — gender & age test	All controls
Migrant*Policy	-0.0217*** (0.00548)	-0.0222*** (0.00548)	-0.0286 (0.0200)	-0.0294 (0.0200)	-0.0557** (0.0257)	-0.0624** (0.0276)	-0.0626** (0.0276)	-0.0891*** (0.0319)
Migrant*2012	0.0302*** (0.00667)	0.0306*** (0.00667)	0.0552** (0.0242)	0.0565** (0.0242)	0.0660** (0.0311)	0.00844 (0.0326)	0.0102 (0.0326)	0.0194 (0.0380)
Migrant*2011	0.0130* (0.00677)	0.0134** (0.00677)	0.0500** (0.0242)	0.0507** (0.0242)	0.0334 (0.0314)	0.0227 (0.0318)	0.0233 (0.0318)	0.00330 (0.0375)
Migrant*2010	-0.0198** (0.00863)	-0.0193** (0.00863)	-0.0178 (0.0243)	-0.0168 (0.0243)	0.0178 (0.0328)	-0.102*** (0.0316)	-0.0990*** (0.0316)	-0.0642* (0.0384)
Migrant*2009	0.0142 (0.00923)	0.0153* (0.00922)	0.0408 (0.0252)	0.0436* (0.0252)	0.0596* (0.0346)	0.000403 (0.0329)	0.00337 (0.0329)	0.0173 (0.0404)
Migrant*2008	0.0349*** (0.00835)	0.0349*** (0.00835)	0.00233 (0.0233)	0.000625 (0.0233)	-0.0323 (0.0316)	0.0349 (0.0307)	0.0341 (0.0307)	-0.00286 (0.0373)
Migrant	0.0719*** (0.00466)	0.0748*** (0.00466)	0.0517*** (0.0171)	0.0622*** (0.0171)	-0.0300 (0.0220)	0.115*** (0.0232)	0.129*** (0.0232)	0.0411 (0.0270)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	Yes	No	Yes	Yes	No	Yes	Yes
Age taking test	No	Yes	No	Yes	Yes	No	Yes	Yes
Age at arrival	No	No	No	No	Yes	No	No	Yes
Constant	-0.109*** (0.00218)	0.280*** (0.0159)	-0.109*** (0.00217)	0.301*** (0.0179)	0.513*** (0.0808)	-0.109*** (0.00217)	0.300*** (0.0180)	0.479*** (0.0895)
Observations	1,050,839	1,050,839	841,058	841,058	33,808	834,737	834,737	27,528
R-squared	0.003	0.004	0.002	0.002	0.005	0.002	0.002	0.005
p-value F-test	0.0000	0.0000	0.0210	0.0158	0.0936	0.0094	0.0114	0.2500

Note: The policy dummy takes on a value of 1 when the test score and teacher assessment has been given in 2014/2015, 2015/2016 or 2016/2017. The migrant dummy in columns A takes on a value of 1 if the child is a 2nd generation immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns B takes on a value of 1 if the child is a 1st generation Western immigrant and a value of 0 if the child is native Dutch. The migrant dummy in columns C takes on a value of 1 if the child is 1st generation non-Western immigrant and a value of 0 if the child is native Dutch. Controls include gender (male equals 1), age at arrival, and age when taking the school exit test. Year fixed effects include dummies for all the years (and thus also the Policy dummy), except for 2013 (i.e. baseline year/year prior to the policy change). Standard errors are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.7 Discussion

This chapter investigates the effect of the inability to observe the school exit test score on the teacher assessments of migrants compared to natives. The identification strategy exploited a policy change which changed the timing

and nature of secondary school track placement. Before the policy change the secondary school track placement depended on both test scores and teacher assessment, but this changed to a track placement that only depended on teacher assessments after the policy implementation. Our main conclusion is that the policy change decreased the relative advantage of migrants compared to natives in terms of teacher assessments, as expressed by the negative interaction coefficient across all migrant groups and specifications. The idea that migrant students will benefit from a track assessment decision that puts more weight on the teacher assessment and less on the exit test score (which could be lower for migrants due to language barriers), which is the case after the policy change, is not supported by the data.

From a policy perspective, the findings suggest that the policy change led to relatively higher teacher assessments for natives. This might be due to the fact that the migrant children already expressed relatively high teacher assessments before the policy change, arguably making the policy change not as necessary for them as was initially thought. The negative effect of the policy change on the relative advantage of migrant children compared to natives with respect to teacher assessments could serve as an argument to return to the old policy, as is also the conclusion of Swart et al. (2019) and Onderwijsraad (2018). Furthermore, the Inspectie van het Onderwijs (2016, 2019) reports that the discrepancy in teacher assessments and test scores has increased in terms of the socio-economic status (SES) of children as well. It is plausible that the underlying mechanisms for migrant background and SES background are similar: students with a high SES background (who are most likely to be native) benefit relatively more from the policy change than students with a low SES background (who are most likely to have a migrant background). It seems that the group of students who were targeted by the policy change (e.g. migrants and low SES students) did not benefit from the policy change in terms of their teacher assessments relative to natives.

The findings confirm the hypothesis on teacher bias and statistical discrimination, i.e. a negative effect of the policy change on the teacher assessments of migrants. This is in line with previous results on the discrepancy between objective and subjective ability measures for students with a migration background (Burgess & Greaves, 2013; Dee, 2005). Furthermore, given the fact that the Netherlands uses an early tracking system and taking into account the negative consequences of early tracking for long run educational and human capital (e.g. Checchi & Flabbi, 2006; Hanushek & Woessmann, 2006; Brunello & Checchi, 2007; Bol & van de Werfhorst, 2013; Ludemann & Schwerdt, 2013; Hopwood et al., 2016), it would be interesting for future research to check whether the negative results could be mitigated by tracking at a later age. Another interesting way forward is to check whether the negative results of the policy change are reflected in worse long term outcomes, such as secondary school exam grades and tertiary education levels.

4.8 Appendix

Figure A4.1

Dutch education system

Note: This figure provides an overview of the Dutch education system. Special education and practical education are not included, since they are not considered to be part of the mainstream education.

Table A4.1 Test scores and track assessments

Test score	Secondary education track placement suggestion	Dutch abbreviation	ISCED level	Ordinal scale
500 – 519	Pre-vocational	Vmbo-b	2	1
519 – 522		Vmbo-b – Vmbo-k	2	2
523 – 528		Vmbo-k	2	3
529 – 532	Theoretical preparatory	Vmbo-g(t)	2	4
533 – 536	Theoretical preparatory – Secondary General	Vmbo-g(t) – Havo	2 – 3	5
537 – 539	Secondary General	Havo	3	6
540 – 544	Secondary General – Pre-University	Havo – Vwo	3 – 4	7
545 – 550	Pre-University	Vwo	4	8

Note: This table provides an overview of the Dutch secondary education tracks. For comparability reasons, the corresponding ISCED levels are displayed in column 2 and the Dutch abbreviations in column 3. The last column shows that practical education corresponds to the lowest track and pre-university to the highest track.

Figure A4.2

Timing test score and teacher assessment pre- and post-policy change

Chapter 5

The effect of specialized refugee schools on the primary school outcomes of refugee children †

† I thank Christian Dustmann, Bart Golsteyn, Inge de Wolf, Mark Kattenberg, several colleagues from the Centre for Research and Analysis of Migration of University College London and participants at the 2018 LEER Conference, the 2018 Transitions in Youth Conference, the 2018 Dutch Economists' Day and 2019 CPB internal seminar for their valuable feedback and suggestions.

5.1 Introduction

Education is one of the most important institutions that can help young immigrants perform at the same level as natives, integrate well and feel at home in the host country. Research has shown that refugees³¹ face more barriers to education than other types of immigrants (McBrien, 2005; Bloch et al., 2015; Suarez-Orozco et al., 2011) and that the educational setting of a country has important implications for the equality of opportunities in terms of educational achievement (Dupriez & Dumay, 2006). Understanding the factors that determine refugees' educational success is therefore a key issue for educational policy. The large inflow of refugees into Europe over the past few years has underlined the pressing nature of this issue. After arrival in the host country, refugee children deal with language problems, a lack of education in recent years, possible traumas, and they need to adjust to a new culture and country.

This chapter exploits the placement into two different school types as a natural experiment to investigate the effect of the school environment on primary school outcomes of refugee children. During the asylum process in the Netherlands, refugees are randomly placed into a housing facility and their children are enrolled at the nearest school. This school can either be a regular primary school or a specialized school for refugee and migrant children. The specialized schools are solely meant for children of newly arrived immigrants and provide a different educational environment than regular schools. At the refugee schools, the main focus is on learning the Dutch language, with personalized learning goals and teachers who have experience with traumatized children and teaching Dutch. At regular schools, immigrants follow the standard curriculum with some additional Dutch classes. Refugee schools can be seen as a parallel schooling system, since the refugees less easily get in contact with native students while they are enrolled at this facility. Therefore, the exposure to native peers and peers with the same background differs greatly between the two types of schools.³²

In the empirical analysis, I confirm that refugees are indeed placed randomly into regular or refugee schools. This random placement makes it possible to estimate the effects of being placed in a regular or refugee school on test scores and teacher assessments at the end of primary school. The main results of the analyses are that there are no significant causal effects of being placed at a refugee school on test scores and teacher assessments relative to being placed at a regular school.

31 Refugees are immigrants with an asylum migration motive; they have fled their home country due to fear of persecution, war or violence. The refugee sample used in this chapter consists of first generation immigrant children who are born in Iran, Iraq, Somalia, Afghanistan, Syria or Eritrea (see section 5.2 for more details).

32 Section 5.2 provides a detailed description of these differences.

Research on outcomes of refugees is scarce compared to the vast literature on immigrants' outcomes. This lack of literature on refugees' outcomes is mainly due to the difficulty to identify refugees from immigrants, a lack of long term data, and the small sample sizes of refugees in surveys when data are available (Evans & Fitzgerald, 2017). Dustmann et al. (2017) find that refugees in Europe were less likely to have a job compared to economic migrants from the same country of origin and with the same level of education. In the US, refugees who entered at ages 18-45 and were in the US for more than six years, worked more often than natives but never reached the same levels of employment as the natives (Evans & Fitzgerald, 2017). Evans and Fitzgerald (2017) also show that the high school graduation and college entry rates are similar for refugees and natives if refugees were younger than 14 upon arrival. The refugees who were older upon arrival perform worse, mainly due to language barriers. Gerritsen et al. (2019) study the effect of age at arrival on the educational attainment of refugees who arrived in the Netherlands between 1995-1999. They find that if refugees arrive one year earlier, their chance of obtaining a higher education degree increases by 3.6 percentage points, with stronger effects for girls. For this same arrival cohort, Maliepaard et al. (2017) find that refugees do not reach the same educational attainment as natives but perform better compared to other immigrants. Their findings on the employment of refugees are in line with the US finding of Evans and Fitzgerald (2017); namely that refugees catch up over time, but never reach the same level of employment as natives.

