The educational performance of immigrant children: Examination of the native-immigrant education gap

Helene Wind Fallesen

The educational performance of immigrant children: Examination of the native-immigrant education gap

Study paper 97

Abstract

By using administrative register data from Denmark, the educational attainment of children of immigrants is examined. The focus is on immigrant children from non-Western countries and to examine the relative importance of being born in Denmark compared to abroad. By comparing upper secondary completion rates for the native children and immigrant children graduating from compulsory school between 1990 and 2007, the size of the native-immigrant gap is examined, and it is studied whether the immigrants are converging to the educational attainment of the natives. The education production function is estimated as a linear probability model with OLS, and one of the main goals of the empirical analysis is to determine the possible sources of the gap. Therefore, a large number of control variables are included stepwise in the estimations to examine how each factor help explain the gap. Two subanalyses are performed. One includes performance in compulsory school as an additional explanatory variable to examine whether the gap arises in compulsory school rather than during upper secondary school. The second subanalysis focuses on the immigrant children born abroad, and how their age at arrival is related to their educational attainment. The results show that a native-immigrant education gap exists in Denmark, and that the upper secondary completion rate, in particular, is much lower for the immigrants born abroad compared to the natives. Furthermore, the immigrant boys have a larger educational disadvantage than the immigrant girls, when the attainment gap is examined separately for boys and girls. Importantly, the gap has been sharply reduced over the past 20 years, and both immigrant children born in Denmark and abroad are approaching the educational attainment of the natives. For the immigrant children born abroad, later arrival reduces the completion rate, and the disadvantage increases sharply for children arriving after the age of seven. Differences in family background are found to explain the main part of the gap, but when the child's previous performance is included, family background becomes less important. Conditional on grades from compulsory school, the gap in upper secondary school disappears entirely suggesting that it is in compulsory school, the integration effort should be intensified.

Table of Contents

1 Introduction

Non-Western immigrants and their children constitute a growing share of the population in many Nordic countries. Successful integration of immigrants is important for a number of reasons. First and foremost, it is important that all individuals living in Denmark have equal opportunities of accomplishing education and labour market success regardless of ethnicity or socioeconomic status. That is the core of the Danish welfare state. Furthermore, Denmark, like many other Western countries faces an ageing population with less people in the working age to support the growing number of elderly (Statistics Denmark, 2014). This leads to a fiscal sustainability problem, but successful integration of immigrants may be one way to alleviate this problem (Schou, 2006).

Education is an important part of the integration process. In this paper, focus is on the children of immigrants from non-Western countries, and this group of children often have low educated parents. A strong relation between parents' education and the educational attainment of children is widely documented (Deding & Hussain 2005, Dickson et al. 2013). The non-Western children of immigrants might therefore have a lower educational attainment than the natives, all else being equal. Colding et al. (2009) and Rangvid (2010), among others, find that a native-immigrant education gap exists in Denmark, and they explore different explanations to the existence of the gap. To my knowledge, none of the Danish studies, however, focuses on the development in the gap across cohorts. In a recent Norwegian study by Bratsberg et al. (2012), this relationship is examined. They find a convergence across time in the educational attainment of immigrants born in Norway to that of the natives. Such a convergence can be seen as an indicator of successful integration, and it is therefore interesting to examine if a similar development in the gap is present in Denmark. The question, however, remains which educational measure should be used to compare the educational performance or attainment of the native and immigrant children. In the literature, different measures have been employed. One way to measure the gap is by comparing standardized test scores from PISA or TIMSS tests for natives and immigrants (Dustmann et al. 2012, Schneeweis 2011). Another way is to compare the educational level attained by a certain age (Riphahn 2003, Bratsberg et al. 2012) or grade point averages in compulsory school (Böhlmark 2008, Nielsen & Rangvid 2012). In Denmark, upper secondary school give students the necessary competencies to either continue to higher education or directly to the labour market dependent on the chosen educational track. Completion of upper secondary school is therefore crucial for later labour market success. The Danish government also emphasizes the importance of upper secondary school, and a key goal is that 95 percent of a class graduating from compulsory school complete upper secondary school (The Welfare Agreement, 2006). Therefore, the chosen measure of educational attainment, in this paper, is completion of upper secondary education.

The purpose of the paper is to examine the educational attainment of immigrant children compared to native children in Denmark. The focus is on immigrant children from non-Western countries and to examine the relative importance of being born in Denmark compared to abroad. It is not only the size of the native-immigrant education gap that will be examined, it will further be studied how the gap has developed across time, and thereby if there are signs of the immigrants catching up with the natives. This is analysed by using a merged dataset of administrative register data from Denmark covering the period of 1990 to 2012. The chosen measure of educational attainment is *completion of upper secondary education within five years of graduating from compulsory school,* which means that the main analysis is based on data for the 18 cohorts graduating from ninth grade in compulsory school between 1990 and 2007.

The analysis will be performed by estimating different versions of the education production function. Based on the theory of human capital accumulation and the empirical evidence on the education gap, estimation equations will be set up to examine the effect of having immigrant background and being born abroad on the upper secondary completion rate. The models will be estimated as linear probability models using OLS. From a policy point of view, it is of utmost importance for optimal design of integration policies to get a better understanding of the sources of the gap. Therefore, a large number of control variables will be included stepwise to examine how each of them helps to explain the gap. In this way, it is possible to examine why the immigrants have a lower educational performance in the first place, and whether the gap solely exists because of differences in family background variables between immigrants and natives. In a similar way, it will be examined whether the development in the gap is a consequence of compositional changes in the immigrant population across time.

Two subanalyses will be conducted. First, for the cohorts graduating from ninth grade between 2002 and 2007, data on grade point averages is additionally available. It is thereby possible to study the relationship between performance in compulsory school and the upper secondary completion rate. If the gap disappears conditional on grade points in ninth grade, this suggests that the gap arises in compulsory school rather than during upper secondary school. From a policy point of view, this gives important insights into at what age the integration effort should be intensified. The other subanalysis will have a particular focus on the children born abroad. Their educational attainment is not only affected by factors such as parental background and country of origin, but also by the age at which they arrive in Denmark. Immigrants arriving in their early childhood may not experience the same disadvantage as immigrants arriving in their teens. Therefore, this part of the analysis will focus on the effect of age at arrival, and whether some critical arrival age can be determined after which, the completion rate is significantly reduced.

In the main part of the analysis, the effect of having immigrant background and being born abroad is not gender specific. This means that my baseline model does not allow for a potential gender difference in the completion rate between native and immigrant children. For some

source countries, however, there is a large gender gap, and a cultural gender bias may mean that immigrant girls need more integration support than immigrant boys do (Rangvid, 2010). On the other hand, some studies find superior educational performance for immigrant girls compared to boys (Støren & Helland, 2010). Therefore, in extended analysis, the effects of having immigrant background, being born abroad and the catching up rates are allowed to vary by gender to be able to examine possible gender differences.

The remaining part of the paper is organised as follows. Section 2 gives an overview of immigration to Denmark and the structure of upper secondary education in Denmark. Section 3 describes the relevant theory, and section 4 sums up the empirical evidence on the native-immigrant education gap and places the contribution of this paper within that. Section 5 concerns the empirical approach. Here the estimation models are set up and the empirical method is discussed in relation to other often-used alternatives. Section 6 explains the construction of the data set and the various variables are presented. Section 7 is a descriptive analysis of what the data directly reveals about the size of and development in the gap. Section 8 contains results of the empirical analysis and here the results will be interpreted, different explanations of the findings will be discussed, and the results will be compared with empirical evidence from related studies. Section 9 holds a discussion of the possible policy implications of the results and finally, section 10 concludes the paper.

2 The Danish context

Before the examination of the native-immigrant education gap begins, it is necessary to understand the Danish context. Therefore, this section introduces the development in the immigration to Denmark and the Danish school system.

2.1 Immigration to Denmark

As in many other countries, the number and composition of immigrants and their children has changed rapidly in the latest decades. Historically, Denmark is not known as an immigration country. Figure 1 illustrates the development from 1980 to 2014 in the number of immigrants living in Denmark, who are born in Denmark and abroad from Western and non-Western countries, respectively. The first thing to note is the rapid increase in the number of immigrants especially from non-Western countries. In 1980, only three percent of the Danish population consisted of immigrants – two percent from Western countries and one percent from non-Western countries. In 2014, these fractions have sharply increased. Now, Western immigrants on their own constitute four percent of the Danish population and immigrants from non-Western countries more than seven percent. In pure numbers, this translates to a six-fold increase in the number of non-Western immigrants born abroad. For the non-Western immigrants born in Denmark the number is more than 16 times larger in 2014 than in 1980 with an increase from 7,653 to

128,027. This development underscores the importance of proper integration of immigrants in Denmark.

Figure 1: Number of immigrants born in Denmark and abroad living in Denmark
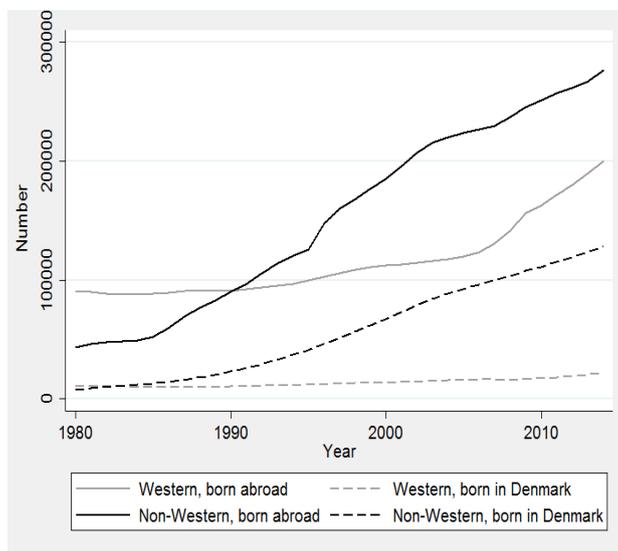
The composition of immigrants has also changed during the years. Around 1970 the first so-called guest workers from Turkey, Pakistan and former Yugoslavia arrived (Nielsen et al., 2003). Before this, immigrants mainly came from other Western countries such as Norway, Sweden, the United Kingdom, Germany and the United States. The immigration of guest workers stopped after the first old crisis in 1973, and hereafter the refugee immigration increased considerably. From figure 1, some of the compositional changes in the immigrant population can also be seen by comparing the number of immigrants born abroad from Western and non-Western countries. In the 1980s, there were twice as many foreign-born immigrants originating from Western countries as from non-Western countries. The number of non-Western foreign-born immigrants did, however, rise rapidly and in 1990, they exceeded the number of Western immigrants born abroad. That the immigration to Denmark is relatively recent also means that the children of the immigrants are very young. Out of the 128,027 Danish-born non-Western immigrants living in Denmark in 2014, 70 percent is under or at the age 18. This furthermore highlights the importance of examining the sources of the low educational attainment of the young immigrants in Denmark.

4

2.2 The educational system in Denmark

The Danish school system consists of nine years of compulsory school[1], a tenth optional year of school, upper secondary education and tertiary education. The focus of this study is on the completion of upper secondary education, and in Denmark, upper secondary school consists of two different tracks – one vocational and one academic. Most of the students wanting to attend upper secondary school is accepted even if they have a low grade point average from compulsory school. The academic track has a duration of two to three years and gives the students the competences to enter advanced education at the tertiary level. Hence, students choosing the academic track will have to continue to a tertiary education to get an education, which directly qualifies them to the labour market. On the other hand, the vocational track gives the students the qualifications of direct use in the labour market. It consists of more than 100 different educations, which qualifies to jobs such as social and health care assistant, carpenter, hairdresser and waiter (Ministry of Education, 2011). Because of the many different educational directions, the duration of the vocational track of upper secondary school varies, but the typical length is three to four years. In Denmark, tuition is free at all levels of the educational system, which means that tuition fee cannot be used as an explanation for not attending upper secondary school. However, studies also show that the problem is not that immigrants are not enrolled in upper secondary school, but that a larger fraction of the immigrants compared to the natives choose to drop out of upper secondary school (Ministry of refugees, immigrants and integration 2005, Colding et al. 2009). This underscores the importance of choosing a measure of educational attainment that includes completion, not only attendance.

The immigration patterns and school system in Denmark has now been briefly explained. Next section includes a presentation of the relevant theory regarding the educational attainment of children with emphasis on the importance of families and the differences between native children and immigrant children.

3 Theory

In economics, educational attainment can be conceptualized through the human capital model and the education production functions. In the human capital literature, education is perceived as an investment in human capital. Becker (1964) was one of the first to use the theory of human capital in an economic context. He defines investments in human capital as activities that influence future income and considers education and training as the most important investments. That is, education is an investment, which will bring economic return in the future, and the underlying assumption is, therefore, that wages will increase, when the individual's skill level

---

[1] In 2009 the Danish legislation was changed from nine to ten years of compulsory schooling, which means that preschool (børnehaveklasse) became compulsory

increases. Based on the theory by Becker (1964), Cahuc & Zylberberg (2004) sets up a simple model to decide the optimal duration of education. The individual has to choose between studying and working, since the assumption is that it is not possible to do both at the same time. The optimal duration of schooling will therefore be determined by the costs and returns associated with getting educated. In the absence of direct costs such as tuition fees, the indirect cost is the loss of earnings, while studying instead of working. The return will depend on the human capital accumulated while getting educated, which again will depend on the individual's efficiency and aptitude. Hence, the theory suggests that the optimal duration of education depends on the expected return and costs, and thereby also on the individual's abilities. In practice, there is uncertainty associated with the future return on education. If immigrants, for instance, due to discrimination expect to receive lower return on education than natives do, this is a possible explanation for the lower educational attainment of immigrants.

The traditional way of modelling the human capital accumulation process is in the framework of the education production function, which was presented by Hanushek (1973). In this literature, educational attainment is modelled as an input-output process, where the various inputs in collaboration creates the output. The outcome is an educational measure and inputs are, for instance, school and family characteristics. That is, both the human capital model and the education production function concern the human capital accumulation of children, but the way education enters and is determined differs in the two theories. In the education production function, educational attainment is determined by different inputs. In the human capital model, education is a choice the individual makes based on economic incentives. Both theories, however, emphasise the importance of families for children's accumulation of skills. Becker (1964) argues that families are important for the human capital accumulation of children and that, differences in preparedness among young children can translate into large differences in educational performance, when they grow older. In the education production function, family characteristics are one of the main input groups. One way parents affect the educational attainment of their children is through their own education and socioeconomic status, and how these are transmitted through generations.

Transmission of status through generations relates to the theory of intergenerational mobility and assimilation. Intergenerational social mobility refers to the relationship between the socioeconomic status of parents and their children (OECD, 2010). The mobility then reflects the extent to which the child is able to change socioeconomic status compared to the parents'. In a society with low mobility, the child's education and occupation is strongly related to that of his parents. In the context of natives and immigrants, the degree of intergenerational mobility is, therefore, important for the potential of assimilation. Gordon (1964) presented some of the early work on assimilation, and the core of the assimilation hypothesis is that differences between immigrants and natives fade across generations. A low degree of intergenerational mobility can therefore be perceived as an obstacle for successful assimilation, when assimilation is measured by immigrants and natives having similar educational attainment or socioeconomic status. Becker &

Tomes (1979) argue that the two main factors determining the degree of intergenerational mobility is the degree of inheritability of endowments, and the propensity of parents to invest in children. Their theory therefore suggests that intergenerational mobility not only relates to abilities or status transmitted through generations, but also to an investment decision of the parents.

The degree of intergenerational mobility is important for the child's accumulation of human capital and in the context of natives and immigrants, it is furthermore interesting to examine whether the mobility is the same across ethnic groups. This is examined by Borjas (1992), who introduces the concept of ethnic capital. He defines ethnic capital as the average skills of the ethnic group in the parent's generation. Borjas argues that the educational attainment of the children of immigrants not only is affected by the parent's input, but also by the ethnic environment, the parents live in. He assumes that ethnicity enters as an externality in the human capital accumulation process, and that differences in skills among ethnic groups may be transmitted through generations and never converge. This means that immigrant children may be influenced by other elements than the native children and therefore, that there may be different effects from different inputs in the human capital accumulation process for natives and immigrants. The direct effect from parental capital may be less for the immigrants pointing to more intergenerational mobility, but when the ethnic capital is accounted for as well, this reduces the intergenerational mobility for the immigrants. This is supported by results by Card et al. (1998).

Studies with focus on the empirical effect of different variables on the educational attainment of children use the education production function. In this analysis, the expected costs and returns on education, which the human capital model proposes as determining the optimal educational level, are therefore not explicitly accounted for[2]. To examine educational attainment in a production function framework in general raises many important questions. What is the appropriate outcome measure? Which input variables should be included, and how should they be measured? Some of these topics are discussed by Hanushek (1979), who points to the conceptual and empirical issues in the estimation of the education production function. Conceptually, there is consensus about a model for educational achievement such as the one below (see e.g. Hanushek (1973) or Todd & Wolpin (2003) for similar equations):

$$A_{it} = f(F_i(t), S_i(t), I_i) \tag{1}$$

Achievement for child i at time t is a function of a group of individual and family characteristics (F_i), school variables (S_i) relevant for child i and child i's innate abilities or endowment (I_i). Both family background variables and school variables must be cumulative until time t, meaning that all current and past family and school inputs have to be included. In other words, equa-

---

[2] Wilson (2001) combines the theory of human capital and education production functions when examining educational attainment in the United States. She finds that even though individuals are affected by economic returns most of the effect of family and school characteristics work through the education process rather than affecting returns to education.

tion (1) is what ideally is estimated, when evaluating the effect of different factors on the educational achievement of an individual. However, data limitations will mean that all relevant inputs never are available. First, the child's innate abilities are, by definition, unobservable and finding an appropriate measure or proxy might be troublesome. Secondly, it is often not possible to include all past and current measures for all relevant family and school input variables. Furthermore, one of the main problems with applying standard production function theory to education is that here there is no homogenous output implying that the appropriate output measure has to be chosen as well as the inputs. That is, I use equation (1) as a point of departure, when the estimation equations used in this analysis are set up in section 5. It will here be described how they differ from the general version of the education production function, and which empirical issues this gives rise to.

In the early work by Hanushek (1973), he also relates the human capital accumulation process to ethnicity. He argues that if there are different effects from the different inputs on the educational performance of the ethnic minority and majority, it is inappropriate to consider them together in the production process, and he therefore estimates the effects in separate analyses. In empirical analyses, the education production function is often both estimated based on a pooled sample of native and immigrant children and separately for immigrants and natives to examine whether the effects differ by immigrant status. In next section, I present some of the empirical evidence from the estimation of the education production function for immigrants and natives.

4   Empirical evidence on the native-immigrant education gap

In the context of examining educational differences between native and immigrant children, the education production function is not only estimated to examine how different factors affect the child's educational attainment. It is furthermore estimated to study how different factors help explain the differences in the educational outcomes between native children and immigrant children. One of the main group of variables in the education production function is family characteristics, which often are found to be strongly related to the educational achievement of a child (Tartari 2015, Adli et al. 2010, Deding & Hussain 2005). Furthermore, most studies regarding the native-immigrant gap agree that differences in family background variables explain an important part of the gap. Dustmann et al. (2012) have compared the educational performance of second-generation immigrants to that of natives in several OECD countries. They find a strong relation between the children's test scores and the parents' educational achievement, and for some countries, the disadvantage for the immigrant children even disappears, when differences in parental background have been accounted for. Todd & Wolpin (2007) get similar results when studying the test score gap in the United States. Rangvid (2010) examines differences in test score gaps by combining PISA data and administrative register data for Denmark. Her re-

sults indicate that less favourable socio-economic background of immigrant children can explain a major part of the test score gaps, but than even after these have been taking into account, the educational performance for both immigrants born in Denmark and abroad is lower than that of the natives. In the theory regarding human capital and ethnicity, it was further argued that immigrants are affected by other factors in the human capital accumulation process meaning that the direct link between parental background and the child's educational achievement might by weaker for immigrants than for natives. This is supported by empirical research. Støren & Helland (2010) examine etnicity differences in the upper secondary completion rate in Norway and find that parents' educational level is of less significance for the children with non-Western background. Their findings suggest that native children benefit more from having highly educated parents, whereas children of immigrants loss less from having parents with lower education. That the parental background variables seem to be of less significance for the immigrants compared to the natives is further supported by Schneeweis (2011), Jakobsen & Smith (2006) and Colding et al. (2009).

The second group of variables, in the education production function, is the school variables. The effect of schools is only included in a few of the studies focusing on the sources of the native-immigrant education gap. Schnepf (2007) analyses the educational disadvantage of immigrants across ten countries and study whether the uneven distribution of immigrants across schools helps explain their educational disadvantage. She finds that it varies for the different countries, but that the effect of having immigrant background is reduced in Switzerland and Germany, when the uneven distribution across schools have been controlled for. The effect of schools is also examined in a Danish context. Rangvid (2007) includes a wide variety of school inputs and characteristics to examine their potential for explaining the native-immigrant education gap in Denmark. Her findings suggest that differences in school quality in schools attended by immigrants and natives may be part of the explanation of the native-immigrant performance gap. It should, however, be noted that her analysis is based on a limited sample size with less than 700 children with immigrant background, and it is therefore questionable, how general the results are.

The child's abilities enter in the education production function as an unobservable variable defined as the child's innate abilities. One way of imperfectly controlling for the child's abilities or skills is to include performance from compulsory school when examining educational attainment in upper secondary school. Grades from compulsory school is found to be the far most predictive factor for early school leaving and non-completion of upper secondary school in Norway (Markussen et al., 2011). In a Danish context, Colding et al. (2005) consider grades in compulsory school as a measure of educational preparedness and argue that the lower grades achieved by immigrant children must take part in the explanation of the difference in educational attainment and the high dropout rates in upper secondary school for the immigrants in Denmark.