One of the main identification problems in previous research on immigrants' and refugees' outcomes (e.g. wages, unemployment benefits) is selection into neighborhoods or schools. The choice to settle in a certain geographical area is most likely influenced by unobservable characteristics, which in turn affect e.g. education and labor market outcomes. One way to address this identification issue is by exploiting a natural experiment,³³ such as the refugee dispersal policies in Sweden (Edin et al., 2003; Åslund & Rooth, 2007; Åslund & Fredriksson, 2009; Åslund et al., 2011), Denmark (Damm, 2009 and 2014; Damm & Dustmann, 2014), and Israel (Gould et al., 2004). In Sweden, lower skilled immigrants have improved labor market outcomes when they live in a bigger enclave, and high-income ethnic groups benefit more from living in an enclave (Edin et al., 2003). The higher the unemployment rates and welfare dependence immigrants are exposed to after arrival, the higher the chance that these immigrants will be unemployed and welfare dependent later in life, with long term consequences (Åslund & Rooth,

33 Within the strand of literature on immigrants' outcomes, there are also numerous papers that use instrumental variables, such as a larger geographical area (e.g. Jensen & Thomsen, 2013), historical settlements (e.g. Altonji & Card, 1991; Barone et al., 2016; Halla et al., 2017), and availability of living space (Steinmayr, 2016).

2007; Åslund & Frederiksson, 2009). The GPA and educational attainment are increasing in the number of highly educated adults of the same ethnicity (Åslund et al., 2011). An increase in the skill level of men with the same nationality in the neighborhood and an increase in the enclave size, raises the annual earnings of immigrants in Denmark (Damm, 2009 & 2014). Damm & Dustmann (2014) find that male immigrants in Denmark are more likely to be convicted when they lived in neighborhoods with a higher share of young people convicted for (violent) crimes. In Israel, higher pre-immigration math scores reduce dropout rates and increase passing rates on high school math exams, with no effects for pre-immigration verbal scores (Gould et al., 2004).

This chapter contributes to the literature by using unique administrative data, by exploiting random placement in schools and specialized refugee schools, thus reporting causal effects, and by focusing on a relatively new and underreported cohort. The results are interesting for policy makers, because they have to choose whether to invest in specialized refugee schools.

The remainder of this chapter is organized as follows. The next section provides further details on the Dutch education context and immigration school placement. Section 5.3 describes the data used in this chapter. Section 5.4 outlines the methodology and discusses identification issues. Section 5.5 discusses the effect of attending a specialized refugee school on the test scores and teacher assessments of refugee children. Section 5.6 summarizes, concludes and discusses the results.

5.2 The Dutch educational context and placement policy

Compulsory school age in the Netherlands ranges from ages 5 to 16³⁴, but usually students enter the first grade when they are 4 years old. Primary school consists of 8 grades. Students continue to secondary education when they are about 12 years old.³⁵ From this point onwards, students enter a tracked system, which means that the Netherlands has a relatively early³⁶ tracking system. Track placement is determined by two signals about the ability of the child: a test score and an assessment from the teacher about which track would be the best fit. The teacher has seen the students throughout a whole year and knows e.g. the learning attitude and learning curve of the students. Furthermore, the teacher knows the student's results from previous years. Based on this information, the teacher can then make

34 When students do not have a secondary vocational, a secondary general or a pre-university degree before the age of 16, they have to stay in school until the age of 18.

35 In the Netherlands, some students repeat a grade. They are about a year older when they continue to secondary education.

36 Only Austria, Germany, Hungary, and the Slovak Republic track at an even earlier age (10 years).

an informed decision about the secondary education level that would fit best. The combination of the test score and teacher track recommendation determines the secondary education track placement.³⁷ There are three main secondary education levels: pre-university education (6 years), secondary general education (5 years), and theoretical and vocational preparatory education (4 years). The secondary education track placement determines the tertiary education enrollment options level and can have long-term consequences (Checchi & Flabbi, 2006; Bol & van de Werfhorst, 2013; Hopwood et al., 2016; Swart et al., 2018; Feron, 2018). Figure A4.1 displays an overview of the Dutch education system.

If immigrants are between 4 and 12 years old upon arrival, they enroll in primary school. The Netherlands has two different types of primary schools: regular schools and specialized refugee schools (which are named ‘newcomer facilities’). The educational settings between these two primary school types are different with respect to multiple aspects. Specialized refugee schools receive additional funding from the government to enable smaller classes and hire teachers who are trained to teach Dutch as a new language, and who know how to recognize and help children with traumas. From a didactical point of view, the refugee schools align the teaching with the needs of the immigrants; they use different teaching methods, put more emphasis on the social skills and socio-emotional development of the immigrant, and the main focus is on learning the Dutch language. Regular schools, on the other hand, only provide some additional help with learning the Dutch language; most of the time the immigrant follows the normal curriculum. Since refugee schools are designed exclusively for newly arrived immigrants, the ethnic composition of the classes of these facilities differs from the ethnic composition of classes at regular schools. An immigrant will not have any native peers at a refugee school and more peers with the same ethnic and cultural background as the immigrant. It is also more likely that immigrants can still talk their mother tongue with peers at a refugee school, whilst they will hear more Dutch at a regular school.

Immigrants have different motives to migrate to the Netherlands, such as work, study, family reunion, or asylum. Only those with an asylum motive (i.e. asylum seekers)³⁸ have the right to receive housing during the asylum process. The

37 Chapter 3 of this thesis discusses a policy change that altered the timing of teacher assessments and test scores from 2014-2015 onwards. The analysis in this Chapter uses data from years prior and post policy implementation and uses a policy dummy as a control variable.

38 An asylum seeker is an immigrant who claims to be a refugee and asks for a permit on this ground. Only when the claim has been evaluated and accepted, an asylum seeker is officially called a refugee. Since this chapter studies outcomes of asylum seekers who have been in the Netherlands for a longer period of time, these asylum seekers have received a permit and thus can be defined as refugees. Hence, this chapter uses the term refugees when discussing the results.

housing of asylum seekers is organized by the Central Agency for the Reception of Asylum Seekers (COA).³⁹ This agency randomly places asylum seekers into available housing locations. Individual wishes and preferences of asylum seekers are not taken into account during this allocation process.⁴⁰ After being placed in a COA housing facility, children of asylum seekers are enrolled at the nearest school. Children of primary school going age are either placed at a regular school or at a refugee school. Refugee children enroll at refugee schools for two years at most, after which they have to switch to a regular school due to financial reasons.⁴¹ In the Netherlands, everyone is free to choose at which school they want to enroll their child. However, conversations with teachers and inspectors who come into contact with parents from asylum seekers' children, point out that these parents are rarely aware of this possibility. Even when they know about it, they often do not have the means to find out which alternatives are better options than the current school. Therefore, I assume that children of asylum seekers who are randomly placed by COA are enrolled at the nearest school and do not switch schools within the first year after arrival. The results of the balance test in Section 5.4 support this assumption.

If immigrants have a different migration motive than seeking asylum, it is plausible that they already have a network or some family in the Netherlands that can support and host them. In that case, the immigrant is not placed by COA and selection into neighborhoods (and subsequently schools) may be a problem. To ensure that I only select immigrants who were randomly placed by COA, I turn to the data of COA on the type of immigrants they housed. Table A5.1 reports the top 5 countries of origin per arrival year that COA housed. Based on this information, I select immigrants from Iran, Iraq, Somalia, Afghanistan, Syria and Eritrea and define this draw from the immigrant population the refugee sample.

5.3 Data

The data used in this chapter are Dutch administrative data containing educational outcomes, a variety of demographic variables and school characteristics for all children who were enrolled in Dutch primary schools for at least one year during

³⁹ COA is the Dutch abbreviation for 'Centraal Orgaan opvang Asielzoekers'.

⁴⁰ Once asylum seekers have received a temporary permit, they will move from COA housing to private housing provided by municipalities. Municipalities have to provide a certain amount of housing facilities each year, based on the current population. This is in line with dispersal policies in Sweden (Edin et al., 2003) and Denmark (Damm, 2009). The allocation to municipality housing can depend on first degree family, work, study, and health issues.

⁴¹ Schools only receive two years of additional funding for each newly arrived immigrant. It is expected that an immigrant child is able to participate in the normal curriculum after this time.

2008 – 2016. The data are gathered on a yearly basis. This chapter analyzes the sample of 1,580 refugees who entered the Netherlands from 2007 onwards,⁴² who were of primary school-going age upon arrival (3-12 years old),⁴³ who did not leave the Netherlands, and for whom both a test score and teacher track assessment is available. 217 of these 1,580 refugees were enrolled at a refugee school at the time of first observation.

Two important assessments in the last grade of Dutch primary education are test scores from a nationwide test and a teacher track assessment. For comparability reasons, I only consider the test scores from the Cito test. This test is the most commonly used test at the end of primary school in the Netherlands; about 84% (78%) of all the students (refugees) who took a test, participated in the Cito test in the observation period of 2008-2016.⁴⁴ The test assesses math and language skills, with anchored items that result in a score from 500-550. The calculation from these anchored items to a score on the 500-550 scale is the same for every year of the observational period of this chapter, hence the test scores are comparable. Besides taking this test, students receive an assessment from their teacher about the secondary education level that fits them best. The assessment by the teacher is an ordinal variable on a 10-point scale,⁴⁵ as displayed in Table A5.2.⁴⁶

I control for variables of refugees that may have an effect on educational outcomes, where I include the refugee's gender, the age upon arrival, the region of origin, and the arrival year. I also include an indicator variable for a Dutch policy change. A policy change occurred in 2015 and dictated that the timing of test scores and teacher assessments changed, resulting in teachers having less information when they had to create an assessment (see also Chapter 4 of this thesis). Feron (2018) points out that due to this policy change the test becomes

42 The first observation is in 2008, so arrivals before 2007 most likely will not give trustworthy data on the first placement since refugees do not stay in COA housing locations for such a long time.

43 I include three year olds as well, since they can be four by the first time they are observed.

44 Before the policy change of September 1st, 2014, that changed the timing of taking the test and receiving the teacher assessment (see next paragraph for more information), the Cito test was the only available test. After the policy change, new tests were available. The content and scaling of the tests are not comparable with one another.