When the native-immigrant education gap is examined, the focus is often on either immigrants born in the host country or foreign-born immigrants. The immigrants born abroad have not received all of their education in Denmark if they arrive after school age, and this must affect their educational attainment. Therefore, it might be meaningful to examine the children born abroad in relation to their arrival age. The effect of age at arrival is the focus of several international studies and it is, particularly, interesting to examine if some critical arrival age exists after which the disadvantage becomes extraordinary large. Van Ours & Veenman (2006) have examined the relationship between age at arrival and educational attainment of immigrants in the Netherlands. They find that the critical arrival age depends on both gender and country of origin. Böhlmark (2008) makes a similar analysis for Sweden and finds a strong relation between age at immigrant and school performance. Later arriving children have a significantly lower performance, and he finds the critical arrival age to be around nine.

As mentioned in the previous section, Hanushek (1973) estimates the education production function separately for the natives and immigrants to allow for different effects from the different inputs on the educational performance. Several researchers chose to follow that procedure, and some furthermore divides the sample based on gender to allow for different effects for boys and girls. This is, for instance, done in a Danish analysis by Nielsen et al. (2003), and they find considerable gender differences. In the educational system, immigrant women do better than men, but when they leave the school system and enter the labour market, they seem to face larger problems than men do. Gender differences are also part of the analysis by Rangvid (2010). She finds the gap between Turkish girls and Danish girls to be significantly larger than between Turkish boys and Danish boy, but do not find significant gender differences for the immigrants from other countries of origin. In Sweden, the gender differences are in favour of the girls (Støren & Helland, 2010). The female students have higher upper secondary completion rates than the male students, and the gender difference is larger among the immigrant students than the native students.

The differences in educational attainment between natives and immigrants have in a number of studies been examined in a Danish context, and some of them have been mentioned briefly above. Studies with focus on second-generation immigrants include Nielsen & Rangvid (2012) and Nielsen et al. (2003), among others. Nielsen & Rangvid (2012) argue that the second-generation immigrants is a heterogenous group, and that their educational performance must be related to their parents' length of stay in Denmark. Focus is not on examining the size of the gap between natives and second-generation immigrants, but to identify the explanations of the differences in the educational outcome among the second-generation immigrants. They find a positive relationship between parents' years since migration and the immigrant children's educational performance. Parents' length of stay in Denmark can, therefore, help explain the educational performance differences among immigrant children born in Denmark. Nielsen et al. (2003) focus on the gap between natives and second-generation immigrants in both the educational system and on the labour market. Raw average statistics reveal that the immigrants are

less successful than the natives on a number of areas. On average, they are less likely to obtain a qualifying education, they have longer waiting time until they get their first job, the first employment spell is shorter, and their wages in the first employment is lower. The results suggest parental capital and neighbourhood effects as some of main factors explaining these gaps. Other studies focus on the educational attainment of both immigrants born in Denmark and abroad compared to the natives'. One of these studies is by Colding et al. (2009), who investigate the native-immigrant gap by using a dynamic discrete model of educational choices. They place the foreign-born immigrants arriving after the age of six in a separate group to examine if they experience larger disadvantages than the immigrants born in Denmark. Their findings suggest that dropout rates from vocational upper secondary education, in general, are much higher for immigrant children than native children, and that immigrants arriving after the age of six experience additional barriers with regard to low transition rates to upper secondary school. The results further show that strengthening family characteristics for the immigrants reduces the dropout rate.

The size of the native-immigrant education gap, and the possible factors explaining the gap, have been researched in a Danish context a number of times, and the list above is far from exhaustive. This study does, however, add to the literature in several ways. Colding et al. (2009) focus on the transition from compulsory school to upper secondary school as is the focus of this analysis, but he does not include performance in compulsory school as an input variable. In this analysis, grades from compulsory school is included, and it is thereby possible to examine if the gap in upper secondary school vanishes, when conditioned on compulsory school performance. This will give important insights in relation to the integration effort, and at what age it should be focused. Furthermore, none of the Danish studies, to my knowledge, focuses directly on the effect of each arrival age for the foreign-born immigrants, and thereby on determination of a critical arrival age. Finally, this study adds to the literature by not only examining the size and sources of the attainment gap in Denmark, but also focusing on how the gap has developed over time, and if there are signs of the immigrants' attainment converging to that of the natives. The development in the gap has been examined in a German and Norwegian context (Riphahn 2003, Bratsberg et al. 2012). In Germany, the rather surprising result is that the educational gap between natives and second-generation immigrants, in general, has increased significantly over time rather than decreased. The results from Norway are, however, more positive. They find that the gap has been sharply reduced over the past two decades, and that the second-generation immigrants are approaching the educational performance of the natives. The difference in the results from Germany and Norway further underscores the relevance of examining the development in the native-immigrant education gap in Denmark.

Empirical evidence on the native-immigrant education gap has now been presented, and in the empirical analysis, the evidence will further be discussed and related to the results found in this analysis. However, before the empirical analysis can be performed, the estimation equations have to be set up, and an appropriate estimation method has to be chosen.

5 Empirical approach

The theory section contained a presentation of the education production function, which due to data limitations never can be estimated in its most general form. Therefore, this section begins with a description of the estimation models and explanations of how they differ from the general education production function presented in section 3. Furthermore, the empirical method will be discussed in relation to its advantages and disadvantages and compared to other often-used alternatives.

5.1 Estimation model specification

The purpose of this paper is to examine the average difference in the upper secondary completion rate between immigrant children born in Denmark and abroad and native children. Equation (1) in section 3 is the education production function in its most general form, which is used to model the educational attainment of children. The theory and empirical evidence show that different input variables may affect immigrant children and native children differently suggesting that separate estimations for the two groups are most appropriate. However, an estimation based on the pooled sample of native and immigrant children enhances efficiency. Furthermore, it becomes possible to directly get, for instance, the average marginal effect of being born abroad compared to in Denmark for the immigrant children by including immigrant specific variables. Therefore, the estimation equations, in this section, are specified with a point of departure in equation (1) but will, additionally, include indicators for immigrant status, since it is based on the pooled sample of native and immigrant children. Another modification is that no school variables are included in the analysis, meaning that the group of school related inputs denoted by $S_i(t)$ is omitted from the models. Furthermore, the baseline model will not include a measure of the child's own abilities. If school variables and the child's abilities do have an effect on the educational attainment, this may introduce an endogeneity problem, which will be discussed in next section regarding the empirical method.

Another issue with applying equation (1) is that it requires all past and current inputs to be included. When working with data from administrative registers, you can follow each individual year after year as long as they live in Denmark. Therefore, the different input variables are measured at different points in time. However, since the same input variable measured at different points in time probably will be highly correlated with itself, I have chosen a measurement time instead. It is primarily the parental inputs that may vary across the child's life such as the parents' marital status or socioeconomic status. Some studies show that the effect from the parental inputs is strongest early in the child's life suggesting that these should be measured as early in life as possible (Cunha & Heckman, 2008). Another argument for measuring, especially, socioeconomic status early is that parents may choose to stay home instead of working after they become parents. If parents with low performing children are more inclined to switch status from working to staying home, this introduces a possible endogeneity problem. In this analysis,

the problem with choosing an early measurement time is that the sample includes both immigrants born in Denmark and abroad. The latest arriving immigrants arrive in Denmark at the age of 15. If parental variables are measured when the child is, for instance, eight years old this means that parental variables will be systematically missing for all children arriving after the age of eight. Therefore, the most appropriate approach seems to be to measure the parent variables when the child is 15, since the measurement time then is the same for all the native and immigrant children in the analysis. That is, instead of including all past and current measures of family characteristics, they will be measured when the child is 15.

Another minor change to the formulation is that F_i in equation (1) reflected both individual and family characteristics, where in the following specification, they are divided into two separate variables. With these modifications, the baseline specification in this analysis represents the educational attainment of individual i for the pooled sample of native children and children with immigrant background and is formulated as

$$y_i = \alpha_0 + \alpha_1 IM_i + \alpha_2 IM_i \cdot ABROAD_i + \alpha_3 IM_i \cdot t + \alpha_4 IM_i \cdot ABROAD_i \cdot t \qquad (2)$$
$$+\alpha_5 X_i + \alpha_6 F_i + \varepsilon_i$$

where IM_i is a dummy indicating if individual i has immigrant background, and $ABROAD_i$ is a dummy reflecting if individual i is born abroad. t corresponds to the graduating cohort[3] individual i belongs to. X_i is a vector of individual characteristics, and F_i is a vector of family related inputs, which can be divided into two subgroups: family structures (FS_i) and family background (FB_i). ε_i is an error term. The baseline specification in equation (2) therefore states that the educational attainment of individual i is determined by individual characteristics, family characteristics, and whether individual i has immigrant background and is born abroad.

In the theory section, it was explained that the output has to be chosen as well as the inputs when estimating the education production function. There are several ways to measure the performance, achievement or attainment of the native and immigrant children. Naturally, the measure will depend on the data available and the purpose of the paper. In cross-country examinations it is, for instance, particularly useful to use the standardized test scores from PISA or TIMSS tests, since they are easily comparable across countries. Another approach is to focus on the completed level of education at a certain age instead of specific achievements in an exam or test, and this may be seen as a more broad measure of attainment. In this paper, *completion of upper secondary education within five years of graduating from compulsory school* is the chosen measure of attainment. A similar measure has been used in studies by Riphahn (2003), Støren & Helland (2010) and Bratsberg et al. (2012). It is important that the measure includes completion and not only attendance since one of the main problems in Denmark is that many drop out of upper secondary education and the problem seem to be more severe for immigrants than for natives

---

[3] t takes the value one if individual i belongs to the cohort graduating in 1990, 2 if he belongs to the cohort graduating in year 1991 and so forth

(Colding et al., 2009). Therefore, y_i is an indicator variable taking the value one, if individual i has completed either vocational or academic upper secondary education within five years of graduating from compulsory school, and zero otherwise.

The core of this study is to compare the educational attainment of the native and immigrant children born in Denmark and abroad. Therefore, dummies for having immigrant background and being born abroad is included. The estimated coefficient on IM_i gives the average difference in the upper secondary completion rate between immigrant children and native children, and the estimated coefficient on $IM_i \cdot ABROAD_i$ gives the additional difference for immigrant children born abroad. Several studies document that the immigrants have a significantly lower performance than the natives in Denmark (Colding et al. 2009, Nielsen et al. 2003, Rangvid 2007). If these coefficients are negative and significant, this will be further supported by my analysis. As an extension to the previous Danish studies, I further examine the development in the native-immigrant gap across cohorts. Therefore, having immigrant background and being born abroad are interacted with time trends reflecting the graduating cohorts. The trend terms capture the average annual change in the effect of having immigrant background and being born abroad, respectively. That is, a positive and significant coefficient on $IM_i \cdot t$ will reflect a positive catching up rate for the immigrants, and that the immigrants born in Denmark are catching up with the natives. The sign on the coefficient on $IM_i \cdot ABROAD_i \cdot t$ will indicate whether the catching up rate is larger or smaller for the immigrant children born abroad. The estimated coefficients on these four variables ($IM_i, IM_i \cdot ABROAD_i, IM_i \cdot t$ and $IM_i \cdot ABROAD_i \cdot t$) will be the coefficients of main interest.

As individual characteristics gender, graduating cohort and country of origin dummies are included. Several studies find that girls have a higher educational attainment or performance than boys, and the gender dummy is included to control for this (McNabb et al. 2002, Castagnetti & Rosti 2009, Machin & McNally 2005). The cohort dummies control for general developments, which are cohort specific and affect the completion rate, such as school system changes. The immigrants are a heterogenous group, and several studies find large educational attainment differences across source countries (Rangvid 2010, Schneeweis 2011). The country of origin dummies enter to account for these differences. Besides individual characteristics, family related inputs are expected to affect the educational attainment of the child. As family structure variables, indicator variables are included for whether parents are divorced or not, number of siblings in the family, and parents' age at childbirth. To live in a nuclear family may have positive effect on a child's performance and oppositely, children of divorced parents may have a lower educational attainment (Tartari, 2015). The number of siblings may affect both positively and negatively. One or two siblings can be positive, if they can help each other with homework etc. On the other hand, a large number of siblings may mean less attention from the parents to each child and affect the child negatively. The latter is supported by the findings by Adli et al. (2010). Family background variables include mother and fathers years of education and socioeconomic status. A strong relation between parents' education and socioeconomic status and the child's

educational achievement is widely documented (Deding & Hussain 2005, Dickson et al. 2013, Weinberg 2001). Furthermore, differences in these are often found as main explanations of the native-immigrant gap (Todd & Wolpin 2007, Dustmann et al. 2012).

In general, all the different variables contained in X_i and F_i are included stepwise in the estimations to examine how controlling for them affects the four parameters of interest. If the native-immigrant gap, for instance, only exists because immigrants have a weaker parental background than natives, $\hat{\alpha}_1$ should no longer be significant after including parents' education and socioeconomic status (FB_i). First, a simple model controlling only for gender and cohorts is estimated to examine the size of the raw native-immigrant education gap. Secondly, countries of origin are included, and thirdly family structure variables are added. Next parents' socioeconomic status and years of education are controlled for – first one at a time and then simultaneously. By including the variables stepwise, it becomes possible to examine how each factor helps to explain the educational attainment differences between natives and immigrants. Furthermore, it is possible to check how robust the trend effects are. If a positive development for the immigrants, for instance, is a consequence of secular changes in the source country composition of immigrants, the catching up rates become insignificant once country of origin dummies have been controlled for.

Equation (2) is estimated for the entire pooled sample of native children and children with immigrant background born in Denmark and abroad, and the first analysis does not include any interactions between, for instance, family related variables and immigrant background. This means that the underlying assumption of the model is that all control variables in X_i and F_i affect the native and immigrant children in the same way. If this is not the case, the model is misspecified, and the coefficients will be biased. As discussed, both theory and empirical evidence suggest otherwise, and it is therefore important to check if the model shows signs of misspecification. To examine this models are estimated separately for the native and immigrant children, and the estimated coefficients are compared. Empirical evidence, particularly, show that the effects of parents' education and socioeconomic status are weaker for immigrants than for natives. If the results from the separate estimations in this analysis support these findings, a model, which does not allow the effects to differ by immigrant background, is misspecified. One solution is to drop estimation based on the pooled sample and only include separate estimation results, but pooled estimations are, as mentioned, preferable since they are more efficient compared to separate estimations. The solution is, therefore, to let the effects of parents' education and socioeconomic status vary by immigrant background by introducing more interaction terms. Hence, in the empirical analysis I will use equation (2), where all variables in X_i and F_i enter without interaction with IM_i as a point of departure and then models will be estimated separately for the immigrant and native subsamples to examine potential model misspecification. If the separate estimations show signs of differences in effects of parents' years of education and socioeconomic status, these variables will both enter on their own and interacted with IM_i

for the remaining part of the analysis allowing for differences in the effects for immigrants and natives.

The educational attainment of the immigrant children may be influenced by their parents in other ways than through the family background and family structure variables, mentioned above. The parents' length of stay in Denmark reflects how long time the parents have had to adjust to the Danish society, and thereby the integration potential of the family. Nielsen & Rangvid (2012) find that parents' years since migration explain some of the educational differences among Danish-born immigrants. Therefore, after the most appropriate model specification has been determined, I will examine if controlling for parents' years since migration affects the average differences in the completion rate between immigrants and natives.

As mentioned in section 4 regarding the empirical evidence, as well as making separate estimations for natives and immigrants, some furthermore divide the estimations based on gender to examine gender differences in the education gap. The baseline formulation already includes a gender dummy, which will capture, if there is a general gender difference in the completion rate, which holds for both native and immigrant children. Some studies, however, find that the gender difference among native children differ from the gender difference among immigrant children (Rangvid 2010, Støren & Helland 2010). The disadvantage of having immigrant background might be larger for boys than for girls, or the immigrant girls may have higher catching up rates than the boys. To be able to examine this, I include gender interactions. This results in a modification of equation (2), where the four variables of interest IM_i, $IM_i \cdot ABROAD_i$, $IM_i \cdot t$ and $IM_i \cdot ABROAD_i \cdot t$ both will enter on their own and interacted with the gender dummy. Note, that this changes the interpretation of the estimated parameters. The coefficient on IM_i will now only give the marginal effect of having immigrant background for the boys and to get the marginal effect for the girls the coefficient on the corresponding gender interaction has to be added ($IM_i \cdot FEMALE$).

A limitation of the baseline specification is that it does not include any measure of the child's abilities. The problem is that a child's innate abilities are unobservable, and an appropriate measure or proxy therefore has to be found. One way of controlling for the child's abilities, when examining attainment in upper secondary school, is to include grades from compulsory school. Markussen et al. (2011) find performance in compulsory school to be the main predictor for upper secondary school completion, which underscores the importance of including grades in the estimations. The data only contains grade information for the cohorts graduating from compulsory school in 2002 and onwards. Therefore, the part of the analysis where grades are controlled for is only based on these latest cohorts. This further implies that the inclusion of trend terms no longer is meaningful because of the short time period. Hence, the specification changes to

$$y_i = \delta_0 + \delta_1 IM_i + \delta_2 IM_i \cdot ABROAD_i + \delta_3 X_i + \delta_4 F_i + \delta_5 I_i + \mu_i \qquad (3)$$

where I_i is grade point average dummies in compulsory school and μ_i is an error term. First, a simple model with only gender, cohort and country of origin will be estimated with and without grades to examine, if grade differences close the native-immigrant gap even without accounting for the family variables. Afterwards, a model additionally including parental variables will be estimated with and without grades. Finally, it will be examined whether the effect of grades differ for natives and immigrants by first making separate estimations for the two subsamples and then allowing for interactions between grades and immigrant background, which corresponds to adding $IM_i \cdot I_i$ to equation (3). Furthermore, IM_i and $IM_i \cdot ABROAD_i$ will again be interacted with gender to examine if the possible gender differences remain after grades have been controlled for.

The last part of the analysis focuses on the immigrant children born abroad. In equation (2) and (3), the difference in the completion rate between immigrants born abroad and natives is averaged across all the immigrants born abroad. They do however differ in a very important way, namely, in their length of stay in Denmark. International studies find a strong link between age at arrival and the educational attainment of immigrants (Böhlmark 2008, Van Ours & Veenman 2006). It is therefore interesting to examine each arrival age's effect on the completion rate in Denmark and to study whether a critical arrival age exists. This calls for a modification of the specifications in (2) and (3), since the effect of age at arrival will be examined both for the entire sample and for the subsample of the grade cohorts to see how including grades influences the effect of age at arrival. The dummy variable $ABROAD_i$ will be replaced by indicator variables for each possible age of arrival and the specification changes to

$$y_i = \gamma_0 + \gamma_1 IM_i + \gamma_2 IM_i \cdot AGEATARR_i + \gamma_3 IM_i \cdot t + \gamma_4 IM_i \cdot ABROAD_i \cdot t \qquad (4)$$
$$+\gamma_5 X_i + \gamma_6 F_i + \eta_i$$

and for the estimations including grade point averages

$$y_i = \theta_0 + \theta_1 IM_i + \theta_2 IM_i \cdot AGEATARR_i + \theta_3 X_i + \theta_4 F_i + \theta_6 I_i + \varphi_i \qquad (5)$$

where $AGEATARR_i$ is dummies indicating the child's age at arrival, and η_i and φ_i are error terms. In this part of the analysis, it will again be studied if the effect of the main variables varies by gender by interacting the gender dummy with IM_i, $IM_i \cdot AGEATARR_i$, $IM_i \cdot t$ and $IM_i \cdot ABROAD_i \cdot t$.

An important thing to note regarding the above specifications is that in most cases the average difference in the completion rate between native children and immigrant children has to be computed based on a weighted average of the other coefficients rather than directly estimated. It is only in the first basic estimation, where only gender and cohorts are controlled for that the effect will be estimated. As soon as country of origin dummies are accounted for, this is no longer possible, since these also cover the entire immigrant population, and therefore, including both country of origin dummies and the variable IM_i will induce linear dependence. In practice, this means that IM_i as a variable is omitted from the estimations, when countries of origin are

17

controlled for. Instead, the difference in the completion rate between the immigrants and the natives will be reflected by the coefficients on the different source countries. That is, the average effect of having immigrant background will be calculated as the average of the coefficients on the source countries weighted by the immigrant population shares[4]. Furthermore, when parents' years since migration is included, the average difference will be based on both the weighted average of coefficients on source countries and the weighted average of coefficients on parents' years since migration.

When the effects from parents' education and socioeconomic status are allowed to vary by immigrant background, this complicates the computation. It now has to include the coefficients on the interaction terms, and the question is how the native-immigrant difference should be evaluated. One possibility is to evaluate it across the native parental distribution. This relates to the Blinder-Oaxaca decomposition method, where a gap between two groups is explained by decomposing it into two parts (Oaxaca, 1973). One part of the gap is explained by differences in the size of the determinants, and the other by differences in the effects of the determinants. Including the interaction terms between immigrant background and parental background allow the effects of parental background to differ for natives and immigrants. Evaluating the average differential in the native parental distribution then corresponds to examining, whether the attainment gap persists when immigrants are given the same parental resources as the natives. The same procedure is used when grades are allowed to vary by immigrant background, and the difference in that context, reflects if native-immigrant differentials exist, when immigrants have the same grade point distribution as natives[5].