45 Note that this Chapter uses a 10-point ordinal scale for the teacher assessments and Chapter 4 an 8-point ordinal scale. Practical education and the assessment vmbo-k/vmbo-g(t)' are not included in the rescaling of test scores to secondary education track assessments. Since I do not compare test scores and teacher assessments in this Chapter, I include these two assessments levels, resulting in a 10-point scale.

46 In the remainder of this chapter I treat the teacher assessment variable as linear, since the OLS regression results are similar to the results of the ordered logistic regressions.

low-stake for some students and the teachers base their assessments on more subjective measures such as gender and SES. In order to avoid measuring this policy change effect instead of the desired placement effect, I include a dummy for the policy change in the regressions.

5.4 Methodology

This chapter analyzes the effect of initial school type placement on primary school outcomes of refugees. This can be captured by a linear relation:

$$Y_i = \beta_0 + \beta_1 Treatment_i + \beta_2 X_i + \varepsilon_i$$

where $Treatment_i$ is a dummy which takes on the value of 1 if the refugee was placed at a refugee school and 0 if the refugee was placed at a regular school, X_i is a vector of control variables including dummies for country of origin, gender, arrival cohort, age at arrival and policy change, and ε_i is an error term. Y_i either stands for the results of the test scores or the teacher assessments.

The estimation of β_1 using OLS may give a biased estimate of β_1 if school placement ($Treatment_i$) correlates with unobserved determinants of education outcomes (ε_i). However, if refugees are randomly placed by COA, and enroll at either the nearest refugee school (treatment) or regular school (control), there is no correlation between school placement and the potential confounders in the error term in a simple OLS regression.

The last column of Table 5.1 reports the differences between the mean values of the control variables for treatment and control group and the corresponding p-values of the t-test. The only difference at the 1% significance level is for refugees who arrived in 2012; a refugee who arrived in 2012 was more likely to be enrolled at a refugee school. The last line of Table 5.1 reports the p-value of the F-test that all the differences are equal to 0 at the same time. With a p-value of 0.1959 the hypothesis that there are no differences cannot be rejected. In line with Altonji et al. (2005), I assume that there is no relationship between $Treatment_i$ and unobserved characteristics (ε_i) given the fact that there is no relationship between $Treatment_i$ and the observed characteristics (X_i).

Table 5.1 Balance table refugees

	Variable	Control group	Treatment group	Difference
Country of origin	Iran	0.079	0.055	-0.023
		(0.269)	(0.229)	(0.019)
	Iraq	0.312	0.346	0.034
		(0.463)	(0.477)	(0.034)
	Somalia	0.415	0.382	-0.033
		(0.493)	(0.487)	(0.036)
	Afghanistan	0.089	0.083	-0.006
		(0.285)	(0.276)	(0.021)
	Syria	0.070	0.083	0.013
		(0.256)	(0.276)	(0.019)
	Eritrea	0.035	0.051	0.015
		(0.184)	(0.220)	(0.014)
Gender	Male	0.507	0.535	0.028
		(0.500)	(0.500)	(0.037)
Arrival cohort	2007	0.104	0.065	-0.040*
		(0.306)	(0.246)	(0.022)
	2008	0.167	0.175	0.008
		(0.373)	(0.381)	(0.027)
	2009	0.219	0.161	-0.057*
		(0.413)	(0.369)	(0.030)
	2010	0.190	0.189	-0.001
		(0.392)	(0.392)	(0.029)
	2011	0.152	0.175	0.023
		(0.359)	(0.381)	(0.026)
	2012	0.039	0.078	0.039***
		(0.193)	(0.269)	(0.015)
	2013	0.067	0.069	0.002
		(0.250)	(0.254)	(0.018)
	2014	0.042	0.060	0.018
		(0.200)	(0.238)	(0.015)
	2015	0.019	0.023	0.004
		(0.137)	(0.150)	(0.010)
	2016	0.001	0.005	0.003
		(0.038)	(0.068)	(0.003)
Age at arrival	3	0.026	0.023	-0.003
		(0.160)	(0.150)	(0.012)
	4	0.051	0.023	-0.028*
		(0.219)	(0.150)	(0.015)
	5	0.088	0.097	0.009
		(0.283)	(0.296)	(0.021)
	6	0.123	0.115	-0.008
		(0.329)	(0.320)	(0.024)

7	0.160	0.189	0.029
	(0.367)	(0.392)	(0.027)
8	0.168	0.166	-0.002
	(0.374)	(0.373)	(0.027)
9	0.159	0.138	-0.021
	(0.366)	(0.346)	(0.027)
10	0.137	0.124	-0.013
	(0.344)	(0.331)	(0.025)
11	0.068	0.111	0.042**
	(0.252)	(0.314)	(0.019)
12	0.018	0.014	-0.005
	(0.134)	(0.117)	(0.010)
p-value F-test: no difference			0.1959
Observations	1,363	217	1,580

*Note: This table reports the mean values of the several categories of control variables for immigrants who enrolled at a regular school immediately (control, column 1) and those who enrolled at a refugee school (treatment, column 2). The third column reports the difference between the mean values of treatment and control. The standard errors are reported in brackets. Significance levels are indicated with asterisks: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The last line reports p-value of the F-test that all the differences are equal to 0 at the same time.*

5.5 Results

The first two columns of Table 5.1 show the descriptive statistics for the control variables, separately for refugees who were placed at a refugee school and for those who were placed at a regular primary school. The sample includes 1,580 refugee children: 1,363 of whom enrolled at a regular primary school and 217 at specialized refugee schools. Regarding the country of origin of the refugees; about 7% originate from Iran, 33% from Iraq, 40% from Somalia, 8.5% from Afghanistan, 7.5 from Syria, and 4% from Eritrea. This means that more than 70% of the refugees were born in either Iraq or Somalia. About half of the refugee sample consists of boys.

Since I only investigate refugees who have finished primary school, the older arrival cohorts are better represented in this chapter. The age at arrival categories 6 – 9 years old make up more than 50% of the refugee sample of this chapter.

Figure 5.1 displays boxplots of the test scores of refugee children separately for each country of origin group and for refugees enrolled at a regular primary school and specialized refugee school separately. The mean test score (represented by the line in the box) is highest for Iranian refugees (around 534) and lowest for Somalian refugees (around 527) for refugees who were enrolled at a regular

primary school. For refugees who were enrolled at a specialized refugee school, refugees from Iraq have the highest mean test score (around 531) and refugees from Iran and Afghanistan score the lowest (around 525). Figure 5.2 reports the percentage of students per assessment level separately for each country of origin and for refugees enrolled at regular primary schools (control) and specialized refugee schools (treatment) separately. Refugees from Iran and Afghanistan receive higher teacher assessments when they were enrolled at regular primary schools. For refugees from Iraq, this is the other way around: they receive higher teacher assessments when they were enrolled at specialized refugee schools. Somalian refugees overall do not differ in their teacher assessments conditional on treatment, but Somalian refugees who were enrolled at a regular primary school are more likely to receive pre-university assessments. For refugees from Syria and Eritrea, there is no clear difference between those who were enrolled at a regular primary school and those who were enrolled at a refugee school.

Figure 5.1
Boxplots of test score per country of origin of refugee children

Note: This Figure shows the boxplots of the test scores of refugee students for each country of origin. The upper bar of the box is the 75th percentile, the lower bar of the box the 25th, the line in the box is the median test score. The vertical line below the box displays the lowest 25% of the scores, the vertical line above the box displays the highest 25% of the test scores. The test scores are the scores from the Cito test and are given on a 500-550 scale. The boxplots on the left show the test scores of refugees enrolled at a regular primary schools, the boxplots on the right show this for refugees enrolled at a specialized refugee school.

Figure 5.2
Division of teacher assessments per country of origin of refugee children

Note: This Figure displays the percentages of refugee children per assessment level for each country of origin. The assessment levels range from lowest (practical education) to highest (pre-university). An overview of the assessment levels and how they compare to international standards is given in Table A5.2.

Table 5.2 shows the OLS regression results of the treatment (i.e. assignment to a refugee school) on the test scores of refugees. Every column includes a different combination of control variables. The main finding is that placement in a refugee school is not significantly related to the test score. There is no significant effect of being placed in a specialized refugee school. The point estimates are also small (less than 1 point on the 500-550 Cito scale), indicating that the insignificance of the results does not appear to be driven by a lack of statistical power. The point estimate indicates that being treated reduces the test score of the Cito test by merely 0.4 – 0.6 points on a 50-point scale. The coefficients remain similar and insignificant across all other specifications.

Table 5.3 reports the OLS regression results of the treatment on teacher assessments of refugee children. Again, there is no significant effect of being placed at a specialized refugee school on the teacher assessment of the refugee children. The results show again that the treatment variable coefficients are small (less than 0.01 on a 10-point ordinal teacher assessment scale) and insignificant across the specifications.

Table 5.2 OLS regressions of enrollment at a refugee school on test scores of refugees

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.575 (0.816)	-0.625 (0.801)	-0.401 (0.817)	-0.539 (0.808)	-0.573 (0.816)	-0.507 (0.790)	-0.625 (0.790)
Ethnicity	NO	YES	NO	NO	NO	YES	YES
Arrival cohort	NO	NO	YES	NO	NO	YES	NO
Age at arrival	NO	NO	NO	YES	NO	YES	YES
Gender	NO	NO	NO	NO	YES	YES	YES
Policy change	NO	NO	NO	NO	NO	NO	YES
Constant	528.0*** (0.302)	531.9*** (1.006)	530.4*** (0.893)	527.9*** (1.724)	528.1*** (0.416)	533.0*** (2.074)	532.5*** (1.960)
Observations	1,580	1,580	1,580	1,579	1,580	1,579	1,579
R-squared	0.000	0.043	0.014	0.031	0.000	0.095	0.088

Note: This table reports the results for the Ordinary Least Squares regressions (OLS) of the treatment dummy on the test scores of refugees. The test scores are reported on a scale of 500-550. Column 1 has no controls included. Column 2 includes a dummy for country of origin: Iran, Iraq, Somalia, Afghanistan, Syria and Eritrea. Column 3 includes a dummy for the arrival year of the immigrant. Column 4 includes a dummy for the age at arrival category of the immigrant. Column 5 includes a dummy whether the child is male. Column 6 includes all of the above. Column 7 replaces the arrival cohort dummy with a dummy whether an immigrant took the test before the policy change in 2014/2015. Significance levels are indicated with asterisks: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5.3 OLS regressions of enrollment at a refugee school on teacher assessment of refugees