5.2   Empirical method

After the estimation equations have been formulated, the next question is which regression model and method should be used for the estimations. In the literature, different methods have been applied for various purposes, and they all have their strengths and weaknesses. The empirical model and method used in this paper will now be elaborated on in relation with some of the often-used alternatives.

---

[4] The standard error associated with the average effect of having immigrant background is computed based on the weighted average of the standard errors of the country of origin coefficients, as well. The computation of the standard errors is based on the following rule: $Var(\sum_j \omega_j Z_j) = \sum_j \omega_j^2 Var(Z_j)$ under the assumption that the OLS estimators for the different country of origins are independent. That is the variance of the sum of the weighted coefficient estimate equals the sum of the population weights squared times the variance on each country of origin coefficient. An equivalent procedure holds when the computation of immigrant background becomes more sophisticated because of the interaction terms.

[5] When the average differential is allowed to vary with gender the computations are the same, the only difference is that the coefficients now are the coefficients related to the gender interactions and that the evaluations are based on the female immigrant population share and female native frequency distributions.

5.2.1 LPM and the alternatives

In this analysis, the models are estimated as linear probability models (LPM) using OLS. The LPM is a model, which is criticized by many econometricians, but also often applied in empirical work because of its simplicity (Wooldridge, 2003). The LPM for the binary response y is specified as

$$P(y = 1|\boldsymbol{x}) = \alpha_0 + \alpha_1 IM_i + \alpha_2 IM_i \cdot ABROAD_i + \alpha_3 IM_i \cdot t + \alpha_4 IM_i \cdot ABROAD_i \cdot t \qquad (6)$$
$$+\alpha_5 X_i + \alpha_6 F_i$$

where $\boldsymbol{x}$ denotes the vector containing all the explanatory variables and each of the explanatory variables are defined as described in the previous section. Equation (6) says that the probability of success is a linear function of the explanatory variables. In this context, the probability of success equals the probability of completing upper secondary education within five years of graduating from compulsory school. The main problem with the LPM is that it can predict probabilities below zero or above one, which obviously does not make sense. Secondly, the linearity implies that the partial effects of any explanatory variables are constant meaning that the effect on a child's educational attainment of having one sibling compared to none is the same as having two siblings compared to one. These two disadvantages can be overcome by using more sophisticated binary response models such as the nonlinear probit or logit models, which are used in several studies examining the native-immigrant gap (see for instance Riphahn 2003, Støren & Helland 2010 or Bartolomeo 2011). However, in this context the interest lies in the marginal effects e.g. the effect of having immigrant background compared to being native Dane, so the possible problem with the predicted probabilities being less than zero or above one is of less relevance. Furthermore, the problem regarding the linear effect of the explanatory variables is handled by introducing binary indicators for the explanatory variables. That is, instead of including a variable called *siblings,* which contains the child's number of siblings, binary indicators are included for each possible number of siblings. In this way, the effect of having an extra sibling is allowed to vary with the number of siblings. Since all the explanatory variables, in this analysis, enter as dummies reflecting different categories, the linear effect implied by the LPM is not a problem.

Probit and logit models are used as alternatives to the LPM, when the dependent variable is a binary variable, and the problems with the LPM are considered too severe. Some of the other models applied in the literature includes ordered and multinomial probit and logit models. The multinomial and ordered models are used in situations with more outcome possibilities. If the outcome variables have a natural ranking, ordered models are used and if not, multinomial models are applied. In a study by Markussen et al. (2011), they use a multinomial logit model to identify the factors predicting early school leaving and non-completion in upper secondary school in Norway. The outcome variable in their analysis is divided into three groups: completing upper secondary school within five years, carrying out all years of upper secondary school but not passing all subjects or either did not start or left before finishing upper secondary school.

By using multinomial regression, they have the possibility of examining which factors affect early leaving and non-completion compared to completion, which may give a more nuanced picture. As the name suggests, using an ordered model requires a ranking of the dependent variable, which may sometimes be difficult in relation to education. The method is used in the Danish analysis by Jakobsen & Smith (2006) focusing on the educational attainment of the children of the Danish guest worker immigrants. The dependent variable in their analysis represents the level of completed qualifying education. They order the educational attainment as no qualifying education, vocational education, short theoretical education, medium theoretical education and long theoretical education. It can here be discussed whether the ranking is appropriate. The ranking of the vocational education compared to the theoretical education is, in particular, troublesome and applying an ordered model may not be the most suitable choice when studying upper secondary school in Denmark.

Another alternative to a simple linear model is to model the obtained educational level as an outcome of a number of sequential multinomial decisions. This is done by Colding et al. (2009), who use a dynamic discrete model to examine educational choices. By applying this model, it is possible to address questions regarding choice of path in the educational system, and they can thereby determine at what stages of the educational career the children of immigrants face barriers to educational progression. Furthermore, they can examine how background characteristics affect educational choices at different points in the educational system, which may help explain the observed differences in educational attainment between native children and children of immigrants. Hence, the dynamic model employed by Colding et al. (2009) is very suitable in a multi-track educational system as the Danish. The results are, however, more difficult to directly interpret compared to the results from a LPM.

A number of different models can be applied in the examination of the education gap in upper secondary school. Some of the models included dependent variables with multiple outcomes. In this analysis, a binary model is chosen to keep the analysis simple, but still insightful. One of the clear advantages of the LPM is that the results are easier to directly interpret compared to results from more advanced models, making it easier for policy makers to understand. Furthermore, if the model is saturated, meaning that the explanatory variables are dummy variables for mutually exclusive and exhaustive categories, the LPM is completely general and estimation with probit, logit and LPM will yield the same results (Wooldridge 2010 and Angrist & Pischke 2009).

5.2.2 OLS

The LPM is estimated using OLS, which produces unbiased estimates, if there is no endogeneity. The possible problem of endogeneity has already been commented on in several of the previous sections. The no endogeneity assumption says that the error term has to be uncorrelated with the explanatory variables (Wooldridge, 2003). All variables that are not included explicitly as an explanatory variable will be contained in the error term. This means that if there are variables

that affect the children's educational attainment, which is not included in the estimation equation, and if these variables are correlated with one or more of the explanatory variables in the estimation equations, a problem of endogeneity will exist. This is, for instance, the case if the omitted school variables affect the child's educational attainment and are correlated with some of the explanatory variables in the estimation equations. The main endogeneity problem in this kind of estimations arises because the child's abilities are unobservable. In the education production function seen by equation (1) in section 3, I_i denoted the child's innate abilities, which by definition are unobservable, and therefore will be contained in the error term. This will result in an endogeneity problem if the child's abilities are correlated with some of the explanatory variables in the model. Parents' education and socioeconomic status enter as measures of the parents' abilities, and there must exist some correlation between the abilities of a parent and his/her child. This implies that in equation (2) where the child's own abilities do not enter as an explanatory variable, the effect of parents' education and socioeconomic status may be biased upward, since they also capture the effect of the child's own abilities. The effect of the family background variables therefore has to be interpreted with caution. One way of controlling for the child's abilities is to include grade point averages from ninth grade, which is done in part of the analysis for the latest graduating cohorts. Therefore, the estimations including grades will be expected to suffer less from endogeneity problems, and the expectation is that the effects of parental background variables will be significantly lower in these estimations.

Even though the models include a wide range of family related control variables, there might still be unobserved differences between immigrant and native families, which are not accounted for. One way of handling this problem is to use family fixed effects, which corresponds to including a dummy variable for each family in the data set. It is in this way possible to control for unobserved time-invariant family characteristics. The method is used by several studies focusing on the native-immigrant education gap (see for instance Bratsberg et al. 2012, Böhlmark 2008 and Nielsen & Rangvid 2012). It is, in particular, useful when the focus is on examining the effect of age at arrival on educational performance, which is the focus in parts of this paper. The parents might take the children's age into account when making the migration decision[6]. This can lead to selection bias if the way parents make the decision is correlated with omitted unobserved family characteristics, which at the same time is correlated the child's educational outcome. It might, for instance, be the case that parents who prioritize their children's education tend to migrate when the children is as young as possible, making the children less likely to become educational disadvantaged. This is discussed by Böhlmark (2008), who estimates the effect of age at arrival with and without family fixed effects, but his results show that the estimates are very similar. To apply family fixed effects is, however, not without problem. It restricts the sample to only include families with more than one child, and it can therefore be discussed, how

---

[6] The migration does not necessarily include a decision to make. Therefore, parents who have the possibility of making a choice regarding migration time such as labour-market immigrants may be more likely to migrate with younger children than parents without the choice such as refugees.

general the results really are. Family fixed effects is therefore not used in this analysis, but the results will be discussed in relation to studies both with and without family fixed effects.

The other issue arising when estimating the LPM with OLS is the problem of heteroskedasticity, which does not affect the unbiasedness of the OLS estimator but makes the standard errors and corresponding t-statistics invalid. The homoskedasticity assumption states that the variance of the error term conditional on the explanatory variables is constant for all combination of outcomes of explanatory variables (Wooldridge, 2003). This means that when the variance of the error term depends on the explanatory variables, the error term exhibits heteroskedasticity. The variance of y conditional on the explanatory variables for a LPM is defined as

$$Var(y|\boldsymbol{x}) = \boldsymbol{x\beta}(1 - \boldsymbol{x\beta}) \qquad (7)$$

where $\boldsymbol{x\beta}$ simply is shorthand for the right-hand side of equation (6). Since the variance of the error term equals $Var(y|\boldsymbol{x})$, equation (7) means that heteroskedasticity is present, when OLS is used to estimate the LPM. A simple way of dealing with the heteroskedasticity is to still use OLS estimation but to use heteroskedasticity-robust standard errors (Wooldridge, 2010). Therefore, all standard errors reported in the empirical analysis in section 8 are heteroskedasticity-robust standard errors.

The strengths and weaknesses of the empirical model and method have now been discussed, and it is, particularly, the endogeneity problem one has to keep in mind when interpreting the results. Before continuing to the analysis, an introduction to the data set and the variables now follows.

6   Data

The data in this paper is derived from three primary sources: the population register, the educational institution register, and the labour force register. From the population register demographic characteristics such as gender, age, immigrant status, family structure, country of origin, and date of immigration is retrieved. The educational institution register contains information on highest educational attainment[7] and data on employment status is available from the labour force register. Each individual has a unique ID number and the population register contains the ID numbers of each child's parents as well. It is, thereby, possible to link data on parents and children from the three different registers. For the cohorts graduating from compulsory school in 2002 and onwards, grades from ninth grade are also available from the compulsory school grade register. A separate data set is therefore constructed for the individuals with grade data available, which is merged with data from the three other registers.

---

[7] Education attained prior to immigration comes from a survey-based register.

6.1 Sample selection

The data is extracted from different registers, but the analysis is not based on every individual contained in the registers, and I therefore make a sample selection. The focus of this study is to examine the fraction of a cohort graduating from ninth grade in compulsory school, who manage to complete upper secondary education within 5 years distinguishing between native children and children with immigrant background. The administrative registers contain information on every individual living in Denmark for each year from 1980. Therefore, the first selection is to only include children (and information about their parents), who graduate from compulsory school in each of the years. Children of immigrants are only included in the analysis, if they originate from a non-Western country[8], since their educational attainment typically differ most from that of the natives. Several studies examining the native-immigrant education gap choose to focus on the educational differences between native children and immigrant children from non-Western or less developed countries (see for instance Colding et al. 2009 or Bratsberg et al. 2012). Furthermore, figure 1 in section 2.1 illustrated the sharp increase in the number of immigrants from non-Western countries underscoring the importance of examining the attainment gap between this immigrant group and the natives. As mentioned, data is available on individuals graduating from compulsory school between 1980 and 2012. In figure 1, it is however seen that the number of non-Western immigrants living in Denmark in the 1980s was limited[9]. Therefore, this analysis only includes the cohorts graduating from ninth grade in 1990 and onwards.

In table 1, I report the reduction in the sample size for each of the selection criterion. The first sample size in the table is equivalent to the total number of native children and children with immigrant background from non-Western countries graduating from ninth grade in 1990 or later. It should be noted that the children with immigrant background are extracted from a data set containing the entire Danish population, whereas the native children are extracted from a data set with a ten percent random sample of the Danish population. This means that the 219,581 children in the table is based on the total immigrant population with non-Western background in Denmark, but only ten percent of the native population. In table 1, it is seen that the first selection criterion is that the individual still must be observed in Denmark five years after the ninth grade graduation. This criterion follows from the chosen measure of educational attainment, which in this paper is *completion of upper secondary education within five years of graduating from compulsory school*. It is thereby necessary to be able to follow the individuals five years after their graduation date. The chosen period of five years is based on different considerations. First, the period has to reflect the average duration of upper secondary school. In section 2.2, it was described that the length of upper secondary school varies, but typically is around three years.

---

[8] Non-Western countries are all countries except: EU28, Andorra, Australia, Canada, Iceland, Liechtenstein, Monaco, New Zealand, Norway, San Marino, Switzerland, USA and Vatican City State

[9] There were less than 1000 individuals with immigrant background per graduating cohort in the 1980s

The period of five years therefore allows for both attending the optional tenth grade before attending upper secondary school, and the opportunity to change educational direction for most of the students. Secondly, there is a trade-off between allowing a longer period and reducing the sample size. If the period was extended to six or seven years, this would exclude the cohorts graduating from compulsory school in 2006 and 2007, since the individuals included in the sample then has to be followed for six or seven years after graduation. In table 1, the sample size reduction based on this criterion is divided between those who are excluded because they belong to the cohorts graduating between 2008 and 2012, and those who are excluded because they either emigrated or died. The main part of the reduction is due to the exclusion of the latest cohorts, why allowing for a longer period than five years would significantly reduce the sample size. Therefore, five years is chosen as the period within which the individuals have to have completed upper secondary education after ninth grade, and the analysis is based on the 18 cohorts graduating from ninth grade in the years between 1990 and 2007.

Table 1: Sample size reduction for each sample selection criterion

Description of selection criteria	Reduction	Sample size after selection criteria
Number of native children and children with immigrant background from non-Western countries graduating ninth grade between 1990 and 2012		**219,581**
Still observed five years after ninth grade	65,911	153,670
- Belong to the cohorts graduating between 2008 and 2012	*61,363*	
- Emigrated or died	*4,548*	
No children with one immigrant and one native parent	3,869	149,801
Not abroad for more than a calendar year during compulsory school or the five following years	3,086	146,715
Maximum age at arrival is 15	1,759	144,956
Between 15 and 18 when graduating ninth grade	586	144,370

NOTE: Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007. Natives are extracted from a data set containing a 10 percent random sample of the Danish population and the immigrants from a data set with the entire Danish population.

In the data, children with immigrant background[10] are denoted as children where none of the parents is both Danish citizens and born in Denmark, and the children with immigrant background are further divided according to whether they are born in Denmark or abroad. Native children are by Statistics Denmark denoted as children with *one or more* parent being Danish citizens and born in Denmark. Following this definition, children with one immigrant and one native parent are native. It can be discussed whether it is most appropriate that the children with mixed background are defined as natives, immigrants or simply are excluded. In some Danish studies, they chose to follow the definition by Statistics Denmark and include children with mixed background in the native sample (Rangvid, 2007). In the international comparison

---

[10] For some children there are inconsistency in their 'immigrant status' meaning that they in the data set may e.g. switch from being recorded as immigrant born abroad to immigrant born in Denmark. The inconsistency are repaired by the following rule: If the immigrant status are the same for the first and last year the person are observed, this is set as the immigrant status. If not, the children is set to being born in Denmark if the first observed age is zero and born abroad if the first observed age is different from zero.

paper by Dustmann et al. (2012), they exclude mixed background children in the main part of their analysis, and when they are included they are defined as immigrants. Bratsberg et al. (2012) make an analysis similar to mine for Norway and exclude children with one foreign-born and one native parent. Since the focus of this paper is to examine the difference in educational attainment between native children and children with immigrant background, it is important that the distinction between the two groups is clear. Therefore, children with one immigrant background parent and one native parent are excluded from the sample. This means that the native children in this analysis are children where *both* parents are Danish citizens and born in Denmark.

Furthermore, the sample is limited to individuals who did not die or spend a full calendar year or more abroad during compulsory school or the five subsequent years. This means that the analysis is based on individuals who has not left Denmark for a longer period after they attended compulsory school or immigrated to Denmark. This selection is made to make the sample as homogenous as possible. Individuals, who have left Denmark for some years and then returned, may be influenced by a number of other factors than the children who stayed in Denmark making it harder to compare their educational attainment. It is seen in table 1 that this selection criterion reduces the sample size with approximately 3000 individuals. The last two selection criteria relates to age. For the immigrant children born abroad the maximum age at arrival is set at 15 since this is the age around which children attend ninth grade. Some individuals chose to complete ninth grade later in their life if they did not complete it as children. The focus of this analysis is on children and therefore, individuals are only included if they were between 15 and 18 years, when they completed ninth grade. In table 1, it is seen that this last criterion only reduces the sample size with 586 individuals meaning that more than 99.5 percent of the individuals in the sample already were within this age interval.

Table 2: Sample composition for each graduating cohort

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Natives	6,112	6,043	5,554	5,251	5,289	4,908	4,769	4,409	4,439	4,186
Immigrants born in Denmark	491	579	568	691	785	928	1,122	1,132	1,090	1,047
Immigrants born abroad	793	897	935	1,164	1,183	1,274	1,506	1,706	1,857	1,920
Total	7,396	7,519	7,057	7,106	7,257	7,110	7,397	7,247	7,386	7,153

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Natives	4,296	4,445	4,529	4,557	4,825	5,033	5,145	5,125	88,915
Immigrants born in Denmark	1,149	1,346	1,476	1,834	2,011	2,349	2,523	2,884	24,005
Immigrants born abroad	2,075	2,180	2,196	2,180	2,325	2,501	2,396	2,362	31,450
Total	7,520	7,971	8,201	8,571	9,161	9,883	10,064	10,371	144,370

NOTE: Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007. Natives are extracted from a data set containing a 10 percent random sample of the Danish population and the immigrants from a data set with the entire Danish population. The students in the sample are those who are still observed five years after graduating ninth grade, do not have parents with mixed background, have not been abroad for more than a calendar year during compulsory school or the five following years, are maximum 15 years old when they arrive in Denmark and are between 15 and 18 years old when they graduate ninth grade.

Based on these selection criteria, the result is a data set consisting of 144,370 individuals. Since the natives are extracted from a data set containing a ten percent random sample of the Danish population and the immigrants from a data set with the entire population, there are 88,915 native children in the sample and 55,455 children with immigrant background. Of the children with immigrant background 24,005 are born in Denmark and 31,450 are born abroad. These are the total number of observations in the sample and reflect the number in each group across all cohorts. The number of immigrants in the graduating cohorts increase steadily for each year between 1990 and 2007, which is seen in table 2. Until 2006, the number of immigrants born abroad exceeds the number of immigrants born in Denmark. For the cohorts graduating from ninth grade in 2007 in the sample, there are 5,125 natives, 2,884 immigrants born in Denmark and 2,362 born abroad.

The sample selection has now been argued for, and the subsequent sections will include descriptions of the dependent variable, the variables of main interest, and the control variables. All variables in the analysis are dummy variables[11], and appendix A contains a complete list of the variables including data sources and detailed definitions.

6.2 The dependent variable

In this analysis, *completion of upper secondary education within five years of graduating from compulsory school* is the chosen measure of educational attainment. Therefore, the dependent variable is an indicator variable taking the value one, if the child has completed either vocational or academic upper secondary education within five years of graduating from ninth grade, and zero otherwise. This is determined based on changes in the education code for the child's highest educational attainment. The year the code changes to upper secondary school is compared to the year, the child graduated from ninth grade to identify whether upper secondary school is completed within five years.

6.3 Variables of interest

The four parameters of interest are the estimated coefficients on the immigrant background dummy, the born abroad dummy and their interactions with the time trend denoted by IM_i, $IM_i \cdot ABROAD_i$, $IM_i \cdot t$ and $IM_i \cdot ABROAD_i \cdot t$ in the estimation equations in section 5.1. As mentioned in the sample selection, a child has immigrant background if none of the parents is both Danish citizens and born in Denmark. By definition, the children who are born abroad have immigrant background[12], and for these children the born abroad dummy takes the value one. Native children are children, where both parents are Danish citizens and born in Denmark, and they enter as a reference category. The trend term denoted by t represents the graduating cohort the indi-

---

[11] Except the trend term

[12] If no information is available about the parents and a child is born abroad, he will be contained in the *born abroad* group

vidual belongs to and takes a value between one and 18 corresponding to the 18 cohorts graduating between 1990 and 2007. The graduating cohort corresponds to the year the education code for the child's highest educational attainment changes to ninth grade for the first time.

6.4 Explanatory variables

6.4.1 Individual characteristics

As individual characteristics, gender is included as well as cohort, and country of origin dummies. The gender dummy takes the value one if the individual is female, and there is included a cohort dummy for each of the possible 18 graduating cohorts. Statistics Denmark defines country of origin of immigrants as the mother's country of birth. If only the father is known, his country of birth is used instead, and when none of the parents is known, the country of origin is defined based on the child's own information[13]. In some studies, the source country is controlled for by using dummies for source regions such as Southern Europe or the Middle East (Bauer & Riphahn, 2007). Others include country of origin dummies for each country present in the data set (Bratsberg et al., 2012). A third approach is to include the largest source countries as dummies and make a residual category for the rest of the countries (Riphahn 2003, Rangvid 2010). This is, in particular, relevant when a large part of the sample originates from a limited number of countries. The immigrants in Denmark is a very diverse group originating from a large number of countries. This means that some source countries have a very low number of immigrants in the sample. Therefore, including a dummy for each possible source country makes little sense. There are 13 source countries with more than 1000 individuals in the immigrant sample, and these countries cover 82.5 percent of the immigrant sample. Table 3 illustrates the distribution of immigrants across the 13 source countries in Denmark for the combined group of immigrants born in Denmark and abroad. Turkey is, by far, the largest source country, from where 23 percent of the immigrants originate. In addition, Pakistan, Lebanon, former Yugoslavia and Bosnia & Herzegovina is among the largest countries of origin, but less than ten percent of the immigrant sample originate from each of these countries. The 13 largest countries each get their own country of origin dummy in the empirical analysis, and the rest of the countries are collected in a residual category called *other countries*. I do this, since the rest of the countries are such a diverse group that a meaningful division of them is hard to find.

In a separate analysis, the immigrant children born abroad are grouped based on arrival age. Age at arrival is determined by combining information on the child's immigration date and year of birth. Dummy variables are included for each possible arrival age from zero to 15, which is the maximum age at arrival.

---

[13] For immigrants born abroad it is defined as the country of birth and for immigrants born in Denmark as the country of citizenship

Table 3: Source country composition for the immigrant subsample

Variable	Mean	Std. Dev.
Turkey	0.231	(0.421)
Pakistan	0.082	(0.274)
Lebanon	0.074	(0.262)
Former Yugoslavia	0.069	(0.253)
Bosnia and Herzegovina	0.068	(0.252)
Iraq	0.056	(0.230)
Vietnam	0.048	(0.213)
Iran	0.045	(0.206)
Sri Lanka	0.037	(0.189)
Morocco	0.032	(0.177)
Somalia	0.031	(0.173)
Afghanistan	0.027	(0.161)
South Korea	0.026	(0.159)
Other countries	0.175	(0.380)
Number of observations		55,455

NOTE: Administrative register information from Statistics Denmark for sample of non-Western immigrants graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).

In parts of the analysis, grades from compulsory school are included. The compulsory school grade register holds information on both continuous assessment grades (*standpunktskarakterer*) and exam grades from ninth grade. Some studies argue that the exam grades are the best measure of academic achievement, and especially the written exams, since they are identical across schools making them easy comparable (Nielsen & Rangvid, 2012). Furthermore, both a teacher and an external examiner grade the exams, making it less likely that the grades are affected by the teacher's relationship or attitude towards the student. However, in Denmark the nine years of schooling is compulsory, but to complete the final exams in ninth grade was not mandatory until 2007 (Danish Economic Councils, 2007). Therefore, it seems natural that the weaker performing students and the students with less support from home would be more inclined to drop the exams, and hence, that only including individuals with exam grades would systematically exclude some of the weaker students. To examine if this is the case, grade point averages are computed and compared for the individuals with and without exam grades, which is depicted in table 4. Grade information is available for 51,360 students in total and out of these, 756 students only have data on continuous assessment grades. Therefore, the grade point averages depicted in table 4 is based on exam grades for the 50,604 individuals with exam grades and on continuous assessment grades for the 756 individuals with missing exam grade information. It is here evident that the individuals without exam grade data, on average, are significantly lower performing than the rest of the sample. This confirms the suspicion that only including individuals with exam grades tend to exclude some of the weaker students and therefore, the 756 individuals will also be included in the grade analysis, but their grade point average will be based

on their continuous assessment grades[14]. The point of including grades is to control for individual abilities, and therefore all individuals with some information on abilities is included. The grades are measured on a seven-point scale, where the possible grades are -02, 00, 02, 4, 7, 10 and 12. The lowest passing grade is 02. Six different dummy categories are made representing the different grade levels and these can be seen in table 6 in section 7, where the descriptive statistics for the grade cohort is elaborated on.

Table 4: Grade point averages for individuals with and without exam grades available

	Mean	Std. Dev.	Observations
Individuals with exam grade data			
Grade point average	5.459	2.188	50,604
Individuals without exam grade data			
Grade point average	2.233	2.069	756

NOTE: Grade point average is measured on the basis of exam grades in ninth grade for individual's with exam grades and on continuous assessment grades in ninth grade for individual's without exam grades. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2).

6.4.2 Family characteristics

Family characteristics are divided into two groups of variables: family structure and family background variables. Family structure variables contain dummies for mother and father's age at childbirth[15], whether the parents are divorced, and how many siblings[16] under the age of 18, the child has. The parents are defined as divorced if the marital status of at least one of them is stated as divorced. Number of siblings are determined from information on number of children for the mother and father. If the number differs for the two, the largest number is used. As family background variables, parents' years of education and socioeconomic status is controlled for. In empirical work, it differs whether the education and socioeconomic status of both parents are included separately, if only one parent is included, or if a combined measure is made based on the highest attainment of the mother and father. Jakobsen & Smith (2006) argue that the education of the mother and father is highly correlated and find that when both are included, either one of them or both become insignificant. They therefore choose to only include mother's education. However, many include both parents separately allowing them to affect the child in different ways (Deding & Hussain 2005, Van Ours & Veenman 2006, Riphahn 2003). For instance, Deding & Hussain (2005) find that a better-educated mother is more important than a better-educated father for children's educational attainment in Denmark. In this analysis, education and socioeconomic status of the parents will therefore be included separately for the mother

---

[14] It might seem inconsistent that the grade point is not measured in the same way for all the students. It is however for 98.5 percent of the students and the only other solution would be to exclude the 756 students from this part of the analysis, which seem like a worse decision since the data suggest that these students in general are lower performing than the average.

[15] Dummies for mother and father's age at childbirth is based on the following intervals: under 18, 18-22, 23-27, 28-32, 33-37, 38-42 and over 42 years old

[16] Dummies for the number of siblings are based on the following categories: no siblings, 1 sibling, 2 sibling, 3 siblings, 4 or more siblings

and the father enabling them to affect the children differently. Parents' education is denoted as the standard time limit of their highest completed education. The educational attainment of the parents enter as dummies based on a division of three levels: Low, middle and high educational level. The parents with the lowest educational level have at most completed primary school (0-9 years of education). Attaining the middle level corresponds to completing upper secondary school (10-12 years of education) and parents in the highest educational category have studied for more than 13 years and obtained a higher education. The socioeconomic status of the parents reflects the parents' connection to the labour market[17], and the overall division is employed, unemployed or outside the labour force. The group of employed are furthermore divided into seven subgroups depending on their job function, and whether they are employed or self-employed[18]. The seven subgroups can be seen in table 5 in section 7, which holds the descriptive statistics. For the group outside the labour force, a separate category is made for those who are enrolled in education. Both parents' years of education and socioeconomic status as well as the marital status of the parents and the number of siblings may change through the child's life and is therefore measured when the child is 15, which was elaborated on in section 5.1. As well as age at arrival, parents' years since migration is controlled for. Again, mother and father's years since migration are included separately, since they not necessarily arrive at the same time and affects the child in the same way. Years since migration is measured when the child is 15 and is based on the parents' immigration dates, meaning that a mother with 15 years since migration corresponds to a mother arriving in Denmark at the child's year of birth.

6.5 Missing indicators

An advantage of using administrative data from statistical registers is that individuals only leave the data set if they die or emigrate. This means that there is not the same concern with missing observations as when using survey data, where people can choose not to answer particular questions. If people with specific characteristics refrain from answering, results from estimations based on the survey may be biased. In this analysis, the problem with missing observations primarily concerns the parent variables, which may be missing either because the parent is dead, does not live in Denmark or because there is no data available for the parent on the variable of interest. The latter is often a problem for education of immigrant parents. If they have not acquired an education in Denmark, it most likely will be missing, since the educational institution register only contains information on education attained in Denmark, and therefore education attained prior to immigration comes from a survey-based register. It obviously does not make sense to exclude all children where one or more of the variables are missing, especially, since the problem is larger for the immigrant children than the native children. Therefore, a missing indicator is included as a separate category for each of the variables. This means that if,

---

[17] The parents' connection to the labour market is measured in November the year before.

[18] In 1996, the subdivision of employees in the administrative registers change from being based on whether they were officials (funktionær), skilled or unskilled workers to being based on job function. Definitions for each group used in this analysis can be found in appendix A.

for instance, mother's education is missing for an individual, this individual will have a zero in all the education categories, and a one in the missing category. This is a solution, which is used in several of the empirical analyses (see for instance Riphahn 2003 and Colding et al., 2005). It does, however, not mean that missing observations are no longer a problem, and one has to be aware of the measurement error the missing observations impose on particularly parents' education.

The construction of the data has now been explained, and the next section will show, what the raw data reveals about the size of and development in the native-immigrant education gap. Furthermore, it will include tables of descriptive statistics, and I will discuss, what the descriptive statistics show about the average characteristics for the native and immigrant subsamples, respectively.

7 Descriptive analysis

One of the purposes of this paper is to determine the size of the native-immigrant education gap and examine how it has evolved across time. This is illustrated in figure 2, where the upper secondary completion rate for each of the cohorts graduating from ninth grade between 1990 and 2007 is plotted separately for native children, immigrant background children born in Denmark and abroad. The upper secondary completion rate for each cohort reflect the fraction of a graduating cohort that completes upper secondary education within five years. Figure 2 supports what was expected and found by previous Danish studies, namely, that the immigrant children, on average, have a lower educational attainment than the native children. It is here evident that the native-immigrant education gap exists in Denmark. The development in the gap is, however, positive and figure 2 further illustrates that the educational attainment of the immigrant children both born in Denmark and abroad is approaching that of the natives.

The completion rate for the native children has been relatively stable across time with a little rise in the later years. If a similar analysis is made some years from now, it will be interesting to see whether there is an apparent improvement in the upper secondary rate after 2006 suggesting that the initiatives regarding upper secondary school in The Welfare Agreement have paid off. For the children with immigrant background the rise in the completion rate is more evident, and especially the children born abroad have experienced a large increase in the completion rate, which also originated from a very low level. In the early 1990s, only 30 percent of the foreign-born immigrant children in a ninth grade graduating class managed to complete upper secondary school within 5 years. In 2007, this fraction has increased to almost 55 percent. As expected, the immigrant children born in Denmark have a completion rate somewhere between that of the natives and the immigrants born abroad. The positive development in the completion rate for the immigrant children born in Denmark is not as evident as for the children born

abroad, but it is particularly clear in the later years that they are catching up with the natives. Out of the group of native children graduating from ninth grade in 2007, 70 percent manage to complete upper secondary education within five years. The same holds for 60 percent of the immigrants born in Denmark, and 55 percent of the immigrants born abroad. The corresponding figures for the cohort of 1990 are 64, 50 and 32 percent. This means that the native-immigrant education gap has been sharply reduced across time, and that the immigrants are approaching the natives. The question, however, remains whether the positive development is a consequence of changes in the immigrant population across time. It might be the case that the immigrants in the later graduating cohorts originate from less disadvantaged source countries or have more resourceful parents, and that this explains the improvement in the completion rate. This will be examined in the empirical analysis in section 8.

Figure 2: Development in upper secondary completion rate
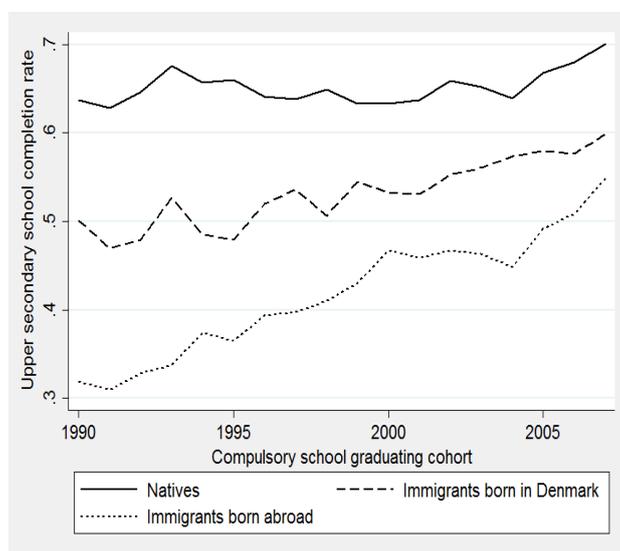
NOTE: Upper secondary school completion rate reflects the fraction of a graduating cohort who complete upper secondary school within five years of graduating from compulsory school (ninth grade). Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).

In the Danish media, it has been discussed whether the immigrant girls born in Denmark in particular is catching up with the native girls (Mikkelsen, 2013). In this context, senior researcher Beatrice Schindler Rangvid has said that she finds it possible that the Danish-born immigrant girls may even outperform the native Danish girls. Oppositely, there might also be cultural gender differences implying that the immigrant girls experience larger disadvantages than the immigrant boys do (Rangvid, 2010). It is therefore interesting to examine if the development in the completion rate differs significantly for boys and girls.

Figure 3: Development in completion rate for boys   *Figure 4: Development in completion rate for girls*
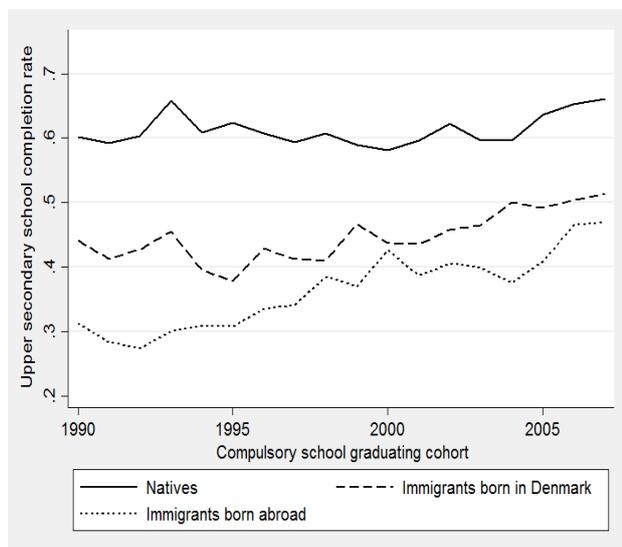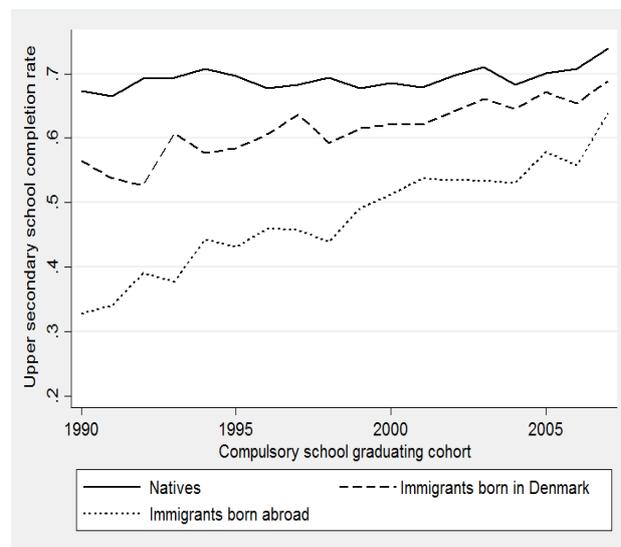
NOTE: Upper secondary school completion rate reflects the fraction of a graduating cohort who complete upper secondary school within five years of graduating from compulsory school (ninth grade). Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).

Figure 3 and 4 display the upper secondary completion rate separately for boys and girls, and it is here evident that there are large gender differences in the completion rate. Both the size of and development in the gap seem to differ considerably for boys and girls. A first thing to note is that girls have a substantially higher completion rate than boys in all three groups. The native girls have a very stable completion rate around 70 percent for the entire period, while the native boys' fluctuate around 60 percent. The main difference between figure 3 and 4 is, however, the size of and development in the native-immigrant gap. For both the boys and the girls with immigrant background, the completion rate increases across time but the rate by which it increases seem to be somewhat larger for the girls. Furthermore, the gap between native and immigrant boys remain large for the entire period, and even though, it is reduced across the period it is still of significant size in 2007. 66 percent of the native boys graduating from ninth grade in 2007 have completed upper secondary within five years after graduation, but this only holds for 51 percent of the Danish-born immigrant boys and 47 percent of the foreign-born immigrant boys.

A more positive picture emerges for the girls. In figure 4, it is evident that the native-immigrant education gap for the girls has been sharply reduced across time, and that the immigrant girls born in Denmark have a completion rate very close to the native girls. In 2007, the difference between the two groups is only 5 percentage points. The immigrant girls born abroad have a slightly lower completion rate, but it has heavily increased across time from 33 percent for the graduating cohorts in 1990 to 64 percent in 2007. The immigrant girls born abroad have a completion rate on the same level as the native boys in 2007, which is somewhat surprising and underscores the large gender differences. To sum up, separate examination of the boys and girls gives a much more nuanced picture of the attainment gap, and the raw data reveals that immigrant boys experience larger disadvantages than immigrant girls do.

It is further interesting to get an idea of why the immigrants have a lower educational attainment in the first place. This can be examined by comparing summary statistics for the native children and the immigrant children, and these are reported in table 5. The completion rates in table 5 are averaged across all cohorts and reflect the average of each line illustrated in figure 2. This results in an average completion rate of 65.2 percent for the natives, 54.8 percent for the Danish-born immigrants and 43.8 percent for the immigrants born abroad. Theory and empirical evidence suggest that disadvantaged family background is a key part of the explanation of the native-immigrant education gap. Therefore, table 5 includes descriptive statistics for the mother and father's education and socioeconomic status. It is seen that a larger fraction of the native children than immigrant children have parents in the highest educational category. 57 percent of the native children have a father with more than 13 years of education, which only holds for one third of the immigrant children. It is important to note that the immigrants suffer from a relatively large number of missing observations in this variable. However, since the immigrants originate from non-Western countries, it is probably more likely that the individuals with missing educational information belong to the lower educational levels. Examining the parents' socioeconomic status instead might give a clearer picture of the family background since the problem with missing information is a lot smaller here. It is here evident that the immigrants, in general, have a weaker socioeconomic background than the natives. Where more than 40 percent of the mothers of the native children are either managers or employees on the highest or medium level, the same holds for less than 10 percent of the mothers of the immigrant children. The same picture emerges when examining the fathers instead.

What is most striking is the large fraction of individuals outside the labour force in the immigrant subsample. Around 10 percent of the native children's mothers and fathers are outside the labour force, but this number is significantly higher for the immigrants. 35 percent of the mothers of immigrant children born in Denmark are outside the labour force and approximately 25 percent of the fathers. For the immigrants born abroad, the figures are even worse with almost 50 percent and 35 percent for the mothers and fathers, respectively. A part of the explanation may be that the immigrants have a hard time using their education acquired abroad in Denmark, but the numbers underscore the weaker family background for the immigrants. The questions, however, remains whether the education gap still exists after differences in family background have been controlled for, which will be examined in the empirical analysis.

Besides the family background variables depicted in table 5, the regressions also include controls for family structure variables such as parent's age at childbirth, number of siblings and if parents are divorced or not. The summary statistics for these variables are less informative and are therefore, due to space concerns, placed in appendix B together with summary statistics for mother's and father's years since migration and the child's age at arrival.

		Immigrant background	
Variable	Natives	Born in Denmark	Born abroad
Completed upper secondary education			
within 5 years of compulsory education	0.652	0.548	0.438
	(0.476)	(0.498)	(0.496)
Female	0.489	0.500	0.476
	(0.500)	(0.500)	(0.499)
Mother's education			
0-9 years	0.196	0.258	0.228
	(0.397)	(0.438)	(0.420)
10-12 years	0.274	0.192	0.206
	(0.446)	(0.394)	(0.404)
13+ years	0.510	0.232	0.342
	(0.500)	(0.422)	(0.474)
Missing	0.020	0.318	0.224
	(0.142)	(0.466)	(0.417)
Father's education			
0-9 years	0.198	0.287	0.160
	(0.398)	(0.452)	(0.366)
10-12 years	0.182	0.159	0.127
	(0.386)	(0.366)	(0.333)
13+ years	0.570	0.327	0.389
	(0.495)	(0.469)	(0.488)
Missing	0.050	0.226	0.324
	(0.219)	(0.419)	(0.468)
Mother's socioeconomic status			
Employed			
Managers and CEOs	0.008	0.002	0.001
	(0.091)	(0.044)	(0.031)
Employees on highest level	0.129	0.029	0.022
	(0.335)	(0.168)	(0.146)
Employees on medium level	0.275	0.050	0.033
	(0.446)	(0.219)	(0.179)
Employees on lowest level	0.215	0.115	0.072
	(0.411)	(0.319)	(0.258)
Unskilled worker	0.065	0.033	0.018
	(0.247)	(0.177)	(0.134)
Other employees	0.092	0.188	0.103
	(0.289)	(0.391)	(0.304)
Self-employed and assisting spouses	0.054	0.044	0.024
	(0.225)	(0.204)	(0.152)
Unemployed			
Unemployed	0.046	0.138	0.102
	(0.209)	(0.345)	(0.303)
Outside the labour force			

Enrolled in education	0.004	0.032	0.060
	(0.064)	(0.177)	(0.237)
Others outside the labour force	0.101	0.359	0.487
	(0.301)	(0.480)	(0.500)
Missing	0.010	0.010	0.078
	(0.101)	(0.101)	(0.268)
Father's socioeconomic status			
Employed			
Managers and CEOs	0.034	0.009	0.004
	(0.182)	(0.093)	(0.059)
Employees on highest level	0.180	0.037	0.028
	(0.384)	(0.189)	(0.165)
Employees on medium level	0.126	0.035	0.020
	(0.332)	(0.183)	(0.139)
Employees on lowest level	0.251	0.162	0.092
	(0.434)	(0.368)	(0.289)
Unskilled worker	0.046	0.040	0.028
	(0.209)	(0.197)	(0.166)
Other employees	0.112	0.181	0.091
	(0.315)	(0.385)	(0.288)
Self-employed and assisting spouses	0.117	0.133	0.058
	(0.321)	(0.339)	(0.233)
Unemployed			
Unemployed	0.035	0.110	0.100
	(0.183)	(0.313)	(0.300)
Outside the labour force			
Enrolled in education	0.001	0.009	0.030
	(0.035)	(0.092)	(0.171)
Others outside the labour force	0.062	0.243	0.344
	(0.242)	(0.429)	(0.475)
Missing	0.036	0.041	0.205
	(0.187)	(0.197)	(0.404)
Number of observations	88,915	24,005	31,450

NOTE: Standard deviations in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).