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.002 (0.187)	-0.003 (0.182)	-0.005 (0.187)	-0.006 (0.187)	0.0003 (0.187)	-0.017 (0.182)	0.003 (0.182)
Ethnicity	NO	YES	NO	NO	NO	YES	YES
Arrival cohort	NO	NO	YES	NO	NO	YES	NO
Age at arrival	NO	NO	NO	YES	NO	YES	YES
Gender	NO	NO	NO	NO	YES	YES	YES
Policy change	NO	NO	NO	NO	NO	NO	YES
Constant	5.325*** (0.0693)	6.513*** (0.229)	5.789*** (0.205)	4.903*** (0.399)	5.290*** (0.0953)	6.254*** (0.477)	6.128*** (0.452)
Observations	1,580	1,580	1,580	1,579	1,580	1,579	1,579
R-squared	0.000	0.057	0.014	0.010	0.000	0.086	0.078

Note: This table reports the results for the Ordinary Least Squares regressions (OLS) of the treatment dummy on the assessments by the teacher of refugees. The teacher assessments are reported on an ordinal scale of 1-10. Column 1 has no controls included. Column 2 includes a dummy for country of origin: Iran, Iraq, Somalia, Afghanistan, Syria and Eritrea. Column 3 includes a dummy for the arrival year of the immigrant. Column 4 includes a dummy for the age at arrival category of the immigrant. Column 5 includes a dummy whether the child is male. Column 6 includes all of the above. Column 7 replaces the arrival cohort dummy with a dummy whether an immigrant took the test before the policy change in 2014/2015. Significance levels are indicated with asterisks: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A final question is whether these findings are driven by selective attrition. This is the case if the refugees who are not observed⁴⁷ are exposed to a different school environment or have different individual characteristics than observed refugees, and this different school environment or these different individual characteristics influence the educational outcomes of refugees. I investigate whether the probability of being observed (i.e. having a test score and teacher assessment) depends on being treated and on several background characteristics, such as country of origin, gender, arrival cohort and age at arrival. I estimate the following model:

$$Observed_i = \alpha_o + \alpha_1 Treatment_i + \alpha_2 Z_i + \varphi_i$$

where $Observed_i$ is a dummy that equals one if both test score and teacher assessment are known for the refugee, $Treatment_i$ is a dummy that equals one if the refugee was placed at a specialized refugee school, Z_i a vector of control variables including treatment, country of origin, gender, arrival cohort and age at arrival, and φ_i a disturbance term. Table A5.3 displays the results of this analysis. The last column shows that there is no selective attrition, conditional on the control variables, since the treatment coefficient is insignificant.

5.6 Conclusion and discussion

The main contribution of this chapter is to provide causal evidence on the effect of type of school on primary school outcomes of refugee children by exploiting random school placement in the Netherlands and using unique administrative data. Due to the Dutch immigrant allocation policy and subsequent school enrollment, refugees are randomly placed in two primary school types. They are either enrolled at a regular primary school or a specialized school, solely for newly arrived immigrants and refugees. The results are clear: being placed at a refugee school does not significantly affect test scores and teacher assessments of refugee children.

The results should be interpreted carefully. The treatment in this study is the school environment, which is a multidimensional concept that might include factors such as class size, peer quality, teacher quality, didactic methods, exposure to the Dutch language, and school funding. Unfortunately, I do not have data on these variables, hence these direct mechanisms cannot be tested separately. It is

⁴⁷ Refugees who left the Netherlands, or for whom only either a test score or a teacher assessment is available.

possible that some of these mechanisms cancel each other out. Previous research sheds some light on the direction of these mechanisms.

First of all, refugee schools have smaller class sizes. Short run negative effects of large class sizes in primary schools are reported by Angrist and Lavy (1999), Krueger (1999) and Urquiola (2006), but Hoxby (2000) finds no impact. Fredriksson et al. (2013) investigate whether primary school class size has a long term impact, reporting that a larger class size results in higher test scores at ages 13 and 16, and even results in higher wages at ages 27 to 42. Other studies find positive long term effects of smaller class sizes on outcomes such as taking exams in college (Krueger & Whitmore, 2001) and earnings (Chetty et al., 2011). Based on these results, the smaller class size of the specialized refugee schools is likely to work as a positive mechanism for the primary school outcomes of refugees.

Second, the class composition of refugee schools includes a higher proportion of immigrant students compared to regular primary schools. Although Hoxby (2000) finds no significant effects of class composition in terms of ethnic background on test scores, other research (Jensen & Rasmussen, 2011; Guryan, 2001) has pointed out that a high immigrant concentration is not beneficial for the educational outcomes of immigrants. The peer quality mechanism is therefore either negative for a higher immigrant ratio or non-existent.

Third, there are no data available on the quality of the teachers who are employed at specialized refugee schools. Previous research has shown that higher teacher quality is beneficial for the educational outcomes of the students (e.g. Van der Steeg & Gerritsen, 2016). Teachers are required to have a qualification in teaching Dutch as a new language (NT2) when they work at a refugee school. Additionally, teaching at a refugee school is generally viewed as more challenging, and teachers who decide to teach at such a facility plausibly are more motivated and dedicated. Data on whether the quality and background characteristics of teachers at regular schools and refugee schools differ is unavailable. Therefore, I do not know whether there is a positive or negative selection with respect to teacher quality into teaching at a refugee school. The Onderwijsraad (2017) points out that in the past not enough teachers had the right competences to teach Dutch (as a new language). Another challenge while teaching immigrant and refugee children, is to recognize and properly deal with traumas. This competence is not taught in a standard module in the teacher training, but rather is a competence that teachers acquire over time. Due to the lack of data on the quality of teachers, it is not possible to hypothesize the sign of the teacher quality mechanism.

Fourth, most of the classes at a refugee school are aimed at learning the Dutch language, compared to a regular curriculum with some extra Dutch classes at a regular school. The literature review by Collier and Thomas (2009) reports that using the mother language benefits the learning process of learning a new language. Hajer and Pulles (2017) describe that it is important to include the

mother tongue of refugees while teaching Dutch to maintain part of the refugee's identity and to use the same neural networks that the child already has in place from learning the mother tongue. They also point out that the mother tongue can be used for other subjects while learning Dutch in order to prevent stagnation of the intellectual development. On the one hand, it is easier for refugees to speak their mother tongue with other refugees at refugee schools. On the other hand, at regular schools they follow different subjects besides Dutch classes. Yet, it remains unclear which of these didactic methods works best.

The combination of these mechanisms might result in the finding that there is no effect of enrollment at a specialized refugee school on test scores and teacher assessments of refugee children. However, educational outcomes are just one of the many possible outcomes that can be studied. Future research could investigate the effect of refugee schools on other outcomes such as health and happiness. Furthermore, more research is needed on the effect of refugee schools on the total population of first generation immigrants. Although there is no significant treatment effect for the refugee sample in this chapter, there might be an effect for other migrant groups.

5.7 Appendix

Table A5.1 Top 5 countries of origin of refugees housed by COA per arrival year

Arrival year	Nr. 1	Nr. 2	Nr. 3	Nr. 4	Nr. 5
2016	Syria	Eritrea	Iraq	Iran	Afghanistan
2015	Syria	Eritrea	Iraq	Ethiopia	Afghanistan
2014	Syria	Eritrea	Somalia	Iraq	Afghanistan
2013	Somalia	Syria	Iraq	Afghanistan	Eritrea
2012	Iraq	Somalia	Afghanistan	Iran	Russia
2011	Afghanistan	Iraq	Somalia	n.a.	n.a.
2010	Afghanistan	Iraq	Somalia	Iran	Russia
2009	Somalia	Iraq	Afghanistan	Iran	Soviet Union
2008	Iraq	Somalia	Afghanistan	Iran	Soviet Union
2007	Iraq	Somalia	Afghanistan	Guinee	Iran

Note: The Central Agency for the Reception of Asylum Seekers (COA) yearly reports the top 5 countries of origin of the inflow of asylum seekers that they house. In 2011, only the top 3 countries of origin are reported.

Table A5.2 Teacher track assessments

Secondary education track	ISCED level	Dutch abbreviation	Ordinal scale
Practical education	2	Praktijkonderwijs	1
Pre-vocational Preparatory	2	Vmbo-k	2
	2	Vmbo-k — Vmbo-b	3
	2	Vmbo-b	4
Pre-vocational — Theoretical preparatory	2	Vmbo-b — Vmbo-g(t)	5
Theoretical preparatory	2	Vmbo-g(t)	6
Theoretical preparatory — Secondary General	2 — 3	Vmbo-g(t) — Havo	7
Secondary General	3	Havo	8
Secondary General — Pre-University	3 — 4	Havo — Vwo	9
Pre-University	4	Vwo	10

Note: this table provides an overview of the Dutch secondary education tracks. For comparability reasons, the corresponding ISCED levels are displayed in column 2 and the Dutch abbreviations in column 3. The last column shows that practical education corresponds to the lowest track and pre-university to the highest track.

Table A5.3 Test on non-random sample attrition

Dependent variable: Observed						
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.0462**	-0.0378*	-0.0206	-0.0315	-0.0466**	-0.0186
	(0.0218)	(0.0211)	(0.0202)	(0.0206)	(0.0218)	(0.0198)
Ethnicity	NO	YES	NO	NO	NO	YES
Arrival cohort	NO	NO	YES	NO	NO	YES
Age at arrival	NO	NO	NO	YES	NO	YES
Gender	NO	NO	NO	NO	YES	YES
Constant	0.409***	0.541***	0.543***	0.674***	0.394***	0.766***
	(0.00849)	(0.0320)	(0.0272)	(0.0581)	(0.0114)	(0.0679)
Observations	3,935	3,935	3,935	3,932	3,935	3,932
R-squared	0.001	0.063	0.117	0.146	0.002	0.185

Note: This table reports the results for the Ordinary Least Squares regressions (OLS) of the treatment dummy on the dummy of being observed in the data. A refugee is observed when both the test score and teacher assessment is known. Column 1 has no controls included. Column 2 includes a dummy for country of origin included: Iran, Iraq, Somalia, Afghanistan, Syria and Eritrea. Column 3 includes a dummy for the arrival year of the immigrant. Column 4 includes a dummy for the age at arrival category of the immigrant. Column 5 includes a dummy whether the child is male. Column 6 includes all of the above.

Chapter 6

Conclusions

In this thesis the phrase *playing the hand you're dealt* has served as inspiration. Where and when one is born determine the cards that are dealt. This thesis has investigated two cards that are dealt to children during their childhood and the effects of these cards on their personality and educational outcomes. The first part of this thesis has analyzed the question: *How does family structure affect children's personality traits?* The second part has looked at the question: *How does the educational setting in the Netherlands affect educational outcomes of migrant and refugee children?* The main conclusions of these two questions will be addressed for each part separately.