In parts of the analysis, grade point averages are included, and this gives the possibility to explore interesting connections. Unfortunately, grade information is only available for the cohorts graduating from compulsory school in 2002 and onwards. This means that the grade estimations only include the six cohorts graduating from compulsory school between 2002 and 2007 and only the individuals for whom, grade information is available. In the full sample, 56,251 of the children belong to the graduating cohorts of 2002 to 2007, and grade data is available for 51,360 children out of these, corresponding to 93 percent of the natives and 89 percent of the immigrants. The 51,360 children will for the remaining part of the paper be denoted as the children in the *grade cohorts*, and table 6 gives the upper secondary completion rate and the grade point

distributions for these children. Comparison of table 5 and 6 shows that the average completion rate is higher for all three groups, when only averaged across the latest cohorts, which is in line with what figure 2 illustrated. The completion rates might, however, be slightly overrated here compared to if they were averaged across the entire population of children graduating between 2002 and 2007 if grade information, in general, is available for the more talented students. I will examine this further in a sensitivity analysis in section 9.

Table 6: Summary statistics for cohorts with grade information: Means (standard deviations in parentheses)

| | | Immigrant background | |
Variable	Natives	Born in DK	Born abroad
Completed upper secondary education			
within 5 years of compulsory education	0.697	0.611	0.549
	(0.460)	(0.488)	(0.498)
Female	0.500	0.502	0.488
	(0.500)	(0.500)	(0.500)
Grade point average			
Below 02	0.020	0.095	0.120
	(0.140)	(0.293)	(0.325)
Between 02 and 4	0.144	0.273	0.304
	(0.351)	(0.445)	(0.460)
Between 4 and 6	0.299	0.311	0.302
	(0.458)	(0.463)	(0.459)
Between 6 and 8	0.354	0.237	0.204
	(0.478)	(0.425)	(0.403)
Between 8 and 10	0.162	0.078	0.063
	(0.369)	(0.268)	(0.243)
Between 10 and 12	0.020	0.006	0.006
	(0.140)	(0.077)	(0.080)
Observations	27,192	12,151	12,017

NOTE: Standard deviations in parentheses. GPA is based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades (*standpunktskarakterer)* in ninth grade are available, and for these students GPA is measured based on these. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2).

By comparing the grade point averages in table 6, it is seen that the immigrant background children, in general, have a poorer performance in compulsory school than the native children. More than one out of three of the immigrant background children has a grade point average below 4, and the same holds for less than 17 percent of the native children. At the other end of the scale, more than half of the native children have a grade point average of 6 or more, but less than one third of the immigrant children have this. The descriptive statistics show that the immigrant children, on average, have a lower performance than the native children measured by grade point averages in compulsory school. In the empirical analysis, I will examine how much

of the native-immigrant gap in upper secondary school that can be explained by differences in performance in compulsory school.

Examination of the raw data shows that a native-immigrant attainment gap exists in Denmark. However, the size of the gap has been reduced across time, and the immigrant girls are, in particular, catching up with the native girls. In the empirical analysis in next section, it will be examined whether these differences are statistical significant. Furthermore, the data revealed that the immigrant children have less favourable socioeconomic background and a lower performance in compulsory school, and it will be examined how much of the gap in upper secondary school that can be explained by these differences.

8   Empirical analysis

This section contains analysis and interpretation of the estimation results. Due to space concerns, the tables only include the estimated coefficients for the most important variables, but the full estimation results are reported in appendix C to J. In the first subsection, one of the goals is to find the sources of the native-immigrant gap, and therefore, the estimations are based on different specifications with different control variables to examine how each group of variables help explain the gap. Additionally, I examine whether the gender differences found in the descriptive analysis are statistically significant. In the second subsection, the focus is on grades, and how much of the gap can be explained by differences in grade point averages in compulsory school. Finally, the last subanalysis focuses on the immigrant children born abroad, and how their educational disadvantage is related to the age at which they arrive in Denmark.

8.1   Baseline results

To get an idea about the size of the raw native-immigrant education gap, first, a simple model is estimated. I report the results from estimation of the simplest version of equation (2) in table 7, where no parental variables are controlled for, and besides the four variables of interest, only gender and cohort dummies are included. The results from this simple estimation can be compared with the results from estimations with more control variables to examine how large a fraction of the gap each control variable accounts for. Note, that this is the only estimation where the average differential between native children and immigrant children is estimated directly instead of computed, which was elaborated on in section 5.1. As mentioned, the coefficient on *immigrant background* gives the average difference in the completion rate between native children and immigrant children born in Denmark, and the coefficient on *immigrant background · born abroad* gives the difference between immigrant children born in Denmark and abroad. Hence, the difference between the natives and the foreign-born immigrants corresponds to the sum of these two coefficients. The average annual change in the effect of having immigrant background and being born abroad is reflected by the coefficients on the trend terms, and they

therefore, indicate whether the immigrants are catching up with the natives or not. The dependent variable in all estimations is *completion of upper secondary education within five years of compulsory school* meaning that each estimated coefficient reflects the marginal effect that control variable has on the probability of completing upper secondary school within five years of compulsory school.

Table 7: OLS estimation results for a simple model

Dependent variable: Completion of upper secondary education within five years of compulsory school	
	(1)
Immigrant status	
(Native children as reference category)	
Immigrant background	-0.169***
	(0.009)
Immigrant background · born abroad	-0.163***
	(0.011)
Immigrant background · trend	0.005***
	(0.001)
Immigrant background · born abroad · trend	0.006***
	(0.001)
Female	0.104***
	(0.003)
Controls	
Cohort	Yes
Observations	144,370
Adjusted R-squared	0.048

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2). The regressions use the linear probability model. The controls include 18 cohort dummies and a constant is included but unreported. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The results in table 7 show that having immigrant background has a large and significant negative effect on the completion rate, which is as expected and found in the descriptive analysis. The estimated average difference in the completion rate between native children and immigrant background children born in Denmark is almost 17 percentage points and for the foreign-born children the difference compared to the natives is as big as 33.2 percentage points (0.169+0.163). However, when examining the estimated trend effects, they are positive and significant for both immigrant children born in Denmark and abroad. The Danish-born immigrants approach the completion rate of the natives with 0.5 percentage points per year, and for the immigrants born abroad, the catching up rate is 1.1 percentage points (0.005+0.006). That is, the results show that there is a sizeable gap between native children and immigrant children born in Denmark, and an even bigger between native children and immigrant children born abroad. However, at the same time the trend terms show that both immigrant groups are catching up with the natives. This is in line with the descriptive analysis and means that the positive development illustrated in figure 2 is statistical significant. It is now interesting to examine if the gap will be reduced

when a number of control variables are included, and thereby, to see how they help explain the gap. Furthermore, it is interesting whether the positive and significant catching up rates persist after controlling for an extensive number of individual and parental characteristics. Therefore, table 8 contains results from six different estimations, where a number of controls are included stepwise to examine each group of variables' effect on the four parameters of interest. The full estimation results are reported in appendix C. For comparison in column (1), I include the results from the simple model. Results from estimations where country of origin and family structure variables are accounted for are reported in column (2) and (3), respectively. Family background variables are stepwise controlled for. In column (4), I report the results where parents' socioeconomic status are accounted for, and in column (5), they are replaced with parents' years of education. Finally, in column (6), the results are based on a model with all the above-mentioned controls, and where socioeconomic status and years of education enter simultaneously.

Table 8: OLS estimation results after controlling for different groups of control variables

Dependent variable: Completion of upper secondary education within five years of compulsory school						
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant status						
(Native children as reference category)						
Immigrant background	-0.169***	-0.138***	-0.133***	0.009**	-0.078***	0.009**
	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Immigrant background · born abroad	-0.163***	-0.190***	-0.166***	-0.133***	-0.183***	-0.147***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Immigrant background · trend	0.005***	0.004***	0.006***	0.003***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend	0.006***	0.004***	0.004***	0.003***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.104***	0.103***	0.105***	0.106***	0.105***	0.106***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Controls						
Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin	No	Yes	Yes	Yes	Yes	Yes
Family structure	No	No	Yes	Yes	Yes	Yes
Mother and father's socioeconomic status	No	No	No	Yes	No	Yes
Mother and father's education	No	No	No	No	Yes	Yes
Observations	144,370	144,370	144,370	144,370	144,370	144,370
Adjusted R-squared	0.048	0.063	0.090	0.131	0.121	0.141

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2). The regressions use the linear probability model. Specification (1) follows from table 7. In specification (2) to (6), the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. The controls include 18 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 10 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The results in column (2) show that when country of origin is accounted for, the Danish-born children of immigrants perform somewhat better than indicated by the results in the simple model in column (1). The performance of the immigrant children born abroad is, on the other hand, practically unchanged, since the estimated difference in their completion rate compared

to the natives now is -32.8 percentage points compared to -33.2 percentage points in column (1). This means that even though there might be large differences among immigrants across source countries in Denmark, accounting for them does not explain much of the disadvantage of having immigrant background and being born abroad. The catching up rates are slightly reduced but still significant, when source countries are included. This suggests that a secular change in the composition of source countries for the immigrant population explains some of the positive development in the educational attainment of immigrants over time. The trend terms do, however, remain positive and of significant size after country of origin have been accounted for.

By comparing the results in column (2) and (3), it is seen that differences in family structures do not explain much of the native-immigrant education gap. However, another picture emerges when parents' socioeconomic status is included in the estimations. This heavily affects the average effect of having immigrant background. Surprisingly, the disadvantage of approximately 13 percentage points in columns (2) and (3) changes to a small, but significant, advantage of 0.9 percentage points, which is seen both in column (4) and (6). Including parents' years of education instead of socioeconomic status does not have as large an effect but still significantly reduces the disadvantage of having immigrant background. Under the assumption that the model is correctly specified, this means that, if the children with immigrant background had parents with the same socioeconomic status as the native children, they would, on average, do as well as the native children in upper secondary school.

In table 5, the summary statistics revealed large average differences between the parental background variables of the immigrant and native children, and the results now confirm that these differences explain the entire gap. However, this is under the assumption that the model is correctly specified, and it is therefore evident to examine if this is actually the case. Some of the empirical evidence showed that the parental background variables were of less significance for the immigrants compared to the natives. A possible explanation is that there is a problem with measurement error in the immigrant parents' education, since education not attained in Denmark only is based on a survey and is unknown for a large fraction of the immigrant parents. Furthermore, education from other countries can be less useful in Denmark, making it harder for the immigrant parents to exploit their full potential. Therefore, the correlation between the parental background and the children's achievement might be lower for immigrant children than for native children. Furthermore, the children of immigrants from non-Western countries probably experience a higher quality of education in the host country compared to what their parents experienced in the source country. If school quality affects children's achievement, this is a possible explanation as well. Regardless of the explanation, it is important to check if the link between parental resources and children's achievement, in this analysis, also is weaker for immigrants than for natives. If this is the case, it calls for a modification of the estimation specifications.
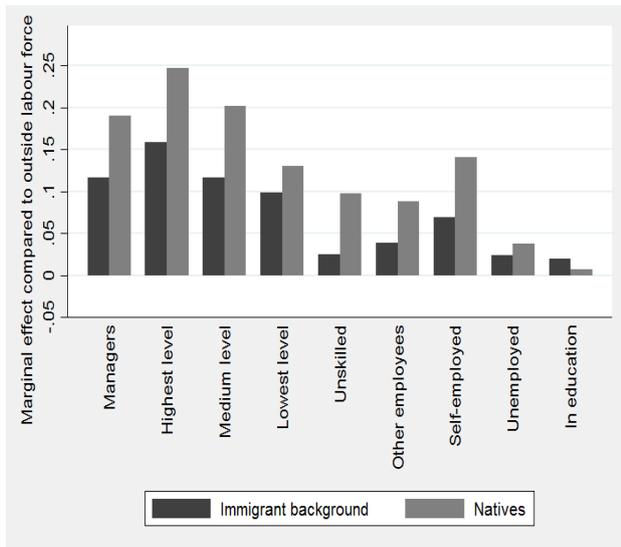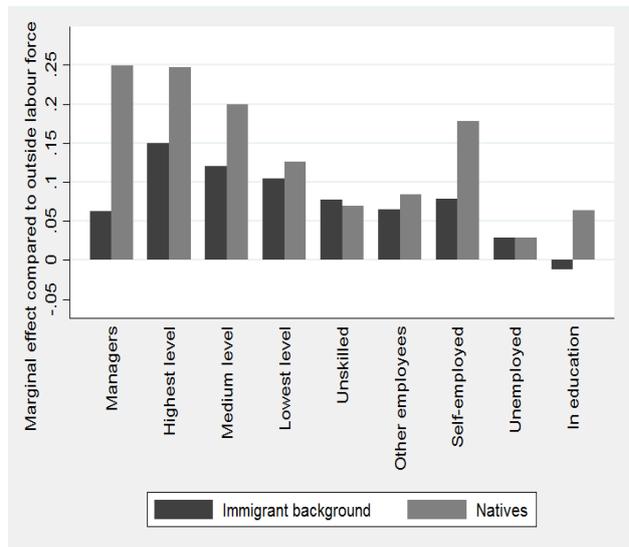
NOTE: Control variables in native subsample is a gender dummy, 18 cohort dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings and a divorce dummy. Additional controls for immigrant subsample are 14 country of origin dummies, born abroad dummy and trend. Reference category is parent outside the labour force for other reasons than because he/she is enrolled in education. Regressions are estimated separately for the native and immigrant subsample, but mother and father's socioeconomic status enter simultaneously. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).
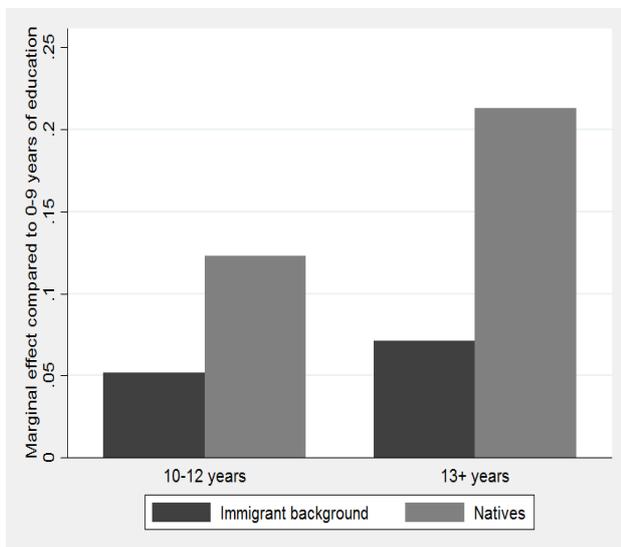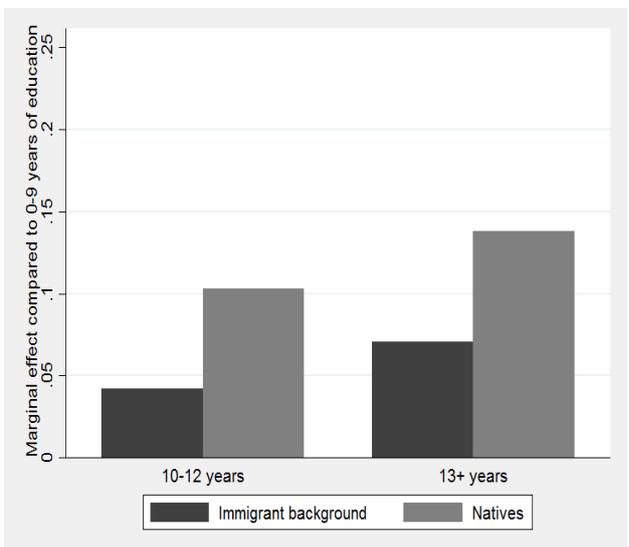
NOTE: Control variables in native subsample is a gender dummy, 18 cohort dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings and a divorce dummy. Additional controls for immigrant subsample are 14 country of origin dummies, born abroad dummy and trend. Reference category is parent with 0-9 years of education. Regressions are estimated separately for the native and immigrant subsample, but mother and father's years of education enter simultaneously. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).

To examine whether mother and father's socioeconomic status and educational background affect immigrant background and native children differently, the regressions are estimated separately for the immigrant and native subsamples. The estimated coefficients for parents' socioeconomic status and education for the two subsamples are illustrated in figure 5 to 8. Figure 5 and 6 depict the marginal effects of having a mother/father in the different socioeconomic status categories compared to having a mother/father outside the labour force. The bars illustrate the marginal effects for natives and immigrant background children, respectively and the full estimation results are reported in appendix D. All the marginal effects are significant at a 5 percent level for both natives and immigrants except the marginal effect of being enrolled in education.

Figure 5 clearly illustrates that the socioeconomic status of the mother affects the native children and immigrant children differently. The effects are stronger and, in some cases, numerically much stronger for the native children. Having a mother on the highest employment level, on average, increases the completion rate by 25 percentage points for the native children, where the effect for a child with immigrant background is 15 percentage points. For the socioeconomic status of the father depicted in figure 6, the differences between native children and immigrant children are not as evident for all the categories. The effect of having a father on the lowest level, unskilled or unemployed does not differ much for the immigrants and natives, but for the higher socioeconomic status levels the differences are numerically large and clear. These results show that the relationship between parents' socioeconomic status and the children's academic attainment is stronger for native children than for immigrant children.

An examination of the marginal effects of the parents' education shows that the same holds here. Figure 7 and 8 illustrate the marginal effects of having a mother/father with 10 to 12, or 13 or more years of education compared to having a mother/father with 0 to 9 years of education. All the effects are significant at a 1 percent level, and the full estimation results are reported in appendix D. Here the same picture, as with the socioeconomic status, emerges. The effects of having higher educated parents are larger for the native children than for the immigrant children and again the difference between the effects is larger for the mother than the father. That is, figure 5 to 8 reveal that parental background variables are less important for the immigrant children's educational attainment than for the native children's, which is in line with the empirical findings by Jakobsen & Smith (2006) and Colding et al. (2009) among others. This supports the hypothesis that a model, which restricts parental background variables to affect native children and immigrant background children similarly, is misspecified. It is therefore interesting to examine if parental background variables still explain the entire educational gap, when the effects from them is allowed to vary by immigrant background. To examine this, I run regressions, where parents' socioeconomic status and years of education both enter on their own and interacted with *immigrant background*. I report the results in table 9 column (2), (4) and (6). For comparison, I include the estimation results from the specifications without interactions between parental background and immigrant background in column (1), (3) and (5).

Dependent variable: Completion of upper secondary education within five years of compulsory school						
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant status						
(Native children as reference category)						
Immigrant background	0.009**	-0.029***	-0.078***	-0.115***	0.009**	-0.029***
	(0.004)	(0.008)	(0.004)	(0.008)	(0.004)	(0.011)
Immigrant background · born abroad	-0.133***	-0.136***	-0.183***	-0.180***	-0.147***	-0.145***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Immigrant background · trend	0.003***	0.003***	0.006***	0.007***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend	0.003***	0.003***	0.005***	0.004***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.106***	0.107***	0.105***	0.106***	0.106***	0.107***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Controls						
Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin	Yes	Yes	Yes	Yes	Yes	Yes
Family structure	Yes	Yes	Yes	Yes	Yes	Yes
Mother and father's socioeconomic status	Yes	Yes	No	No	Yes	Yes
- plus interaction with immigrant background	No	Yes	No	No	No	Yes
Mother and father's education	No	No	Yes	Yes	Yes	Yes
- plus interaction with immigrant background	No	No	No	Yes	No	Yes
Observations	144,370	144,370	144,370	144,370	144,370	144,370
Adjusted R-squared	0.131	0.132	0.121	0.125	0.141	0.143

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2). The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, in specification (2), (4) and (6) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The controls include 18 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 10 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Pairwise comparisons of the results in table 9 show that for all specifications, the disadvantage of having immigrant background becomes larger, when parental background variables are allowed to vary by immigrant status. This suggests that the effects of parental background is over-controlled for in the specifications without interactions. Furthermore, comparison of column (1) and (2), or (5) and (6) show that the small advantage of the immigrants when socioeconomic status was controlled for now is turned to a disadvantage again. That the weaker family background of immigrants explains the entire educational gap was based on a misspecified model. The difference in parental resources for natives and immigrant children do, however, still account for a major part of the gap for the Danish-born immigrants. In a model controlling for gender, cohorts, country of origin and family structures the native-immigrant differential was 13.3 percentage points (table 8 column (3)), but when family background variables additionally were included and entered in the right way, the difference was reduced to 3 percentage points

(table 9 column (6)). For the immigrant children born abroad, differences in family background also explain some of the gap, but a substantial part remains unexplained. From column (6) it is seen that the foreign-born children still experience a remaining disadvantage of 17.4 percentage points (0.029+0.145). That is, for the foreign-born children a large average difference in the completion rate compared to the natives remains even after controlling for an extensive range of family related control variables. This calls for a further examination of the children born abroad. The effects of being born abroad is in table 6 to 8 averaged across all the foreign-born children, but they differ in one important manner, namely, in the years they have been in Denmark. There must be a natural link between the child's length of stay in Denmark and his educational attainment. Therefore, later analysis will have particular focus on the immigrant children born abroad and the importance of arrival age.

Another important result to notice from table 8 and 9 is that the estimated trend effects remain significant for all the different specifications. The catching up rate for the Danish-born children with immigrant background varies between 0.3 and 0.6 percentage points, and it is slightly larger for the immigrant children born abroad varying between 0.6 to 1.1 percentage points depending on the specification. Based on the results in column (6) in table 9, which is the model with the most controls, the average disadvantage of having immigrant background is reduced by 0.4 percentage points per year, and the completion rate of the immigrant children born abroad catches up with that of the natives by 0.7 percentage points per year (0.004+0.003). Based on the figures in the descriptive analysis, it was expected that the trends would be positive. The figures did, however, not reveal anything about the reasons for the development. If the positive development solely was a consequence of secular changes in the source country composition of immigrants, the trend terms would lose their significance, when country of origin dummies were included. Furthermore, if the positive trends were a result of immigrants in the later cohorts having more resourceful parents, the significance would vanish once parental background variables have been controlled for. Since the catching up rates remain positive and significant, none of the above can explain the positive development suggesting that the positive trend is sign of improvement in the educational attainment of the immigrant population across cohorts.

That the catching up rate is larger for the foreign-born immigrants than the Danish-born is as expected from figure 2 but somewhat counterintuitive. The immigrants born abroad, in general, have a lower educational attainment, and it seems natural that they would approach the natives at a slower pace than the Danish-born immigrants. The higher catching up rate for the immigrants born abroad could be a consequence of changes in the arrival age composition across cohorts. If immigrants arriving at a higher age have lower attainment, and there are fewer *late arrivers* in the latest graduating cohort, this could explain the development. If this is the case, the estimated coefficient on the born abroad trend will lose its significance once age at arrival have been accounted for, and this will be examined in section 8.3.

In this section, I estimated models where parents' socioeconomic status and years of education enter both on their own and simultaneously. The marginal effects from these variables are, due to space concerns, not included in table 9 but is reported in appendix E. The results there show that both education and socioeconomic status have significant effects on the child's educational attainment when entered separately as in the models behind the results in column (2) and (4). Some natural correlation must exist between a parent's years of education and socioeconomic status, and therefore, some of the effects might become insignificant, when both education and socioeconomic status are included simultaneously. However, the estimated coefficients of both education and socioeconomic background remain significant, when they enter simultaneously, which is the case in the model behind the results in column (6), and therefore, this is found to be the best model. In the remaining part of the analysis, this model, where both socioeconomic status and education of the parents is included and the effect from them is allowed to vary by immigrant background, will be referred to as *the full model*.

Several analyses focus on comparing the educational attainment of native children and immigrant children born in Denmark, since both groups have attained all of their education in Denmark and therefore should be easily comparable. The Danish-born immigrant children are, however, also a heterogeneous group, and they particularly differ in the length of time their parents' have spent in Denmark. Their parents' length of stay in Denmark may affect their integration potential, and thereby, their educational attainment, and this is the focus in a recent Danish paper by Nielsen & Rangvid (2012). Therefore, the full model is re-estimated with the inclusion of dummies for parents' years since migration to see how this affects the results in my analysis. The effects are, however, insignificant for almost all categories[19], and the results are therefore not included, but is reported in the last column in appendix E. In the analysis by Nielsen & Rangvid (2012), they find the significance of mother and father's years since migration to heavily depend on the estimation specification. My full model specification includes controls for both parents' education and socioeconomic status, and Nielsen & Rangvid (2012) find that the effects of parents' years since migration, in general, is reduced when parents' education and socioeconomic status are accounted for. The insignificance of parents' years since migration in the estimation based on the full model is, therefore, not that surprising.

So far, the empirical analysis has not included comments regarding the gender difference in the completion rate. The results in table 7 to 9 show that girls have an educational advantage over the boys that is very stable and unaffected by the inclusion of more variables. Girls, on average, have a 10 percentage points higher completion rate than boys. Whether the higher completion rate for the girls is due to the girls getting higher grades in compulsory school will be examined in next section. The gender difference is only accounted for by including a gender dummy without any interactions. This means that the difference in the completion rate between boys and

---

[19] Changing the reference category for parents' years since migration makes them more significant but then a number of the source country dummies become insignificant. There therefore must be a high degree of correlation between the two groups of variables, which further supports the decision to leave them out of the analysis.

girls is not allowed to vary by immigrant background. In the descriptive analysis, the development in the completion rate across cohorts was illustrated both on average and separately for boys and girls, and the figures showed that the development in the gap varied substantially by gender. It was here evident that the girls, on average, have a higher completion rate than the boys, and this is already captured in the models by the inclusion of a gender dummy. The figures, however, also illustrated that the difference between native and immigrant children seemed to be larger among boys than girls. This motivates the use of gender interactions. By including interaction terms, I can examine whether the effect of having immigrant background varies significantly by gender. Therefore, I re-estimate the models, and the four variables of interest now enter both on their own and interacted with gender. This allows for different effects of having immigrant background and being born abroad for boys and girls as well as allowing for gender differences in the catching up rate for both immigrants born in Denmark and abroad.

I estimate two different models, where the four variables of interest is interacted with the gender dummy. First, a simple model with cohort and country of origin dummies as the only additional controls are estimated to examine if the size of the gap varies significantly by gender. The results from this estimation is reported in table 10 column (2). Secondly, I estimate the full model with gender interactions on the four variables of interest to examine if the gender differences persist after controlling for parental variables and the results are reported in column (4). In column (1) and (3), I for comparison report the results from similar specifications but without gender interactions. The full estimation results is reported in appendix F. Note that allowing for gender interactions changes the interpretation of the coefficients. In column (1) and (3) where there in the underlying model not is gender interactions on the four variables of interest, the estimated coefficient on *female* reflects the higher average completion rate for girls compared to boys, which holds for both native children and immigrant children. Furthermore, the effect of having immigrant background is not gender specific. This means that the difference in the completion rate between native girls and immigrant girls is the same as between native boys and immigrant boys and is found directly by the coefficient on *immigrant background* and *immigrant background · born abroad.* This is not the case in column (2) and (4). Here the coefficients on *immigrant background* and *immigrant background · born abroad* only relates to the boys and to get the marginal effects for the girls *immigrant background · female* and *immigrant background · born abroad · female* has to further be added. Likewise, the coefficient on *female* only reflects the gender difference in the completion rate for the native children, and to get the gender differences for the immigrant children the interaction terms has to be added, as well.

47

Dependent variable: Completion of upper secondary education within five years of compulsory school				
	(1)	(2)	(3)	(4)
Immigrant status				
(Native boys and girls as reference category)				
Immigrant background	-0.138***	-0.167***	-0.029***	-0.039***
	(0.004)	(0.005)	(0.011)	(0.014)
Immigrant background · female		0.058***		0.021
		(0.007)		(0.018)
Immigrant background · born abroad	-0.190***	-0.156***	-0.145***	-0.113***
	(0.011)	(0.015)	(0.011)	(0.016)
Immigrant background · born abroad · female		-0.069***		-0.066***
		(0.022)		(0.023)
Immigrant background · trend	0.004***	0.004***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · trend · female		0.002		0.003*
		(0.001)		(0.001)
Immigrant background · born abroad · trend	0.004***	0.002*	0.003***	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend · female		0.004**		0.003*
		(0.002)		(0.002)
Female	0.103***	0.079***	0.107***	0.084***
	(0.003)	(0.003)	(0.002)	(0.003)
Controls				
Cohort	Yes	Yes	Yes	Yes
Country of origin	Yes	Yes	Yes	Yes
Family structure	No	No	Yes	Yes
Mother and father's socioeconomic status	No	No	Yes	Yes
- plus interaction with immigrant background	No	No	Yes	Yes
Mother and father's education	No	No	Yes	Yes
- plus interaction with immigrant background	No	No	Yes	Yes
Observations	144,370	144,370	144,370	144,370
Adjusted R-squared	0.063	0.065	0.143	0.145

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2). The regressions use the linear probability model. Specification (1) and (3) follows from specification (2) in table 8 and specification (6) in table 9. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, in specification (3) and (4) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The coefficient on *immigrant background · female* in specification (2) and (4) is computed as the average of the coefficients on the source countries interacted with gender and weighted by the immigrant female population shares. In addition in specification (4) the coefficient on *immigrant background · female* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native female frequency distribution. The controls include 18 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 10 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The difference in the completion rate between native and immigrant background children is found to significantly differ with gender in the simple model in column (2) in table 10. In the model without gender interactions, the immigrants born in Denmark, on average, had a completion rate 13.8 percentage points lower than the natives, which is seen in column (1). This however transforms into a disadvantage of 16.7 percentage points for the boys and only 10.9 percentage points for the girls. This is consistent with the descriptive analysis, which showed larger differences in completion rates between native boys and immigrant boys than native girls and immigrant girls.

One of the most positive baseline findings was that the catching up rates remained positive and of significant size after the inclusion of an extensive number of control variables. The catching up rates are also interacted with the gender dummy here to examine whether the positive development, for instance, only applies to the girls. The results in column (4) show that the size of the catching up rates significantly differ with gender. In the model without gender interaction depicted in column (3), the gap between the native and the immigrant children born in Denmark was reduced by 0.4 percentage points per year. When the catching up rate is allowed to vary by gender, it is still significant for both genders, but it is twice as large for the girls than for the boys. This means that the immigrant girls born in Denmark approach the natives at a faster pace than the boys, when an extensive number of control variables have been accounted for. For the immigrant children born abroad, the additional catching up rate only applies to the girls, and the immigrant boys born abroad thereby has a catching up rate of the same size as the immigrant boys born in Denmark. That is, the educational attainment of the immigrant boys born in Denmark and abroad approaches that of the native boys by 0.3 percentage points per year. The immigrant girls born in Denmark approaches their native counterparts by 0.6 percentage points per year, and the catching up rate for the immigrant girls born abroad is as high as 0.9 percentage points. That is, the size of the native-immigrant gap significantly varies with gender but when differences in family characteristics have been accounted for some of the gender differences vanish. The gender differences in the catching up rates do, however, remain significant, suggesting that the immigrant girls are catching up with their native counterparts at a faster pace than the boys are. In next section, it will be examined if the gender differences persists after controlling for the student's own performance as measured by grade points.

To include the effects of the gender specific trend terms, table 11 holds the marginal effects for an *average immigrant* meaning an immigrant background child belonging to the graduating cohort of 1998. From the first column, it is seen that the average differential between native children and immigrant background children both born in Denmark and abroad is smaller for girls than for boys. The gender difference is, however, most evident for the immigrant children born in Denmark, since the difference for the children born abroad is 26.9 percentage points for the boys and 24.4 percentage points for the girls. The baseline analysis showed that a major part of the native-immigrant gap could be explained by differences in parental resources. From the second column in table 11 it is seen that parental differences basically explain the entire gap for

both immigrant boys and girls born in Denmark. The negative effect of being born abroad is, however, somewhat larger for girls than for boys meaning that differences in parental resources between natives and immigrants explain a larger part of the gap for the foreign-born boys than girls.

Table 11: Marginal effects of having immigrant background and being born abroad for an average immigrant

	Simple model	Full model
Immigrant background		
Boys	-0.131 (-0.167+0.004·9)	-0.012 (-0.039+0.003·9)
Girls	-0.073 (-0.167+0.058+0.004·9)	0.015 (-0.039+0.003·9+0.003·9)
Immigrant background and born abroad		
Boys	-0.269 (-0.167-0.156+0.004·9+0.002·9)	-0.125 (-0.039-0.113+0.003·9)
Girls	-0.244 (-0.167+0.058-0.156-0.069+0.004·9+0.002·9+0.004·9)	-0.137 (-0.039-0.113-0.066+0.003·9+0.003·9+0.003·9)

NOTE: Estimates follow directly from table 10. *Simple model* refers to the model in table 10 column (2) and *full model* to the model in column (4). Coefficients are only included if they are significant at the 10 percent level. *Average immigrant* refers to an immigrant belonging to the graduating cohort of 1998, which is why the trend terms are added and multiplied by nine.

To sum up, the results in this analysis showed that a substantial attainment gap in upper secondary school exists in Denmark, but importantly, the immigrants are catching up with the natives. Examining the boys and girls separately reveals that the gap is larger for the boys, and that they are catching of with the natives at a slower pace than the girls are. The results further show that differences in parental resources explain a major part of the native-immigrant education gap, especially, for the immigrant children born in Denmark. However, as discussed in section 5.2.2 one should be aware of the possible endogeneity, which may mean that the effects of parents' education and socioeconomic status are biased upwards. Therefore, it would be optimal to control separately for the child's own abilities, which is expected to reduce the large effects of parental background.

8.2   Results from grade estimations

One way to imperfectly control for the unobserved abilities of a child is to include grades from compulsory school. Grades are only available for the cohorts graduating after 2001. Therefore, this part of the analysis is only based on the cohorts graduating from ninth grade between 2002 and 2007, and the trend terms are omitted from the estimations because of the short observation period. The purpose of including grade point averages is to explicitly control for the child's own abilities. Until now, abilities have not entered as a control variable on its own. They were partly captured by the parental background coefficients, which therefore were biased upwards. By including grades from compulsory school, it furthermore becomes possible to examine whether the gap arises in compulsory or during upper secondary school. If the gap disappears, after controlling for grades from compulsory school, it indicates that the gap arises in compulsory school rather than in upper secondary school. Therefore, including grades from compulsory school in the regressions gives the opportunity to study several interesting relations. As before, the grades are both included in a simple model and the full model to see how this affects both

the average differential between the immigrants and natives, but also to examine how it affects the marginal effect of grades to include more control variables.

Table 12: OLS estimation results after controlling for grades

Dependent variable: Completion of upper secondary education within five years of compulsory school					
	(1)	(2)	(3)	(4)	(5)
Immigrant status					
(Native children as reference category)					
Immigrant background	-0.062***	0.050***	0.049***	0.037*	0.068***
	(0.004)	(0.003)	(0.012)	(0.018)	(0.016)
Immigrant background · born abroad	-0.113***	-0.031***	-0.030***	-0.078***	-0.024***
	(0.008)	(0.007)	-(0.007)	(0.008)	(0.007)
Female	0.104***	0.042***	0.042***	0.111***	0.050***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Grade point average					
(Below 02 as reference category)					
Between 02 and 4		0.178***	0.147***		0.176***
		(0.008)	(0.018)		(0.008)
Between 4 and 6		0.474***	0.402***		0.456***
		(0.008)	(0.018)		(0.008)
Between 6 and 8		0.720***	0.664***		0.681***
		(0.008)	(0.017)		(0.008)
Between 8 and 10		0.820***	0.777***		0.762***
		(0.008)	(0.017)		(0.008)
Between 10 and 12		0.814***	0.779***		0.744***
		(0.011)	(0.018)		(0.011)
Controls					
Cohort	Yes	Yes	Yes	Yes	Yes
Country of origin	Yes	Yes	Yes	Yes	Yes
Family structure	No	No	No	Yes	Yes
Mother and father's socioeconomic status	No	No	No	Yes	Yes
- plus interaction with immigrant background	No	No	No	Yes	Yes
Mother and father's education	No	No	No	Yes	Yes
- plus interaction with immigrant background	No	No	No	Yes	Yes
Grade point average interacted with immigrant background	No	No	Yes	No	No
Observations	51,360	51,360	51,360	51,360	51,360
Adjusted R-squared	0.046	0.284	0.285	0.114	0.299

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2). Grade information is available for 93 percent of the native subsample and 89 percent of the immigrant subsample. GPA is based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades (*standpunktskarakterer)* in ninth grade are available, and for these students GPA is measured based on these. The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. Additionally, in specification (3) the coefficient of *immigrant background* is averaged across six grade point average levels plus missing weighted by the native frequency distribution. In addition, in specification (4) and (5) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The controls include 6 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 9 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The results are reported in table 12 and based on different versions of equation (3). For comparison in column (1) and (4), I include results from a simple and the full baseline model. The results in column (2) are based on a simple model with gender, cohort, country of origin and grades as the only controls. By comparing the results in column (1) and (2), it can be seen whether the gap disappears when grades are controlled for, even without controlling for parental background differences. The results in column (3) are based on similar controls but here the effect of grades is allowed to vary by immigrant background. Column (5) holds the results based on a full model specification, which additionally controls for grades. The full estimation results for table 12 is found in appendix G.

Since the population now has changed to only include the latest cohorts and those with information on grades, the results from the estimations without grades in column (1) and (4) may differ some from what we found in last section, where it was based on the full population. From column (1), it is seen that the Danish-born immigrants have an average completion rate 6.2 percentage points lower than the natives. In comparison, the disadvantage of the immigrants was 13.8 percentage points in the same model based on the cohorts graduating between 1990 and 2007. The educational disadvantage for the immigrants born in Denmark is thereby found to be substantially smaller, when only the latest cohorts are included and the same holds for the immigrants born abroad. From column (4), it is furthermore seen that differences in family background between immigrants and natives explain the entire gap for immigrants born in Denmark and a major part of the gap for immigrants born abroad. The question, however, remains to which extent this is a result of higher average educational attainment of the immigrants in the latest cohort. It might be the case that grade data to a greater extent is available for the better students meaning that the disadvantage of the immigrants is underrated in this part of the analysis. I examine this in section 9, which contains a discussion of the possible implications of the results.

The purpose of this part of the analysis is to examine how the size of the gap changes, when grades from compulsory school is included. It is seen by comparing column (1) and (2) that even in a simple model without any family controls, the disadvantage of having immigrant background is completely eliminated and turned to a small, but significant, advantage when grades are included in the analysis. This means that immigrants would have a 5 percentage points higher upper secondary completion rate than natives, if they had the same grade point distribution in compulsory school. The baseline findings showed that, in particular, the immigrant children born abroad experienced a substantially lower completion rate than the native children. Even this disadvantage disappears and is turned to an advantage, when I condition on grades in compulsory school, since the difference in the completion rate for immigrants born abroad and natives now is 1.9 percentage points (0.050-0.031). This means that both immigrants born in Denmark and abroad would outperform the natives with respect to upper secondary attainment if they had the same grade point distribution in compulsory school as the natives. By examining the estimated effects of each grade level in column (2), it is seen that grades in

compulsory school have a huge impact on the upper secondary completion rate. As expected, the size of the effect increases as the grade point averages increase, and in all cases, the effect is numerically substantial. Having a grade point average between 6 and 8 compared to below 02, increases the children's probability of completing upper secondary school within five years of compulsory school by as much as 72 percentage points.

The results show that the gap already arises in compulsory school, and that the performance here is an important indicator of the later success in upper secondary school. However, this is under the assumption that the model is correctly specified and we saw in the baseline analysis that not allowing the effect of parental background to vary by immigrant status had a large impact on the effects of having immigrant background. Since the effect of parental background was found to vary by immigrant status, the same might be the case for the effects of grades. Therefore, to ensure that the change from a disadvantage to an advantage not is due to a model misspecification, separate regressions are estimated for the immigrant and native subsample, and the full estimation results can be found in appendix H.

Figure 9: Marginal effects of grade point averages
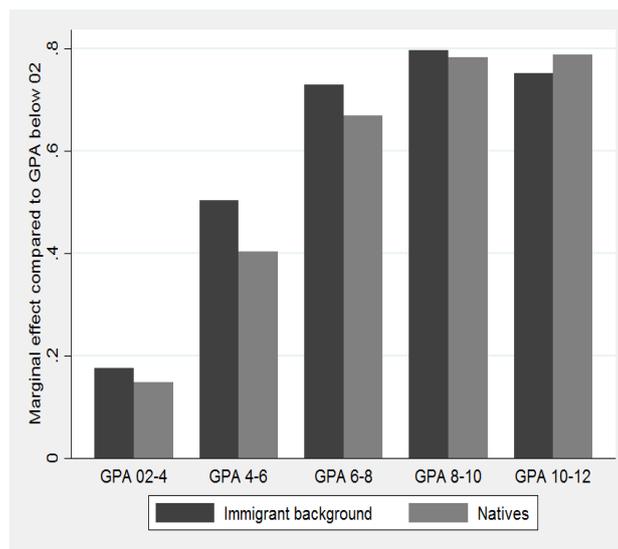
NOTE: GPA is grade point average based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades (*standpunktskarakterer)* are available, and for these students GPA are measured based on these. All marginal effects are significant at a 1 percent level. The controls include a gender dummy and 6 cohort dummies for the native subsample and additionally 14 country of origin dummies and a born abroad dummy for the immigrant subsample. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2).

In figure 9, the marginal effects of having a grade point average in the different intervals compared to having a grade point average below 02 is illustrated for the immigrants and natives,

respectively. The figure suggests that the relationship between grade point average in compulsory school and upper secondary completion is slightly stronger for the immigrants. However, it is particularly for the lower grades that the effect differs for immigrants and natives and the differences are not that clear for the higher grade levels. From figure 9, it is therefore not perfectly clear whether the right specification should allow for the effects of grades to vary by immigrant status or not. Therefore, I estimate a regression with interactions between immigrant background and grades to examine how this affects the other estimates. This is the model underlying the results in column (3) in table 12, and it is seen that changing the specification has basically no effect on the estimated differential between natives and immigrant background children. The full estimation results can be found in appendix G, and there it is further seen that the coefficients on the grade levels interacted with immigrant background are hardly significant and therefore an estimation without interactions with grades seems to be more appropriate.