Part I

The first part of this thesis has investigated the effect of family structure on children's personality traits. The main question of this part is: *How does family structure affects personality traits?* Personality traits have been shown to be of vital importance for later in life outcomes (e.g. Borghans et al., 2008). In order to study the effect of family structure on personality traits, I have used the British Cohort Study. This study reports longitudinal data on personality traits and various demographic variables from a large British cohort. The personality traits included in this thesis are a measure for behavioral problems, self-esteem, locus of control, disorganized behavior, asocial behavior, neuroticism, and introversion. I have investigated two elements of family structure, namely the birth gap and sibling gender, and their effect on personality traits. These two elements of family structure are two types of cards that are dealt to children and over which they have no influence.

Chapter 2 examined the causal effect of the *birth gap* (i.e. the age difference) between siblings on personality traits of children. When the birth gap is small, siblings will compete over the same kind of resources from their parents at the same time. However, it might also be easier to play with and learn from each other when the birth gap is small. In this case, parents can provide the same type of resources simultaneously to several children, which will save them time and effort. Due to these contrasting mechanisms, it is hard to predict the effect of the birth gap on personality traits. The identification strategy employs miscarriages between the first and second child as an instrument for the birth gap. The main results show that an increase in the birth gap by one year increases the disorganized (0.11 s.d.) and neurotic behavior (0.15 s.d.) of the youngest child in two-child households.

Separate analyses for boys and girls show that the increased neurotic behavior due to a larger birth gap remains significant for both genders (0.12 s.d. for boys and 0.22 s.d. for girls). Additionally, boys become more disorganized with a larger birth gap, while girls become more anti-social and introverted. The negative effects of a larger birth gap on personality suggest a negative effect on outcomes related

to personality, such as education. However, the results of a larger birth gap on a variety of achievement tests and educational outcomes are inconclusive.

Chapter 3 investigated the effect of *sibling gender* on personality traits of children. Boys and girls have been shown to display stereotypical behavior which in turn can affect personality traits, but also longer term outcomes such as education, wages and family formation (e.g. Brenoe, 2018; Butcher & Case, 1994; Hauser & Kuo, 1998; Kaestner, 1997; Peter et al., 2018). The identification strategy exploits the randomness of the gender of a child. Additionally, possible endogeneity issues with respect to the number of children in a household and the gender (combination) of the existing child(ren) are avoided by running the analyses separately for low and high SES groups, since SES is arguably an important predictor of the number of children in a household. The main results show that for both oldest and youngest children in a two-child household, and for both boys and girls, having a sister results in more positive personality traits. Girls with a sister are less disorganized (0.29 - 0.32 s.d.) and anti-social (0.20 - 0.23 s.d.) than girls with a brother, while boys with a sister have fewer behavioral problems (0.10 - 0.15 s.d.) than boys with a brother. Having a sister is also reflected in more positive longer term life outcomes.

The results show that oldest boys have fewer behavioral problems at ages 5, 10 and 16 (0.11 - 0.15 s.d.) and are less disorganized (0.13 s.d.) when they have a younger sister instead of a brother. Younger boys also have fewer behavioral problems at ages 5 and 10 (0.11 s.d.) when they have a sister, but they also have lower self-esteem (0.20 s.d.). Oldest girls have fewer behavioral problems (0.13 - 0.15 s.d.), are less disorganized (0.32 s.d.), and show less asocial (0.20 s.d.) and introverted (0.11 s.d.) behavior, but also a lower self-esteem (0.15 s.d.) when they have a sister. Youngest girls have fewer behavioral problems (0.10 s.d.), are less disorganized (0.29 s.d.), and show less asocial (0.23 s.d.) and introverted (0.09 s.d.) behavior when they have a sister. These effects remain significant conditional on SES, which suggests that the correlations in the main analysis can be interpreted as causal relations.

In addition, the Bonferroni multiple hypothesis testing method is applied to correct for possible type I errors (i.e., statistically significant outcomes due to chance). Using the Bonferroni multiple hypothesis test, the effects for oldest boys on behavioral problems remain significant. The results for the youngest boys remain significant with respect to self-esteem. For the oldest low SES girls, the results on self-esteem, and disorganized and asocial behavior remain significant, just as the result on disorganized behavior for high SES girls. The results on disorganized and asocial behavior of the youngest girls from the low SES group remain significant as well. In sum, the results show that there are robust positive effects of having a sister for the development of personality traits that have been shown to have positive longer run effects.

It is plausible that sibling gender does not only affect personality traits, but also other longer term outcomes, such as education and health. Overall, boys are less likely to have been in contact with the police (0.12 – 0.13 s.d.) and more likely to have a healthy BMI (0.12 – 0.14 s.d.) when they have a sister. A similar pattern for long run outcomes is found for girls: oldest girls from both SES groups and youngest girls from the low SES group are less likely to have been in contact with the police (0.12 – 0.19 s.d.) and are more likely to have a healthy BMI (0.06 – 0.07 s.d.) when they have a sister.

The results of Chapter 2 and 3 provide the following answer to the question *‘How does family structure affect children’s personality?’*: a larger birth gap increases disorganized and neurotic behavior of the youngest child in a two-child household. On the other hand, having a sister decreases the chance to develop negative personality traits. Having a sister also decreases the chance to have been in contact with the police and having an unhealthy BMI. The family structure one grows up in thus has long term consequences for children.

One of the strengths of the studies in Chapter 2 and 3 is the availability of a wide variety of personality traits and other life outcomes, covering a large part of the life span. This allows for an analysis of the effects of birth gap and sibling gender in both the short and long term. Furthermore, the employment of miscarriages as an instrumental variable and the randomness of sibling gender allow for causal interpretations of the results. One of the limitations and a recommendation for future research is that mechanisms which may help to understand the underlying factors that might play an important role could not be studied with the current dataset.

Part II

The second part of this thesis has analyzed how the educational system in the Netherlands affects the educational outcomes of migrant and refugee children. The main question of this part is: *How does the educational setting affect educational outcomes of migrant and refugee children?* Children with a migration background need to adjust to a new culture, learn a new language and catch up compared to native children. The design of the educational setting can be an important factor that may help or hinder them with this adjustment. The data used in Part II are Dutch student-level administrative data of all children who were enrolled in Dutch primary schools in the period 2008-2016. These data report the school exit test scores and teacher assessments at the end of primary school, as well as a variety of demographic variables. I have studied two characteristics of the Dutch educational setting. The first one is how a teacher creates his/her assessment of the cognitive level of a child: an important factor in determining the secondary school track enrollment. The second aspect is whether placement at a specialized refugee

school helps the development of refugees relative to placement at a regular school: up until now there is no research available on the effects of such a placement. Both aspects are highly debated in policy circles, and especially in the Netherlands more information is needed on the effects of teacher assessment and placement at a refugee school.

Chapter 4 investigated the following question: *What is the effect of the inability to observe the primary school exit test scores on the teacher's assessment of immigrant children?* The identification strategy exploits a policy change and uses a difference-in-differences framework to compare the difference in teachers' assessments and test scores of migrant students for different migrant groups relative to natives (pre- and post-policy change). This policy change entails a change in the timing of the observation of the school exit test score and the creation of the teacher's assessment, such that from 2014/2015 onwards teachers were not able to observe the school exit test score anymore while giving their assessment of the students. The idea behind this policy change was to put more weight on the teacher's assessment of the students in the secondary school track placement decision. Since the exit test score presumes a certain level of knowledge about the Dutch language, one would expect that the exit test scores of migrant children does not fully reflect their capacities. If this is true, the policy change should have had a positive effect on the teacher assessments of students with a migrant background, since the teachers cannot observe the negatively biased exit test scores of migrants anymore. However, the main results indicate that the policy change had the opposite of the intended effect. After the policy implementation, the share of immigrants who received a higher teacher assessment than their test score would predict increased less than the share of natives who received a higher teacher assessment. Therefore, the relative advantage immigrants had over natives in terms of the discrepancy between teacher assessments and test scores decreased after the policy change.

The difference-in-differences (DiD) framework allows to measure the size of the policy effect for each migrant group compared to natives. The results indicate that migrants receive a higher teacher assessment compared to the test score relatively to natives. After the policy change, the teacher assessment on average increases compared to the test score. However, the increase in teacher assessment after the policy is higher for natives compared to migrants. This results in a lower relative advantage of migrants compared to natives in terms of higher teacher assessments compared to the exit test score after the policy change. The share of first generation Western and non-Western immigrants who receive a teacher assessment that is *lower* than their test score increases by 0.08-0.09 compared to the share of natives after the policy implementation.

The identifying assumption of the DiD results is based on the existence of a common trend prior to the policy change. The results of an event study indicate

that the common trends assumption holds for the first generation Western and non-Western migrants. Additionally, to check whether the main results are driven by the selection of schools that partake in the school exit test and the type of test, we compare results of the total population to results of a subsample, consisting of schools which have participated in the same school exit test score throughout the observational period. The DiD estimators of the subsample are not significantly different from those of the total population, therefore the results of the total population arguably are not driven by selectivity.

The answer to the question *“How does the inability to observe primary school exit test scores affect teacher assessment of migrant children?”* is therefore that the policy change had a negative effect on the relative advantage that immigrants had over natives with respect to higher teacher assessments compared to exit test scores. This result is in line with the negative effect of this policy change on secondary school track placement and track switching (Swart et al., 2019; Feron, 2018). The main strength of this research is the use of administrative data. This allows for an interpretation of the results for all students enrolled in Dutch primary education, instead of a subsample. In addition, the use of a policy change in combination with a check for common trends allows a causal interpretation of the results. The findings of the study are relevant since they provide important guidance for policy makers. One of the limitations of this research is that it analyses the intended secondary school placement at the end of primary school, and not the actual placement. An interesting follow-up study would be to check at which level these students are enrolled in the third grade of secondary school, in order to check for longer term effects of the policy change.

Chapter 5 analyzed the question: *What is the effect of specialized refugee schools on refugees’ test scores and teacher assessments?* To investigate this question, I exploit the Dutch immigration allocation policy as a natural experiment. This allocation policy entails that refugees are randomly placed in housing locations and their children are then enrolled at the nearest primary school. This can either be a regular primary school or a specialized refugee school. The refugee schools are solely meant for newly arrived immigrants and provide a different educational environment than regular schools. At the specialized schools, the main focus is on learning the Dutch language, with personalized learning goals and teachers who have experience with traumatized children and teaching Dutch for non-natives. At regular schools, immigrants follow the standard curriculum with additional Dutch classes. Some argue that specialized refugee schools are better for refugee children, because of their tailored approach and experience with language and traumas. Others argue that refugee schools are a parallel schooling system, which might have a negative impact on immigrant children’s outcomes due to negative peer effects and lack of contact with native students.