Column (5) in table 12 represents the results from estimation of the full model, which additionally controls for grades. It is, however, the question whether it is the most appropriate model when grades are included as well, since this should reduce the effects of the parental background variables. When the full model depicted in column (5) is estimated separately for natives and immigrants[20], the parental background variables are found to have much less significance than it had when grades were not included, which can be seen in appendix G. In the immigrant subsample with grades there are hardly any significant effects from parental variables and for the natives there are some significant effects, but they are numerically very small. It is also seen by comparing the results in column (2) and (5) that even though specification (5) includes an extensive number of control variables compared to specification (2) the estimated effects of the main variables are not that different. However, the estimated grade effects are marginally smaller, which may be explained by the coefficients in column (2) also partly reflecting parental abilities, when these are not separately controlled for. The conclusion is that the gap arises in compulsory school, and performance in compulsory school is the main predictor of the later success in upper secondary school. The results show that, irrespective of the specification, immigrants have a slightly higher completion rate than natives with similar grade point average, and therefore, that the gap is completely explained by differences in grade point averages in compulsory school.

From table 12, it is further seen that the gender difference in the completion rate persists even after grade points have been controlled for. This means that the higher completion rate for the girls is not solely a consequence of the girls getting higher grades than the boys do in compulsory school. Girls, on average, have a completion rate 5 percentage points higher than boys with the same grade point average.

---

[20] The immigrant subsample includes country of origin dummies and born abroad dummy, which is not included in the native subsample.

Table 13: OLS estimation results after controlling for grades and gender interaction

Dependent variable: Completion of upper secondary education within five years of compulsory school		
	(1)	(2)
Immigrant status		
(Native boys and girls as reference category)		
Immigrant background	0.050***	0.015***
	(0.003)	(0.004)
Immigrant background · female		0.071***
		(0.006)
Immigrant background · born abroad	-0.031***	-0.034***
	(0.007)	(0.010)
Immigrant background · born abroad · female		0.006
		(0.014)
Female	0.042***	0.008
	(0.004)	(0.005)
Grade point average		
(Below 02 as reference category)		
Between 02 and 4	0.178***	0.175***
	(0.008)	(0.008)
Between 4 and 6	0.474***	0.469***
	(0.008)	(0.008)
Between 6 and 8	0.720***	0.716***
	(0.008)	(0.008)
Between 8 and 10	0.820***	0.818***
	(0.008)	(0.008)
Between 10 and 12	0.814***	0.813***
	(0.011)	(0.011)
Controls		
Cohort	Yes	Yes
Country of origin	Yes	Yes
Family structure	No	No
Mother and father's socioeconomic status	No	No
- plus interaction with immigrant background	No	No
Mother and father's education	No	No
- plus interaction with immigrant background	No	No
Grade point average interacted with immigrant background	No	No
Observations	51,360	51,360
Adjusted R-squared	0.285	0.285

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2). Grade information is available for 93 percent of the native subsample and 89 percent of the immigrant. GPA is based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades in ninth grade are available, and for these students GPA is measured based on these. The regressions use the linear probability model. Specification (1) follows from specification (2) in table 12. In specification (2), the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. The coefficient on *immigrant background · female* is computed as the average of the coefficients on the source countries interacted with gender and weighted by the immigrant female population shares. The controls include 6 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 9 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

It is interesting to examine whether the gender difference holds for both natives and immigrants when gender is interacted with having immigrant background and being born abroad, and grade differences are accounted for. In the baseline analysis it was found that the size of and the development in the native-immigrant gap was gender specific and persisted even when parental controls were included. The gender differences were though most evident in the catching up rates, which are left out of this analysis. Therefore, I now examine, what happens to the results when the effects of having immigrant background and being born abroad are allowed to differ by gender[21]. The results are reported in table 13 column (2). For comparison in column (1), I include the results from a model with similar controls but without the gender interactions. Full estimation results are found in appendix I.

One of the interesting things to note from column (2) is that when the gender effect is allowed to vary by immigrant background, the coefficient on *female* becomes insignificant. This means that conditional on grades, native girls do not have a higher completion rate than native boys do, and the superior performance of the girls were only related to the immigrant girls. Without gender interactions on having immigrant background and being born abroad, the immigrants born in Denmark had a 5 percentage points higher completion rate than the natives with the same grade point average, which is seen in column (1). When gender interactions are included this, however, transforms into an advantage primarily related to the immigrant girls. Immigrant girls born in Denmark have a completion rate 7 percentage points higher than their native counterparts. The advantage for the Danish-born immigrant boys over the native boys is only 1.5 percentage points. That is, by examining the gender differences in the native-immigrant gap, I find that native boys and girls have the same completion rate conditional on grades, and that it, in particular, is the immigrant girls born in Denmark, who have higher completion rate than their native counterparts. The overall finding of the grade regression is that performance in compulsory school is the main predictor of upper secondary school completion, and that the native-immigrant gap in upper secondary school can be entirely explained by native-immigrant differences in compulsory school grade point averages.

8.3 Results from age at arrival estimations

The baseline results showed that even after controlling for differences in family background variables, a sizeable attainment gap remains between the natives and the immigrants born abroad. The purpose of this subanalysis is to take a closer look at the immigrant children born abroad. This is done by replacing the *born abroad* dummy with dummies for each arrival age from zero to 15. I can, thereby, examine if children arriving at different ages have different educational attainments instead of evaluating the effect of being born abroad as an average across all arrival ages. It is particularly interesting to study if some critical arrival age can be establish, where children arriving after this age has an especially large disadvantage. Furthermore, the

---

[21] It is also examined whether the effect from grades vary with gender by including interaction gender with the grade levels. The interaction terms are however insignificant and the results are, therefore, not included in the paper.

results from the grade estimations showed that the large educational difference between children born abroad and native children vanished, when grades were accounted for and I will, therefore, examine how including grades changes the effects from the different arrival ages.

In table 14 column (1), I report the results from the full model estimation based on equation (4). The only change from the baseline model is that age at arrival dummies have replaced the born abroad dummy[22]. The estimated age coefficients give the additional effect of arriving in Denmark at the different ages compared to having immigrant background but being born in Denmark. The results in column (2) and (3) are based on the grade cohorts and thereby on equation (5). For comparison, I estimate a model with gender, cohort and country of origin as the only additional controls. The results of this estimation is reported in column (2), and column (3) holds the results of a similar estimation but where grades from compulsory school additionally are accounted for. The results in column (1) show that even after controlling for parental background differences, there is a negative effect of arriving in Denmark instead of being born in Denmark regardless of the age at arrival. The effects are significant for all possible arrival ages, but the size of the effects does differ substantially with the age. The marginal effects of each possible arrival age is furthermore plotted in figure 10 to illustrate how the effect varies with age and to more easily establish if a critical arrival age exists. By studying table 14 and figure 10, it is seen that the disadvantage of arriving at the age of one is practically the same as at the age of five. This suggests that as long as the child arrives before school age, they are not much affected by arriving some years after they are born. After school age, the disadvantage of arriving one year later increases. An immigrant arriving at the age of seven, on average, has a completion rate 10 percentage points lower than a Danish-born immigrant and the disadvantage increases to 13 and 15 percentage points if the child arrives at the age of eight or nine instead. The results suggest that childhood immigrants arriving after the age of seven, in particular, experience a substantial disadvantage, which increases with later arrival. A child immigrating at the age of 15 have a completion rate 28 percentage points lower than an immigrant child born in Denmark and 31 percentage points lower than a native child.

The catching up rate for the children with immigrant background has remained at 0.4 percentage points but in the baseline analysis, I found an additional catching up rate for the immigrant children born abroad, meaning that they were catching up with the natives at a higher rate than the immigrants born in Denmark were. When the arrival age dummies replace the born abroad dummy, the additional catching up rate for the immigrants born abroad disappears. This means that the higher catching up rate for the foreign-born immigrants was a consequence of a change in the distribution of the immigrant students across arrival age. The results in column (1) therefore show that the catching up rate, after arrival age is accounted for, is 0.4 percentage points for both the immigrants born in Denmark and abroad.

---

[22] Full estimation results for table 14 is found in appendix J

Table 14: OLS estimation results after controlling for age at arrival

Dependent variable: Completion of upper secondary education within five years of compulsory school			
	(1)	(2)	(3)
Age at arrival			
(Born in Denmark as reference category)			
0 years	-0.039**	0.004	0.048*
	(0.020)	(0.029)	(0.025)
1 years	-0.088***	-0.051***	-0.019
	(0.016)	(0.018)	(0.015)
2 years	-0.074***	-0.058***	-0.026
	(0.016)	(0.020)	(0.017)
3 years	-0.071***	-0.062***	-0.045***
	(0.017)	(0.021)	(0.018)
4 years	-0.091***	-0.092***	-0.059***
	(0.016)	(0.018)	(0.015)
5 years	-0.090***	-0.080***	-0.031**
	(0.016)	(0.017)	(0.015)
6 years	-0.118***	-0.133***	-0.063***
	(0.016)	(0.018)	(0.015)
7 years	-0.106***	-0.112***	-0.038**
	(0.016)	(0.018)	(0.015)
8 years	-0.131***	-0.142***	-0.061***
	(0.015)	(0.018)	(0.016)
9 years	-0.156***	-0.156***	-0.042***
	(0.015)	(0.018)	(0.015)
10 years	-0.151***	-0.159***	-0.021
	(0.015)	(0.018)	(0.016)
11 years	-0.183***	-0.200***	-0.026
	(0.015)	(0.019)	(0.017)
12 years	-0.175***	-0.185***	-0.006
	(0.015)	(0.019)	(0.017)
13 years	-0.194***	-0.175***	0.002
	(0.015)	(0.022)	(0.019)
14 years	-0.240***	-0.254***	-0.057***
	(0.015)	(0.025)	(0.022)
15 years	-0.280***	-0.262***	-0.061**
	(0.015)	(0.031)	(0.029)
Immigrant status			
(Native children as reference category)			
Immigrant background	-0.030***	-0.054***	0.052***
	(0.011)	(0.004)	(0.003)
Immigrant background · trend	0.004***		
	(0.001)		
Immigrant background · born abroad · trend	0.001		
	(0.001)		
Female	0.106***	0.103***	0.042***

	(0.002)	(0.004)	(0.004)
Controls			
Cohort	Yes	Yes	Yes
Country of origin	Yes	Yes	Yes
Family structure	Yes	No	No
Mother and father's socioeconomic status	Yes	No	No
- plus interaction with immigrant background	Yes	No	No
Mother and father's education	Yes	No	No
- plus interaction with immigrant background	Yes	No	No
Controlling for grade points	No	No	Yes
Observations	144,370	51,360	51,360
Adjusted R-squared	0.146	0.048	0.284

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 in specification (1) and between 2002 and 2007 with grade information available in specification (2) and (3) (Sample selection described in notes to table 2). The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, in specification (1) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The controls include 18 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 10 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Figure 10: Age at arrival
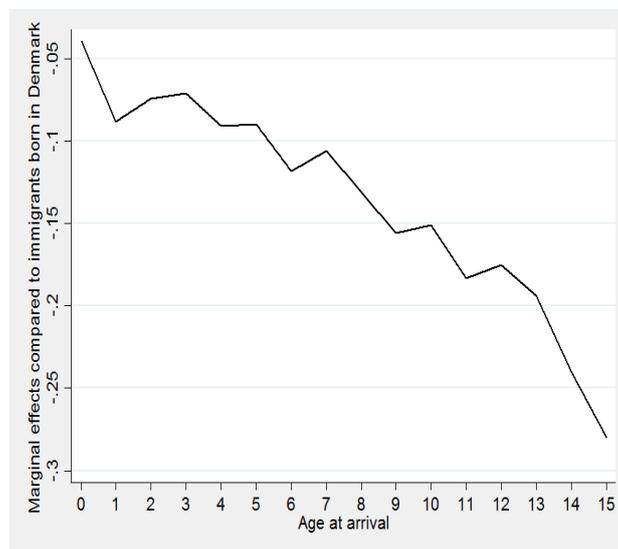
NOTE: Reference category is born in Denmark. Marginal effects from table 14 column (1). Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2).

In the baseline analysis, where gender was interacted with having immigrant background, being born abroad, and the two catching up rates, the results showed that after controlling for several individual and family characteristics, there was no significant gender difference in the

effect of having immigrant background. However, for the immigrant children born abroad, the results showed a larger disadvantage for the immigrant girls born abroad than for the immigrant boys born abroad. By allowing for gender interactions in this subanalysis, I can examine how the gender difference for the immigrants born abroad is related to arrival age. Additionally, I can examine what happens to the gender specific catching up rates when age at arrival is accounted for. I, therefore, estimate a model with the same controls as in column (1), but where immigrant background, the two catching up rates and each of the arrival ages are interacted with gender. Due to space concerns the results of this estimation is not included but reported in appendix J column (2). The results show that the gender difference for the immigrants born abroad particularly is related to the latest arriving immigrants. A boy arriving at the age of 13 have a completion rate 19.3 percentage points lower than a native boy (0.038+0.155). The difference for the corresponding girls is 27.2 percentage points (0.038+0.155+0.079). The results further show that the catching up rates are 0.3 percentage points for immigrant boys and 0.6 percentage points for immigrant girls both born in Denmark and abroad. Hence, the immigrant girls are also catching up with their native counterparts at a faster pace than the boys are, when arrival age has been accounted for.

The results in column (2) and (3) is based on the grade cohorts. Comparing the effects of each age at arrival based on the full sample and the grade cohorts in column (1) and (2) show that the effects are rather similar. In both estimations, it is from the age of seven to eight a big jump is found, and after this age that the disadvantage increases steeply with later arrival. By additionally including grade point averages, it becomes possible to examine whether the lower educational attainment of the foreign-born immigrants is explained by differences in performance in compulsory school. The results from the estimation including grade points are reported in column (3), and it is here seen that it highly affects the effect of age at arrival. Several of the arrival ages become insignificant, and if they remain significant, they are very small compared to in column (1) and (2). The disadvantage of being born abroad compared to being native either disappears or becomes marginal for all arrival ages. For instance, is the difference in the completion rate between a native child and an immigrant arriving at the age of 14 only 0.5 percentage points (0.052-0.057), when grade differences are controlled for. The results show that there is a strong link between later arrival and lower educational attainment and that the critical arrival age is around seven. The gap is, however, explained by differences in performance in compulsory school meaning that if the immigrants born abroad had the same grade point average in compulsory school as the natives, they would have similar upper secondary completion rates.

To sum up, the baseline analysis showed that a substantial native-immigrant education gap exists in upper secondary school in Denmark, and that the immigrant boys have particularly low completion rates compared to the native boys. Fortunately, the results also showed that the gap has been sharply reduced across time and that the immigrants born in Denmark and abroad both are catching up with the natives. The immigrant girls are, however, catching up with their native counterparts at a faster pace than the boys are. In the baseline analysis, I found a higher

catching up rate for the immigrants born abroad than for the immigrants born in Denmark. This however vanished, when arrival age was accounted for, suggesting that it was simply a consequence of changes in the arrival age composition of the foreign-born immigrants. I found a strong link between arrival age and educational attainment and the disadvantage, particularly, increases for the children arriving after the age of seven. Examination of the sources of the gap showed that family background differences explain the main part of the gap between the native children and immigrant children born in Denmark. For the immigrant children born abroad a significant part of the gap remained unexplained. This, however, changed when compulsory school performance was included in the analysis. When grade point averages from compulsory school is accounted for, they explain the entire gap, and family background becomes less important. Even for the latest arriving immigrants the attainment gap in upper secondary school is entirely explained by performance differences in compulsory school.

8.4   Comparison to other studies

The results of the empirical analysis will now be compared to the results from similar studies. It will be discussed to what extent my findings are consistent with the results from similar studies and on which areas, they differ. In my empirical analysis, a sizeable gap between natives and immigrants in Denmark is documented and this is in accordance with previous Danish studies (Colding et al. 2009, Rangvid 2010 and Jakobsen & Smith 2006). Furthermore, the results suggest a smaller difference in the educational attainment between immigrant girls and native girls than immigrant boys and native boys. This is opposite of what Rangvid (2010) finds. Her hypothesis is that the native-immigrant gap will be larger for girls than for boys due to cultural bias against female education in some countries of origin. Her empirical evidence supports this for Turkish students, but the gender difference is insignificant for the other countries of origin. Her analysis is based on data from the PISA 2000 test, which includes a limited number of immigrant students[23], and it is therefore questionable how general her results are. Findings from Sweden are, however, in accordance with my results. Støren & Helland (2010) find a higher upper secondary completion rate for female students, and that the gender difference applies to the immigrant children to a larger extent than to the Scandinavian children.

When the sources of the gap is examined, differences in family background is found to account for a main part of the gap, particularly, between natives and Danish-born immigrants. This is not a surprising result, and studies often find differences in family background to be the key explanation of the gap. Colding et al. (2009) find that fewer immigrants would drop out of upper secondary school, if immigrant children had parental characteristics equal to an average native child. They find that with similar family background, the difference between the natives and the Danish-born immigrants[24] is marginal, but a large gap remains between the natives and the

---

[23] Her analysis includes 1892 immigrant students in total and only 433 from Turkey, which is the largest country of origin
[24] They include immigrants born abroad but arriving in Denmark between the age of zero and five in the category of second-generation immigrants

foreign-born immigrants. This is consistent with my findings. In general, the importance of family background to explain the native-immigrant gap is stressed in several Danish and international papers as mentioned in the empirical evidence section.

One of the most positive results of the analysis is that the gap has been reduced over time. The estimated catching up rate is positive and significant for both immigrants born in Denmark and abroad regardless of the specification. This means that the positive development not is a consequence of changes in source country composition of the immigrants or the immigrant parents becoming more resourceful. In a similar study for Norway, Bratsberg et al. (2012) also find positive catching up rates for both immigrant groups, but their findings suggest that the rate is lower for the foreign-born immigrants. I find the rate to be higher for the foreign-born immigrants but that this is a consequence of changes in the arrival age composition, and when arrival age have been accounted for, the catching up rates are the same for the immigrants born in Denmark and abroad. Both this study and the Norwegian, thereby, agree that a positive catching up rate exists for both the immigrants born abroad and in the host country meaning that they are approaching the educational attainment of the natives. Riphahn (2003) finds evidence of the opposite, when examining the development in the native-immigrant gap in Germany. Her findings suggest a divergence rather than a convergence, which is a dismal result. The negative trend do, however, vanish when country of origin is controlled for suggesting that it is a source country compositional change, which causes the divergence. Furthermore, the German analysis is based on data from surveys conducted in 1996 and before. In Norway and Denmark, the positive development is much clearer in the later years meaning that a German analysis based on more recent numbers may yield brighter results for the immigrants.

Family differences were found to be the main explanation of the gap, but when grades from compulsory school are included, the importance of family characteristics is significantly reduced. In the analysis with grades, the gap between natives and immigrants is found to be entirely explained by differences in grade point averages in compulsory school. This is supported by the results of the Danish Economic Councils (2007), who find that the larger dropout rate for immigrants in upper secondary school to a large extent can be explained by their lower grades in compulsory school. The results are, furthermore, consistent with evidence from Norway. Markussen et al. (2011) find that grades from compulsory school are the far most predictive factor for upper secondary school completion, and Bratsberg et al. (2012) agree that the native-immigrant gap closes conditional on grades. As a robustness check, Bratsberg et al. (2012) applies family fixed effects. This reduces the marginal effects of the grade point averages slightly, but grade differences still explain the entire gap.

For the immigrants born abroad, there is a natural link between their arrival age and their educational attainment. The results in this analysis show that the educational disadvantage for the immigrants born abroad increases with later arrival, as expected. This is in accordance with findings from Sweden, where the results further are found to be robust to the application of

family fixed effects (Böhlmark, 2008). For the Netherlands, Van Ours & Veenman (2006) also find the disadvantage of the foreign-born immigrants to increase with arrival age but the significance of the coefficients vary both by gender and country of origin, and no clear critical arrival age can therefore be established. My findings suggest a critical arrival age around seven, which is consist with the results from Norway (Bratsberg et al., 2012). In Sweden, the critical age is found to be nine (Böhlmark, 2008). The overall conclusion based on my findings and the results from Norway and Sweden are, therefore, that it is the children arriving after school age, who experience an additional disadvantage.

The findings in this analysis is, generally, in accordance with findings from other Danish and international studies. In next section, different reservations in regards to the results will, however, be discussed and a sensitivity analysis will be performed.

9   Discussion

It is important to test the robustness of the results to secure that inaccurate conclusions are not drawn. A large fraction of the immigrants from Denmark originate from Turkey, and if they are very low performing, the attainment gap may be entirely driven by the low performing Turks. I will, therefore, check if the gap in upper secondary school remains when the immigrant children from Turkey are excluded. Furthermore, there might be a relation between for whom grade data is available and their educational performance. If grade data, particularly, is missing for immigrants with lower educational attainment, the disadvantage of the immigrants in the grade estimations may be underrated (or the advantage overrated).

The immigrants in Denmark constitute a very heterogeneous group of individuals originating from a large number of source countries. Turkey is, though, by far the largest country of origin and a major portion of both the immigrants born in Denmark and abroad originates from here. If immigrants from Turkey, in general, have lower educational attainment than the other immigrants, they may drive the entire native-immigrant gap. In a Danish study with focus on source country differences, the Turkish immigrants are found to be among the lowest performing (Rangvid, 2010). Therefore, it might be the case that what seems to be a disadvantage for all the immigrants, actually, only is driven by a substantial disadvantage for the immigrants from Turkey. It is, therefore, interesting to examine what happens to the results if the Turkish immigrants are omitted from the analysis. 23.1 percent of the immigrants in the sample originate from Turkey, which was seen in table 3 in section 6.4.1. Examining the immigrants born in Denmark and abroad separately, 37.4 percent of the immigrants born in Denmark originate from Turkey, whereas 12.2 percent of the foreign-born immigrants are Turkish. This implies that omitting the Turks from the estimation reduces the sample of immigrants born in Denmark from 24,005 to 15,024 and immigrants born abroad from 31,450 to 27,632.