The results of the balance test on country of origin, gender, arrival cohort and age at arrival confirm that the refugee children are indeed randomly placed into either regular primary schools or specialized refugee schools. This random placement is in line with findings in Sweden and Denmark on the random distribution of refugees over municipalities (e.g. Edin et al., 2003; Åslund & Rooth, 2007; Åslund & Fredriksson, 2009; Åslund et al., 2011; Damm, 2009 and 2014; Damm & Dustmann, 2014). The main findings show that there is no significant effect of being placed at a refugee school on test scores and teacher assessments. The point estimates are also small, indicating that the insignificance of the results does not appear to be driven by a lack of statistical power. In addition, these results are not driven by attrition, since the chance of being observed is not significantly affected by the treatment, conditional on the control variables.

One of the strengths of this study is that it exploits a random allocation mechanism, which solves endogeneity issues with respect to selection (a common problem in the immigration literature). One of the limitations of this study is that the results only hold for a small subset of all the immigrants, namely refugees from Iran, Iraq, Somalia, Afghanistan, Syria and Eritrea. There are other children with migrant backgrounds who enroll at specialized refugee schools and for whom this placement might have a different effect on test scores and teacher assessments. Furthermore, this paper does not study the effect of enrollment of refugee children at regular schools on the educational outcomes of natives. Both these questions are interesting ways forward for future research.

Part II of this thesis has investigated the question: *How does the educational setting affect educational outcomes of migrant and refugee children?* The results show that the educational setting can have an influence on the educational outcomes of vulnerable students, in this case migrant and refugee children. I find a negative influence of a policy change that caused teachers to be unable to observe school exit test scores while creating their assessment on the relative advantage of immigrants over natives with respect to the discrepancy of teacher assessments and test scores. I find no significant effects of placement at a refugee school on test scores and teacher assessments of refugee children. With the usage of administrative student-level data it becomes possible to draw conclusions like these, even though it is not possible yet to test for all the relevant mechanisms (as is the case in Chapter 5).

The background of the page is a complex, abstract geometric pattern. It features a series of concentric, wavy lines that create a sense of depth and movement. In the center, there is a prominent diamond-shaped pattern that resembles a woven or mesh-like structure. The overall color palette is a range of greys, from light to dark, which gives the design a modern and sophisticated feel.

Valorisation addendum

Part I

The first part of this thesis has investigated how family composition, in particular birth gaps and sibling gender, affects personality traits. These two aspects of family composition are interesting to study since children do not have an influence on them, while policy makers can use the insights to better tailor family policy.

The results of the studies show that a larger birth gap increases disorganized and neurotic behavior of the youngest child in a two-child household, while having a sister decreases the chance to develop negative personality traits. Since personality traits can have an effect on long term outcomes, such as crime and health, it is important to mitigate possible adverse effects.

Policy makers can play a role in decreasing and/or preventing the development of negative personality traits. First of all, they should create awareness for children at risk. These children should be signaled as early as possible. Since the development of negative personality traits are already visible at age 5, the children who display negative behavior can enroll in a trajectory of more intensive guidance with the help of e.g. parents and teachers. Second, incentives to decrease the birth gap should be implemented. This can be done by creating awareness of the adverse risk of a higher birth gap through e.g. advertising or by implementing financial incentives. An example of a policy that successfully decreased the birth gap is the higher maternity leave benefit for mothers with children with a smaller birth gap in Sweden (Pettersson-Lidbom & Skogman Thoursie, 2009). Third, the birth gap can also be reduced when more miscarriages are prevented. One important factor that is associated with a higher chance of having a miscarriage is a higher age of the mother. The same interventions, creating awareness and providing financial incentives, can be used to reduce the mother's age when getting pregnant.

Part II

The second part of this thesis has looked at the effect of the Dutch educational setting on the educational outcomes of migrant and refugee children. Lower educational attainment has been associated with several long-term negative outcomes, such as lower wages, more health issues and a shorter life expectancy. Therefore, it is important to create educational policies that increase educational outcomes, especially for vulnerable groups, such as refugee children and children with a migration background. The results of the second part show that a recent policy change with respect to the timing of the primary school exit test had a negative effect on the relative advantage of migrant children over natives with respect to the discrepancy between teacher assessments and test scores. The second main result of this part is that placement at a specialized refugee school had no significant effects on teacher assessments and test scores of refugee children.

The results have important implications for policy. First of all, the change in the timing of the school exit test has negative effects on the teacher assessments of migrant children. As a result, they have a relative disadvantage with possible long-term consequences in terms of equality. To counteract this disadvantage, policy makers could decide to revert to the old policy or create a new one in which children with a migrant background are not disadvantaged in terms of teacher assessments. The null-results of placement at a refugee school should be interpreted carefully. They do not yet provide a conclusive answer to the question which educational environment yields better results for immigrants. There is a possibility that mechanisms (like effects of smaller classes and peer effects) cancel each other out. More research is needed to understand whether there is no effect at all or whether underlying mechanisms create contradicting results (with no overall effect). What is important to note is the fact that regardless of the school environment, refugee children still score lower on the school exit test and receive lower teacher assessments than their native Dutch peers. Investing in refugee schools or regular school with refugees by e.g. hiring extra personnel, increasing salaries or creating smaller classrooms, is a wise way forward to reduce this achievement gap.

Final remarks

This thesis has looked at two cards that are dealt in life, namely family structure and educational setting. As a child, one does not have any influence on these cards, but policy makers can affect the cards and the effects of the cards on outcomes for children. Especially the use of big data, such as the student-level administrative data in Part II, provide new and relevant insights for policy makers on how they should influence (the effects of) the cards dealt to children. The identification strategies used in this thesis provide causal interpretations of the results. This is very important in the discussion on which policies work best, since non-causal results might provide false conclusions and result in adverse policies. A drawback of the two datasets used in this thesis is the unavailability of certain variables in order to check for underlying mechanisms. Thus, one of the main challenges for future research is to disentangle these mechanisms. In addition, more research is needed on factors that enhance or decrease the effects of the studies in this thesis, like the family structure or educational settings.

Bibliography

The background of the page is a grayscale abstract design. It features a series of concentric, wavy lines that create a sense of depth and movement. In the lower right quadrant, there is a prominent pattern of interlocking diamond shapes, which appears to be a close-up of a woven or mesh-like structure. The overall effect is a complex, layered geometric composition.

- Alesina, A., Carlana, M., La Ferrara, E. & Pinotti, P. (2018). *Revealing stereotypes: Evidence from immigrants in schools*. NBER Working Paper 25333, National Bureau of Economic Research, Stanford
- Altonji, J. G. & Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives, in Abowd J. M., & Richard, B. (ed.) *Immigration, Trade and Labor*. The University of Chicago Press, Chicago, pp. 201–234
- Altonji, J. G., Cattan, S. & Ware, I. (2017). Identifying sibling influence on teenage substance use. *Journal of Human Resources*, 52(1), pp. 1–47
- Altonji, J. G., Elder, T. E. & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113, pp. 151–184
- Angrist, J. & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114, pp. 533–575
- Åslund, O. & Fredriksson, P. (2009). Peer effects in welfare dependence: quasi-experimental evidence. *Journal of Human Resources*, 44(3), pp. 798–825
- Åslund, O. & Rooth, D. (2007). Do when and where matter? Initial labour market conditions and immigrant earnings. *The Economic Journal*, 117, pp. 422–448
- Åslund, O., Edin, P., Fredriksson, P. & Grönqvist, H. (2011). Peers, neighborhoods, and immigrant student achievement: evidence from a placement policy. *American Economic Journal: Applied Economics*, 3, pp. 67–95
- Barclay, K. (2015). A within-family analysis of birth order and intelligence using population conscription data on Swedish men. *Intelligence*, 49, pp. 134–143
- Barone G., D'Ignazio, A., De Blasio, G. & Naticchioni, P. (2016). Mr. Rossi, Mr. Hu and politics: The role of immigration in shaping natives' political preferences, *Journal of Public Economics*, 136, pp. 1–13
- Baubock, R. & Tripkovic, M. (2017). *The Integration of Migrants and Refugees*. European University Institute, Robert Schuman Centre for Advanced Studies, Florence
- Biehl, A., Hovengen, R., Groholt, E., Hjelmeth, H., Strand, B. H. & Meyer, H. E. (2014). Parental marital status and childhood overweight and obesity in Norway: A nationally representative cross-sectional study. *BMJ Open*, 4:e004502. doi: 10.1136/bmjopen-2013-004502
- Bingley, P. Lundborg, P. & Lyk-Jensen S. (2017). *Brothers in arms: Spillovers from a draft lottery*. Technical Report Discussion Paper 10483, IZA, Bonn
- Bjerkedal, T., Kristensen, P., Skejeret, G. & Brevik, J. I. (2007). Intelligence test scores and birth order among young Norwegian men (conscripts) analyzed within and between families. *Intelligence*, 35(6), pp. 503–514
- Black, S., Devereux, P. & Salvanes, K. (2007). *Older and wiser? Birth order and IQ of young men*. NBER working paper 13237

- Black, S.E., Breining, S., Figlio, D.N., Guryan, J., Karbownik, K., Nielsen, H.S., Roth, J. & Simonsen, M. (2017). *Sibling spillovers*. NBER Working Paper 23062, National Bureau of Economic Research, Stanford
- Bleske-Rechek, A. & Kelley, J. (2014). Birth order and personality: a within-family test using independent self-reports from both firstborn and later born siblings. *Personality and Individual Differences*, 56, pp. 15–18
- Bloch, A., Chimienti, M., Counilh, A., Hirsch, S., Tattolo, G., Ossipow, L. & Wihtol de Wenden, C. (2015). *The children of refugees in Europe: aspirations, social and economic lives, identity and transnational linkages*. Final report – working paper, SNIS, Geneva
- Bol, T. & Van de Werfhorst, H.G. (2013). Educational systems and the trade-off between labor market allocation and equality of educational opportunity. *Comparative Education Review*, 57(2), pp. 285–308
- Borghans, L., Golsteyn, B.H.H., Heckman, J.J. & Humphries, J.E. (2016). What Grades and Achievement Tests Measure. *Proceedings of the National Academy of Science*, doi: 10.1073/pnas.1601135113
- Breining, S.N., Joseph, J., Doyle, J., Figlio, D.N., Karbownik, K. & Roth, J. (2017). *Birth order and delinquency: Evidence from Denmark and Florida*. NBER Working Paper 23038, National Bureau of Economic Research, Stanford
- Brenoe, A. (2018). *Origins of gender norms: sibling gender composition and women's choice of occupation and partner*. IZA Working Paper Series No. 294
- Brown, S. L. (2006). Family structure transitions and adolescent well-being. *Demography*, 43(3), pp. 447–461
- Brunello, G. & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22, pp. 781–861
- Buckles, K. & Munnich, E. (2012). Birth spacing and sibling outcomes. *Journal of Human Resources*, 47(3), pp. 613–642
- Burgess, S. & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31, pp. 535–576
- Butcher, K. & Case, A. (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, 109(3), pp. 531–563
- Byrne, L. K., Cook, K. E., Skouteris, H. & Do, H. (2011). Parental status and childhood obesity in Australia. *International Journal of Pediatric Obesity*, 6, pp. 415–418
- Bzostek, S. H. & Beck, A. N. (2011). Familial instability and young children's physical health. *Social Science and Medicine*, 73(2), pp. 282–292
- Checchi, D. & Flabbi, L. (2006). *Intergenerational mobility and schooling decisions in Italy and Germany*. IZA Discussion Paper 2876, IZA, Bonn
- Chen, A. Y. & Escarce, J. J. (2014). Family structure and childhood obesity: An analysis through 8th grade. *Maternal Child Health Journal*, 18, pp. 1772–1777

- Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Whitmore Schanzenbach, D. & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR, *Quarterly Journal of Economics*, 126, pp. 1593–1660
- Collier, V. & Thomas, W. (2009). *Educating English learners for a transformed world*. Dual Language Education of New Mexico, Fuente Press, Albuquerque
- Conde-Agudelo, A., Rosas-Bermudez, A. & Norton, M. (2016). Birth spacing and risk of autism and other neurodevelopmental disabilities: a systematic review. *Pediatrics* 137(5): e20153482
- Conley, D. (2000). Sibship sex composition: effects on educational attainment. *Social Science Research*, 29(3), pp. 441–457
- Cyron, L., Schwerdt G. & Viarengo, M. (2017). The effect of opposite sex siblings on cognitive and noncognitive skills in early childhood. *Applied Economics Letters*, pp. 1–5
- Damm, A. (2009). Ethnic enclaves and immigrant labor market outcomes: quasi-experimental evidence. *Journal of Labor Economics*, 27(2), pp. 281–314
- Damm, A. & Dustmann, C. (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review*, 104(6), pp. 1806–1832
- Damm, A. (2014). Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants. *Journal of Urban Economics*, 79, pp. 139–166
- Danielsbacka, M., Tanskanen, A.O., Jokela, M. & Rotkirch, A. (2011). Grandparental child care in Europe: Evidence for preferential investment in more certain kin. *Evolutionary Psychology*, 9(1), pp. 3–24
- De Boer, H., Bosker, R. J. & van der Werf, M. P. C. (2010) Sustainability of teacher expectation bias effects on long-term student performance, *Journal of Educational Psychology*, 102, pp. 168–179
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2), pp. 158–165
- Dupriez, V. & Dumay, X. (2006). Inequalities in school systems: Effect of school structure or of society structure? *Comparative Education*, 42(2), pp. 243–260
- Dustmann, C., Fasani, F., Fattini, T., Minale, L. & Schönberg, U. (2017). On the economics and politics of refugee migration. *Economic Policy*, 32(91), pp. 497–559
- Edin, P., Fredriksson, P. & Åslund, O. (2003). Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment. *Quarterly Journal of Economics*, 118(1), pp. 329–357
- Evans, W. & Fitzgerald, D. (2017). *The economic and social outcomes of refugees in the United States: Evidence from the ACS*. NBER Working Paper No. 23498, National Bureau of Economic Research, Cambridge

- Fairlie, R.W., Hoffmann, F. & Oreopoulos, P. (2014). A community college instructor like me: race and ethnicity interactions in the classroom. *American Economic Review*, vol. 104(8), pp. 2567–2591
- Feinstein, L. (2000). *The relative economic importance of academic, psychological and behavioural attributes developed on childhood*. CEPDP, 443, Centre for Economic Performance, London School of Economics and Political Science, London
- Feron, E. (2018). *The role of cognitive tests and teachers in the transition from primary to secondary education*. Datawyse / Universitaire Pers Maastricht, Maastricht. <https://doi.org/10.26481/dis.20180629ef>
- Fredriksson, P., Öckert, B. & Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics*, 128(1), pp. 249–285
- Gennetian, L. (2005). One or two parents? Half or step siblings? The effect of family structure on young children's achievement. *Journal of Population Economics*, 18(3), pp. 415–436
- Gerritsen, S., Kattenberg, M. & Kuijpers, S. (2019). *The impact of age at arrival on education and mental health*. CPB Discussion paper, CPB, Den Haag
- Gibbons, S. & Chevalier, A. (2008). Assessment and age 16+ education participation. *Research Papers in Education*. 23(2), pp. 113–123
- Glock, S., Krolak-Schwerdt, S. & Pit-ten Cate, I. M. (2015). Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory, *European Journal of Psychology of Education*, 30, pp. 169–188
- Golsteyn, B.H.H. & Magnée, C.A.J. (2017). Does birth spacing affect personality? *Journal of Economic Psychology*, 60, pp. 92–108
- Gould, E. D., Lavy, V. & Paserman, M.D. (2004). Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel. *Quarterly Journal of Economics*, 119(2), pp. 489–526
- Gray, R. H., Simpson, J. L., Bitto, A. C., Queenan, J. T., Li, C., Kambic, R. T., Perez, A., Mena, P., Barbato, M., Stevenson, W. & Jennings, V. (1998). Sex ratio associated with timing of insemination and length of the follicular phase in planned and unplanned pregnancies during use of natural family planning. *Human Reproduction*, 13(5), pp. 1397–1400
- Griffin, K. W., Botvin, G. J., Scheier, L. M., Diaz, T. & Miller, N. L. (2000). Parenting practices as predictors of substance use, delinquency, and aggression among urban minority youth: Moderating effects of family structure and gender. *Psychology of Addictive Behaviors*, 14(2), 174–184.
- Guryan, J. (2001). *Desegregation and black dropout rates*. NBER Working Paper No. 8345, National Bureau of Economic Research, Massachusetts
- Hajer, M. & Pulles, M. (2017). Taalonderwijs aan nieuwkomerskinderen. *Beter begeleiden*, pp. 18–23

- Halla, M., Wagner, A. F. & Zweimüller, J. (2017). Immigration and voting for the extreme right, *Journal of the European Economic Association*, 15, pp. 1341–1385
- Hanushek, E. A. & Woessmann, L. (2006). Does educational tracking affect performance and inequality? Differences in differences evidence across countries. *The Economic Journal*, 116, C63–C76
- Harknett, K. (2009). Why are children with married parents healthier? The case of pediatric asthma. *Population Research and Policy Review*, 28(3), pp. 347–365
- Hauser, R. & Kuo, H. (1998). Does the gender composition of sibships affect women's educational attainment? *The Journal of Human Resources*, 33(3), pp. 644–657
- Hopwood, B., Hay, I. & Dymont, J. (2016). The transition from primary to secondary school: Teachers' perspectives. *The Australian Educational Researcher*, 43, pp. 289–307
- Hotz, V. J. & Pantano, J. (2015). Strategic parenting, birth order, and school. *Journal of Population Economics*, 28(4), pp. 911–936
- Hoxby, C. (2000). The effects of class size on student achievements: new evidence from population variation. *Quarterly Journal of Economics*, 115, pp. 1239–1285
- Inspectie van het Onderwijs (2016). *De staat van het Onderwijs 2016*. Inspectie van het Onderwijs, Utrecht
- Inspectie van het Onderwijs (2018). *De staat van het Onderwijs 2018*. Inspectie van het Onderwijs, Utrecht
- Inspectie van het Onderwijs (2019). *De staat van het Onderwijs 2019*. Inspectie van het Onderwijs, Utrecht
- Jensen, C. & Thomsen, J. P. F. (2013). Can party competition amplify mass ideological polarization over public policy? The case of ethnic exclusionism in Denmark and Sweden, *Party Politics*, 19, pp. 821–840
- Jensen, P. & Rasmussen, A. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30(6), pp. 1503–1515
- Joensen, J.S. & Nielsen H.S. (2018). Spillovers in education choice. *Journal of Public Economics*, 157, pp. 158–183
- Kaestner, R. (1997). Are brothers really better? Sibling sex composition and educational achievement revisited. *The Journal of Human Resources*, 32(2), pp. 250–284
- Kalmijn, M. & van de Werfhorst, H. (2016). Sibship size and gendered resource dilution in different societal contexts, *PLoS ONE*, 11(8), e0160953
- Kanazawa, S. (2012). Intelligence, birth order and family size. *Personality and Social Psychology Bulletin*, 38(9), pp. 1157–1164

- Krein, S. F. & Beller, A. H. (1988). Educational attainment of children from single-parent families: Differences by exposure, gender, and race. *Demography*, 25(2), pp. 221–234
- Krueger, A. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2), pp. 497–532
- Krueger, A. & Whitmore, D. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project Star. *The Economic Journal*, 111(468), pp. 1–28
- Langton, C. E. & Berger, L. M. (2011). Family structure and adolescent physical health, behavior, and emotional well-being. *Social Service Review*, 85(3), pp. 323–357
- Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10–11), pp. 2083–2105
- Ludemann, E. & Schwerdt, G. (2013). Migration background and educational tracking. Is there a double disadvantage for second-generation immigrants? *Journal of Population Economics*, 26, pp. 455–481
- Maliepaard, M., Witkamp, B. & Jennissen, R. (2017). *De economische en sociale participatie van asielzoekers*. WODC, Den Haag
- McBrien, J.L. (2005). Educational needs and barriers for refugee students in the United States: a review of the literature. *Review of Educational Research*, 75(3), pp. 329–364
- McHale, S. M., Updegraff, K. A. & Whiteman, S. D. (2012). Sibling relationships and influences in childhood and adolescence. *Journal of Marriage and Family*, 74, pp. 913–930
- McHale, S.M., Crouter, A.C. & Whiteman, S.D. (2003). The family contexts of gender development in childhood and adolescence. *Social Development*, 12, pp. 125–148
- McLanahan, S. & Sandefur, G. D. (1994). *Growing up with a single parent: What hurts, what helps*. Cambridge, MA: Harvard University Press.
- Onderwijsraad (2017). *Vluchtelingen en onderwijs. Naar een efficiëntere organisatie, betere toegankelijkheid en hogere kwaliteit*. Advies, Onderwijsraad, Den Haag
- Onderwijsraad (2018). *Toets wijzer*. Advies, Onderwijsraad, Den Haag
- Peter, N., Lundborg, P., Mikkelsen, S. & Webbink, D. (2018). The effect of a sibling's gender on earnings and family formation. *Labour Economics*, 54, pp. 61–78
- Petterson-Lidbom, P. & Skogman Thoursie, P. (2009). *Does child spacing affect children's outcomes? Evidence from a Swedish reform*. IFAU Working Paper 2009 (7)

- Pit-ten Cate, I. M., Krolak-Schwerdt, S. & Glock, S. (2016). Accuracy of teachers' tracking decisions: Short- and long-term effects of accountability, *European Journal of Psychology of Education*, 31, pp. 225–243
- Pollet, T.V., Fawcett, T.W., Buunk, A.P. & Nettle, D. (2009). Sex-ratio biased towards daughters in among lower-ranking co-wives in Rwanda. *Biology Letters*, 5(6), pp. 765–768
- Rao, N. & Chatterjee, T. (2018). Sibling gender and wage differences, *Applied Economics*, 50(15), pp. 1725–1745
- Rijksoverheid (2016a) Toelating voortgezet onderwijs gebaseerd op schooladvies. <https://www.rijksoverheid.nl/onderwerpen/toelating-middelbare-school/inhoud/toelating-voortgezetonderwijs-gebaseerd-op-schooladvies>
- Rijksoverheid (2016b) Verplichte eindtoets basisonderwijs. <https://www.rijksoverheid.nl/onderwerpen/toelating-middelbare-school/inhoud/verplichte-eindtoetsbasisonderwijs>
- Roberts, B.W., Kuncel, Shiner, N.R., Caspi, R.A. & Goldberg, L.R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), pp. 313–345
- Rohrer, J.M., Egloff, B. & Schmuckle, S.C. (2015). Examining the effects of birth order on personality. *Proceedings of the National Academy of Sciences of the United States of America*, 112(46), pp. 14224–14229
- Salmon, C., Cuthbertson, A. & Figueredo, A. (2016). The relationship between birth order and prosociality: an evolutionary perspective. *Personality and Individual Differences* 96, pp. 18–22
- Segal, C. (2008). Classroom behavior, *Journal of Human Resources*, 43(4), pp. 783–814
- Steinmayr, A. (2016). *Exposure to refugees and voting for the far right: (Unexpected) Results from Austria*. IZA Discussion Paper No. 9790, IZA, Bonn
- Suárez-Orozco, C., Bang, H.J. & Kim, H.Y. (2011). “I felt like my heart was staying behind”: psychological implications of family separations and reunifications for immigrant youth. *Journal of Adolescent Research*, 26(2), pp. 222–257
- Swart, L., Van den Berge, W. & Visser, D. *Eindtoets draagt bij aan beter passend schooladvies*. CPB Policy Brief, Centraal Planbureau, Den Haag
- Timmermans, A. C., Kuyper, H. & van der Werf, M. P. C. (2015). Accurate, inaccurate or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, 85, pp. 459–478
- Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia, *Review of Economics and Statistics*, 88(1), pp. 171–176

- Van der Steeg, M. & Gerritsen, S. (2016). Teacher Evaluations and Pupil Achievement Gains: Evidence from Classroom Observations. *De Economist*, 64, pp. 419–443
- Watanabe, E., Lee, J. S. & Kawakubo, K. (2011). Associations of maternal employment and three-generation families with preschool children's overweight and obesity in Japan. *International Journal of Obesity*, 35, pp. 945–952
- Wilcox, A. J., Weinberg, C. R. & Baird, D. D. (1995). Timing of sexual intercourse in relation to ovulation—effects on the probability of conception, survival of the pregnancy, and sex of the baby. *New England Journal of Medicine*, 333(23), pp. 1517–1521

The background of the page is a grayscale abstract design. It features a series of concentric, wavy lines that create a sense of depth and movement. In the center, there is a pattern of interlocking diamond shapes, resembling a woven or mesh-like texture. The overall effect is a complex, layered geometric composition.

Curriculum Vitae

Cécile Magnée (1993) obtained her bachelor's degree in Economics and Business Economics in 2014. During her bachelor studies she worked as a student assistant on different projects at the Economics and Finance departments of Maastricht University. She spent half a year abroad at Hong Kong University of Science and Technology as part of an exchange programme in fall 2013. She wrote her bachelor thesis as part of the MaRBLe programme: a Maastricht University's excellence programme that brings multidisciplinary scientific research to the bachelor phase. She then continued to study a master's degree in Economic and Financial Research at Maastricht University, which she obtained in 2016. During her two-year research master, she spent half a year as an intern at the Scientific Research Council in The Hague studying the effects and policy implications of the economic crisis of 2008 on the Dutch economy and society. Afterwards she worked as a researcher at the Educational Research and Development Department at Maastricht University studying the effect of internships on thesis grades of master students.

Cécile started her PhD in September 2016 at Maastricht University in cooperation with the Dutch Inspectorate of Education. In 2018, she participated in the KNAW Arts/Science Academy Honours Programme: a programme that brings young researchers and artists together, and she obtained her University Teaching Qualification at Maastricht University. Furthermore, she spent three months abroad at the Centre for Research and Analysis of Migration (CReAM) at University College London in spring 2019.

Since September 2019, Cécile is employed as a scientific researcher at CPB Netherlands Bureau for Economic Policy Analysis where she works on projects related to migration and decentralized governments.

The background of the image is a complex, abstract pattern. It features a series of concentric, wavy lines that create a sense of depth and movement. In the lower right quadrant, there is a distinct diamond or lattice pattern that blends into the overall wavy design. The colors are various shades of gray, giving it a monochromatic yet textured appearance.

ROA Dissertation Series

ROA Dissertation Series

1. Lex Borghans (1993), Educational Choice and Labour Market Information, Maastricht, Research Centre for Education and the Labour Market.
2. Frank Cörvers (1999), The Impact of Human Capital on International Competitiveness and Trade Performance of Manufacturing Sectors, Maastricht, Research Centre for Education and the Labour Market.
3. Ben Kriechel (2003), Heterogeneity Among Displaced Workers, Maastricht, Research Centre for Education and the Labour Market.
4. Arnaud Dupuy (2004), Assignment and Substitution in the Labour Market, Maastricht, Research Centre for Education and the Labour Market.
5. Wendy Smits (2005), The Quality of Apprenticeship Training, Conflicting Interests of Firms and Apprentices, Maastricht, Research Centre for Education and the Labour Market.
6. Judith Semeijn (2005), Academic Competences and Labour Market Entry: Studies Among Dutch Graduates, Maastricht, Research Centre for Education and the Labour Market.
7. Jasper van Loo (2005), Training, Labor Market Outcomes and Self-Management, Maastricht, Research Centre for Education and the Labour Market.
8. Christoph Meng (2005), Discipline-Specific or Academic? Acquisition, Role and Value of Higher Education Competencies, Maastricht, Research Centre for Education and the Labour Market.
9. Andreas Ammermüller (2007), Institutional Effects in the Production of Education: Evidence from European Schooling Systems, Maastricht, Research Centre for Education and the Labour Market.
10. Bart Golsteyn (2007), The Ability to Invest in Human Capital, Maastricht, Research Centre for Education and the Labour Market.
11. Raymond Montizaan (2010), Pension Rights, human capital development and well-being, Maastricht, Research Centre for Education and the Labour Market.
12. Annemarie Nelen (2012), Part-Time Employment and Human Capital Development, Maastricht, Research Centre for Education and the Labour Market.
13. Jan Sauermann (2013), Human Capital, Incentives, and Performance Outcomes, Maastricht, Research Centre for Education and the Labour Market.
14. Harald Ulrich Pfeifer (2013), Empirical Investigations of Costs and Benefits of Vocational Education and Training, Maastricht, Research Centre for Education and the Labour Market.
15. Charlotte Büchner (2013), Social Background, Educational Attainment and Labor Market Integration: An Exploration of Underlying Processes

and Dynamics, Maastricht, Research Centre for Education and the Labour Market.

16. Martin Humburg (2014), *Skills and the Employability of University Graduates*, Maastricht, Research Centre for Education and the Labour Market.
17. Jan Feld (2014), *Making the Invisible Visible, Essays on Overconfidence, Discrimination and Peer Effects*, Maastricht, Research Centre for Education and the Labour Market.
18. Olga Skriabikova (2014), *Preferences, Institutions, and Economic Outcomes: an Empirical Investigation*, Maastricht, Research Centre for Education and the Labour Market.
19. Gabriele Marconi (2015), *Higher Education in the National and Global Economy*, Maastricht, Research Centre for Education and the Labour Market.
20. Nicolas Salamanca Acosta (2015), *Economic Preferences and Financial RiskTaking*, Maastricht, Research Centre for Education and the Labour Market.
21. Ahmed Elsayed Mohamed (2015), *Essays on Working Hours*, Maastricht, Research Centre for Education and the Labour Market.
22. Roxanne Amanda Korthals (2015), *Tracking Students in Secondary Education, Consequences for Student Performance and Inequality*, Maastricht, Research Centre for Education and the Labour Market.
23. Maria Zumbuehl (2015), *Economic Preferences and Attitudes: Origins, Behavioral Impact, Stability and Measurement*, Maastricht, Research Centre for Education and the Labour Market.
24. Anika Jansen (2016), *Firms' incentives to provide apprenticeships – Studies on expected short- and long-term benefits*, Maastricht, Research Centre for Education and the Labour Market.
25. Jos Maarten Arnold Frank Sanders (2016), *Sustaining the employability of the low skilled worker: Development, mobility and work redesign*, Maastricht, Research Centre for Education and the Labour Market.
26. Marion Collewet (2017), *Working hours: preferences, well-being and productivity*, Maastricht, Research Centre for Education and the Labour Market.
27. Tom Stolp (2018), *Sorting in the Labor Market: The Role of Risk Preference and Stress*, Maastricht, Research Centre for Education and the Labour Market.
28. Frauke Meyer (2019), *Individual motives for (re-) distribution*, Maastricht, Research Centre for Education and the Labour Market.
29. Maria Ferreira Sequeda (2019), *Human Capital Development at School and Work*, Maastricht, Research Centre for Education and the Labour Market.

30. Marie-Christine Martha Fregin (2019), *Skill Matching and Outcomes: New Cross-Country Evidence*, Maastricht, Research Centre for Education and the Labour Market.
31. Sanne Johanna Leontien van Wetten (2019), *Human capital and employee entrepreneurship: The role of skills, personality characteristics and the work context*, Maastricht, Research Centre for Education and the Labour Market.
32. Cécile Alice Jeanne Magnée (2020), *Playing the hand you're dealt, The effects of family structure on children's personality and the effects of educational policy on educational outcomes of migrant children*, Maastricht, Research Centre for Education and the Labour Market.