Table 15: OLS estimation results after excluding immigrants with Turkish background

Dependent variable: Completion of upper secondary education within five years of compulsory school				
	(1)	(2)	(3)	(4)
Immigrant background	-0.169***	-0.107***	-0.029***	0.001
	(0.009)	(0.011)	(0.011)	(0.012)
Immigrant background · born abroad	-0.163***	-0.182***	-0.145***	-0.150***
	(0.011)	(0.013)	(0.011)	(0.014)
Immigrant background · trend	0.005***	0.004***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend	0.006***	0.004***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.104***	0.096***	0.107***	0.100***
	(0.003)	(0.003)	(0.002)	(0.003)
Controls				
Cohort	Yes	Yes	Yes	Yes
Country of origin	No	No	Yes	Yes
Family structure	No	No	Yes	Yes
Mother and father's socioeconomic status	No	No	Yes	Yes
- plus interaction with immigrant background	No	No	Yes	Yes
Mother and father's education	No	No	Yes	Yes
- plus interaction with immigrant background	No	No	Yes	Yes
Observations	144,370	131,571	144,370	131,571
Description of sample	Full	Without Turkey	Full	Without Turkey
Adjusted R-squared	0.048	0.038	0.143	0.138

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2). Specification (2) and (4) are based on the same sample selection but excludes immigrant children with Turkey as their country of origin. The regressions use the linear probability model. In specification (3) and (4), the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The controls include 18 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 10 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

In table 15, I report the results of the estimations without Turkey in column (2) and (4), and as comparison, I also report the results based on the entire sample in column (1) and (3). The first two estimations give the size of the raw gap and has cohort and gender dummies as the only additional controls. The results in column (3) and (4) are based on the full model. The first thing to note by comparing the results pairwise is that the difference in the upper secondary completion rate between immigrants born in Denmark and natives is smaller when Turkey is left out of the estimations. In the full sample, the difference in the completion rate between natives and immigrants born in Denmark is 16.9 percentage points, whereas the difference without Turkey is reduced to 10.7 percentage points. The difference compared to the natives for the foreign-born immigrants are also slightly reduced from 33.2 percentage points to 28.9 percentage points. The results show that the native-immigrant gap is smaller when Turkey is excluded meaning that

the immigrants from Turkey have a lower completion rate than the other immigrants, on average. A gap however remains, and the native-immigrant education gap is, therefore, not solely driven by low educational performance of the Turks. From column (4), it is seen that when Turkey is excluded, differences in family background explain the entire gap between natives and immigrants born in Denmark. It is further seen by comparing column (3) and (4) that the size of the positive catching up rate does not change, when Turkey is excluded.

The conclusion is that a significant part of the difference between the natives and the immigrants born in Denmark are driven by the immigrant children from Turkey having low educational attainment. The gap do, however, not disappear when they are excluded meaning that the remaining group of immigrants also have a lower completion rate than the natives. Furthermore, the positive catching up rate remains whether Turkey is excluded or not.

Table 16: Upper secondary completion rates for latest cohorts and grade cohorts

| | | | Immigrant background | |
	Observations	Natives	Born in DK	Born abroad
Cohorts graduating between 2002 and 2007	56,251	0.667	0.576	0.489
Cohorts graduating between 2002 and 2007 with grade data	51,360	0.697	0.611	0.549

NOTE: Upper secondary completion rate reflects the fraction of a graduating cohort who complete upper secondary school within five years of graduating from compulsory school (ninth grade). Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 (Sample selection described in notes to table 2).

Another concern regarding the results relates to the grade regressions and has been briefly commented on. The grade regressions are obviously only based on children with grade data available. Information on grades is available for 51,360 out of the 56,251 individuals graduating between 2002 and 2007 and covers 93 percent of the native subsample and 89 percent of the immigrant subsample. The estimations including grades showed that immigrant children had a higher upper secondary completion rate than native children if they had same grade point average in compulsory school, but the advantage may be overrated if some of the weaker immigrant students are omitted because of missing grade data. One way to examine if the grade cohorts differ from the entire latest cohorts is to compare the completion rate for the two groups and these are reported in table 16. Both natives and immigrants born in Denmark have a 3 percentage points higher completion rate, when only averaged across the individuals with grade data. For the immigrants born abroad the difference is 6 percentage points, and this suggests that it, in particular, is for the immigrant children born abroad that the disadvantage may be underrated or the advantage overrated in the grade estimations. One way to examine this is to compare estimation results based on the entire cohorts of 2002 to 2007 and the grade cohorts. Therefore, the full model is re-estimated based on these two groups, which is reported in table 17. The results are found to be strikingly similar. Based on both samples, differences in family characteristics explain the entire gap for the immigrants born in Denmark and only a small disadvantage for the immigrants born abroad of approximately 5 percentage points remains. The

positive effect of the immigrants in the grade estimations do, therefore, not seem to be overrated and be a consequence of omitting the weaker immigrants from the analysis.

Table 17: OLS estimation results for latest cohorts and grade cohorts

Dependent variable: Completion of upper secondary education within five years of compulsory school		
	(1)	(2)
Immigrant background	0.037*	0.040**
	(0.018)	(0.017)
Immigrant background · born abroad	-0.078***	-0.0917***
	(0.008)	(0.008)
Female	0.111***	0.121***
	(0.004)	(0.004)
Controls		
Cohort	Yes	Yes
Country of origin	Yes	Yes
Family structure	Yes	Yes
Mother and father's socioeconomic status	Yes	Yes
- plus interaction with immigrant background	Yes	Yes
Mother and father's education	Yes	Yes
- plus interaction with immigrant background	Yes	Yes
Observations	51,360	56,251
Description of sample	2002-2007 cohorts with grade data	2002-2007 cohorts
Adjusted R-squared	0.114	0.135

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 (Sample selection described in notes to table 2). Specification (1) is based on the individuals graduating ninth grade between 2002 and 2007 with grade data available and specification (2) on the total sample of individuals graduating between 2002 and 2007. The regressions use the linear probability model. In both specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The controls include 6 cohort dummies, 14 country of origin dummies, 8 age group dummies for parents' age at childbirth, 5 dummies for number of siblings, a divorce dummy, 9 dummies for different levels of parents' socioeconomic status and 3 dummies for parents' educational level as well as missing value indicators. Unreported constant is included in all regressions. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The third and final concern to address is the endogeneity problem, and how it affects the results. The problem of endogeneity has already been commented on several times. The empirical approach used in this analysis do not fix the endogeneity problem and one, therefore, has to be aware of the possibility of biased estimates, when interpreting the results. This means that the results should be seen as associations rather than clear causal relationships. However, including as many relevant variables as possible reduces the endogeneity, and an extensive number of individual and family controls are included in this analysis. The endogeneity problem is, particularly, a concern in the estimations where the child's own abilities not are explicitly controlled for. In the estimations where both country of origin, grade point averages in compulsory school

and family structure and family background variables are accounted for, the endogeneity problem should be minor. That grade differences in compulsory school is a main explanation of the gap in upper secondary school should, therefore, be a robust result.

To sum up, the native-immigrant attainment gap in upper secondary school is larger, when Turkey is included, but the low performance of the Turks does not drive the entire gap. Furthermore, the results of the grade estimations do not seem to be overrated, since estimations based on the grade cohorts and the entire latest cohorts yield strikingly similar results. The endogeneity is the main concern and implies that some of the results should be interpreted with caution but it does not change the main conclusion.

10 Conclusion

Proper integration of immigrants becomes increasingly important, as they constitute a growing share of the Danish population. Education is an important part of the integration process, and this paper examines the educational attainment of the children of immigrants from non-Western countries. The analysis is based on the 18 cohorts graduating from ninth grade in compulsory school between 1990 and 2007 and shows that immigrant children are less likely to complete upper secondary school within five years than native children are. Both immigrant children born in Denmark and abroad have a completion rate lower than the natives', but the disadvantage is particularly large for the immigrants born abroad. Separate examination of the boys and girls further show that the immigrant boys, in particular, is lagging behind the native boys. A positive picture, however, emerges when examining the development in the attainment gap across the 18 cohorts. The gap has been considerably reduced, and the positive development is not simply a consequence of compositional changes in the immigrant population. I find robust evidence of a positive and significant trend, which remains after source countries and parental background have been controlled for. The upper secondary completion rate for both immigrant children born in Denmark and abroad is found to catch up with that of the natives by 0.4 percentage points per year. The catching up rate is positive and significant for both immigrant groups, but when gender differences in the catching up rates are allowed, the rate is found to be larger for girls than for boys. This means that the immigrant girls are catching up with their native counterparts at a faster pace than the immigrant boys are.

It is not only interesting whether the native-immigrant gap exists, and how it has developed across time, but also why the educational differences exist in the first place. Differences in family background between natives and immigrants are found to be the key explanation, and when these differences are controlled for, the remaining gap between native children and immigrant children born in Denmark is marginal. A substantial gap, however, remains between the natives

and the foreign-born immigrants, and the size of this gap is found to be strongly related to arrival age. Later arrival is linked to lower completion rates. Seven is determined as the critical arrival age meaning that the disadvantage for the children born abroad is found to increase sharply if they arrive after the age of seven. For the six cohorts graduating between 2002 and 2007, grade point averages from compulsory school is further available. When performance differences in compulsory school have been accounted for, the gap is completely eliminated, and family background becomes much less important. Even the disadvantage for the immigrant children born abroad disappears and is turned to a small but significant advantage. This suggests that if the performance differences between the native and immigrant children in compulsory school can be evened out, the immigrant children will outperform the natives in upper secondary education. That is, the lower upper secondary completion rate for the immigrants seem to occur because the immigrants, on average, have a poorer starting point than the natives, when they enter upper secondary school. Hence, if policy makers succeeded in improving the educational performance of immigrant children in compulsory school it would translate into higher upper secondary completion rate for the immigrant children, as well.

The development in the educational attainment of the immigrants across cohorts is positive, but a suggestion to future research could be to examine if it translates into success on the labour market. Nielsen et al. (2003) find no significant economic return on an extra year of education for immigrant women born in Denmark. The purpose of increasing the educational level of the immigrants is to give them better conditions, when they enter the labour market. If higher educational attainment of immigrants does not transform into better labour market outcomes then this is an area that calls for further examination.

References

Adli, R., Louichi, A. & Tamouh, N., 2010. The sibling size impact on the educational achievement in France. *Education Economics*, 18(3), pp. 331-348.

Angrist, J. D. & Pischke, J.-S., 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Bartolomeo, A. D., 2011. Explaining the gap in educational achievement between second-generation immigrants and natives: the Italian case. *Journal of Modern Italian Studies,* 16(4), pp. 437-449.

Bauer, P. & Riphahn, R., 2007. Heterogeneity in the intergenerational transmission of educational attainment: Evidence from Switzerland on natives and second-generation immigrants. *Journal of Population Economics,* 20(1), pp. 121-148.

Becker, G. S., 1964. *Human capital: a theoretical and empirical analysis, with special reference to education.* 1. ed. New York: National Bureau of economic Research.

Becker, G. S. & Tomes, N., 1979. An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy,* 87(6), pp. 1153-1189.

Borjas, G. J., 1992. Ethnic Capital and Intergenerational Mobility. *The Quarterly Journal of Economics,* 107(1), pp. 123-150.

Bratsberg, B., Raaum, O. & Røed, K., 2012. Educating children of immigrants: Closing the gap in Norwegian schools. *Nordic Economic Policy Review,* pp. 211-251.

Böhlmark, A., 2008. Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Economics,* Volume 15, pp. 1366-1387.

Cahuc, P. & Zylberberg, A., 2004. *Labor Economics.* Cambridge: MIT Press.

Card, D., DiNardo, J. & Estes, E., 1998. *The more things change: Immigrants and the children of immigrants in the 1940s, the 1970s, and the 1990s,* Massachusetts: NBER working paper 6519, National Bureau of Economic Research.

Castagnetti, C. & Rosti, L., 2009. Effort allocation in tournaments: The effect of gender on academic performance in Italian universities. *Economics of Education Review,* Volume 28, pp. 357-369.

Colding, B., Hummelgaard, H. & Husted, L., 2005. How Studies of the Educational Progression of Minority Children Are Affecting Education Policy in Denmark. *Social Policy & Administration,* 39(6), pp. 684-696.

Colding, B., Husted, L. & Hummelgaard, H., 2009. Educational progression of second-generation immigrants and immigrant children. *Economics of Education Review,* Volume 28, pp. 434-443.

Cunha, F. & Heckman, J. J., 2008. Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *The Journal of Human Resources,* 43(4), pp. 738-782.

Danish Economic Councils, 2007. *Dansk Økonomi - Efterår,* s.l.: s.n.

Deding, M. & Hussain, M. A., 2005. Educational Attainment in Denmark: The role of parents' education and childhood living conditions. *Journal of Applied Aocial Science Studies,* 125(3), pp. 1008-1019.

Dickson, M., Gregg, P. & Robinson, H., 2013. Early, Late or Never? When Does Parental Education Impact Child Outcomes?. *IZA Discussion Paper,* Issue 7123.

Dustmann, C., Frattini, T. & Lanzara, G., 2012. Educational achievement of second-generation immigrants: an international comparison. *Economic policy,* 69(69), pp. 143-185.

Gordon, M. M., 1964. *Assimilation in American Life.* New York: Oxford University Press.

Hanushek, E. A., 1973. *Education and Race: An Analysis of the Educational Production Process.* 2. ed. Lexington: Lexington Books.

Hanushek, E. A., 1979. Conceptual and Empirical Issues in the Estimation of Educational Production Function. *The Journal of Human Resources,* 14(3), pp. 351-388.

Jakobsen, V. & Smith, N., 2006. The Educational Attainment of the Children of the Danish 'Guest Worker' Immigrants. *Danish Journal of Economics,* 144(2), pp. 18-42.

Machin, S. & McNally, S., 2005. Gender and student achievement in English schools. *Oxford Review of Economic Policy,* 21(3), pp. 357-372.

Markussen, E., Frøseth, M. W. & Sandberg, N., 2011. Reaching for the Unreachable: Identifying Factors Predicting Early School Leaving and Non-Completion in Norwegian Upper Secondary Education. *Scandinavian Journal of Educational Research,* 55(3), pp. 225-253.

McNabb, R., Pal, S. & Sloane, P., 2002. Gender Differences in Educational Attainment: The Case of University Students in England and Wales. *Economica,* 69(275), pp. 481-503.

Mikkelsen, M. R., 2013. *Berlingske.* [Online]
Available at: http://www.b.dk/nationalt/indvandrerpiger-naar-samme-uddannelsesniveau-som-danske-piger
[Accessed 6 April 2015].

Ministry of Education, 2011. *Erhvervsuddannelserne - fakta og muligheder,* s.l.: s.n.

Ministry of refugees, immigrants and integration, 2005. *Udlændinge på ungdomsuddannelserne - frafald og faglige kundskaber,* s.l.: s.n.

Nielsen, H. S. & Rangvid, B. S., 2012. The impact of parents' years since migration on children's academic achievement. *IZA Journal of Migration,* 1(1), pp. 1-23.

Nielsen, H. S., Rosholm, M., Smith, N. & Husted, L., 2003. The School-to-Work Transition of 2nd Generation Immigrants in Denmark. *Journal of Population Economics,* 16(4), pp. 755-786.

Oaxaca, R., 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review,* 14(3), pp. 693-709.

OECD, 2010. *Economic Policy Reforms - Going for growth.* s.l.:OECD publishing.

Rangvid, B. S., 2007. Sources of Immigrants' Underachievement: Results from PISA-Copenhagen. *Education Economics,* 15(3), pp. 293-326.

Rangvid, B. S., 2010. Source country differences in test score gaps: Evidence from Denmark. *Education Economics,* 18(3), pp. 269-295.

Riphahn, R. T., 2003. Cohort Effects in the Educational Attainment of Second Generation Immigrants in Germany: An Analysis of Census Data. *Journal of Population Economics,* 16(4), pp. 711-737.

Schneeweis, N., 2011. Educational institutions and the integration of migrants. *Journal of Population Economics,* 24(4), pp. 1281-1308.

Schnepf, S. V., 2007. Immigrants' educational disadvantage: an examination across ten countries and three surveys. *Journal of Population Economics,* 20(3), pp. 527-545.

Schou, P., 2006. Immigration, Integration and Fiscal Sustainability. *Journal of Population Economics,* 19(4), pp. 671-689.

Statistics Denmark, 2014. *Statistisk årbog.* København: Rosendahls - schultz Grafisk A/S.

Støren, L. A. & Helland, H., 2010. Ethnicity Differences in the Completion Rates of Upper Secondary Education: How Do the Effects of Gender and Social Background Variables Interplay?. *European Sociological Review,* 26(5), pp. 585-601.

Tartari, M., 2015. Divorce and the cognitive achievement of children. *International Economic Review,* 56(2), pp. 597-645.

The Welfare Agreement, 2006. *Aftale mellem regeringen (Venstre og Det Konservative Folkeparti) og Socialdemokraterne, Dansk Folkeparti og Det Radikale Venstre om initiativer til sikring af fremtidens velstand og velfærd og investeringer i fremtiden.* s.l.:s.n.

Todd, P. E. & Wolpin, K. I., 2003. On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal,* 113(485), pp. F3-F33.

Todd, P. E. & Wolpin, K. I., 2007. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital,* 1(1), pp. 91-136.

Van Ours, J. & Veenman, J., 2006. Age at immigration and educational attainment of young immigrants. *Economic Letters,* 90(3), pp. 310-316.

Weinberg, B. A., 2001. An Incentive Model of the Effect of Parental Income on Children. *Journal of Political Economy,* 109(2), pp. 266-280.

Wilson, K., 2001. The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions. *Southern Economic Journal,* 67(3), pp. 518-551.

Wooldridge, J. M., 2003. *Introductory econometrics: A modern approach.* 2. ed. Australia: Thomson/South-Western.

Wooldridge, J. M., 2010. *Econometric analysis of cross section and panel data.* 2. ed. Cambridge: the MIT press.

A. Definitions and primary data source for all variables

Variable	Definitions	Primary Data Source
The dependent variable		
Completion of upper secondary education within five years of graduating from compulsory school	Dummy for completing upper secondary education within five years of graduating from compulsory school. The year the individual completes upper secondary school is found based on changes in the education code. The year the education code for highest degree attained changes to an upper secondary education is compared to the compulsory school graduating cohort to determine if the completion is accomplished within five years of compulsory school.	The educational institution register, DST
Main variables		
Immigrant background	Dummy for having parents where none of them is both Danish citizens and born in Denmark constructed from immigrant status code	Population register, DST
Born abroad	Dummy for not being born in Denmark constructed from immigrant status code	Population register, DST
Trend	Variable varying from 1 to 18 dependent on which graduating cohort from 1990 to 2007 the individual belongs to. Graduating cohort is constructed based on year and education code of highest degree attained. Graduating cohort is denoted as the year the highest degree attained equals ninth grade for the first time.	The educational institution register, DST
Individual characteristics		
Female	Dummy for gender	Population register, DST
Cohort	Dummy for year individual graduates ninth grade. Graduating cohort is constructed based on year and education code of highest degree attained. Graduating cohort is denoted as the year the highest degree attained equals ninth grade for the first time.	The educational institution register, DST
Country of origin	Dummy for source country constructed from country of origin codes	Population register, DST
Age at arrival	Dummy for age at arrival constructed by information on immigration date	Population register, DST
Grade point average	Dummy for grade point average constructed by information on exam grades in ninth grade. For individuals without exam grades, but with continous assesment grades in ninth grade, these are used instead.	Compulsory school grade register, DST
Family characteristics		
Parents' age at child birth	Dummy for parents age at child birth constructed by combining information on birth cohort for parents and child	Population register, DST
Divorce	Dummy for divorce constructed based on information on marital status of parents. Takes the value one if at least one of the parents is recorded as being divorced when the child is 15.	Population register, DST
Siblings	Dummy for number of siblings under the age of 18 when the child is 15 constructed from information on number of children for mother and father. If the number of children of the mother and father differs, the largest number is used.	Population register, DST
Parents' years of education	Dummy for mother and father's years of education constructed from an education code of highest degree attained for mother and father measured when the child is 15	The educational institution register, DST
Parents' socioeconomic status		
Managers and CEOs	Dummy for managers and CEOs constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Employees on highest level	Dummy for senior officials, leading officials and employees on the highest level constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Employees on medium level	Dummy for other officials and employees on medium level constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST

Employees on lowest level	Dummy for skilled workers and employees on basic level constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Unskilled worker	Dummy for unskilled workers constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Other employees	Dummy for employees without further specification and other employees constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Self-employed and assisting spouses	Dummy for self-employed and assisting spouses constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Unemployed	Dummy for unemployed constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Enrolled in education	Dummy for enrolled in education constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Others outside the labour force	Dummy for being outside the labour force for other reasons than that he/she is enrolled in education constructed from information on mother and father's primary connection to the labour market and measured when the child is 15	Labour Force Registers, DST
Parents' years since migration	Dummy for parents years since migration constructed by information on immigration date of the mother and father and measured when child is 15. That is parents belonging to the category who arrived 15 years ago arrived approximately at their child's birth.	Population register, DST
Missing indicators	Dummy for lack of information on variable. Used in all categories where information is missing for some of the individuals	

B. Summary statistics for family structure variables, age at arrival and parents' years since migration: Means (standard deviations in parentheses)

Variable	Natives	Immigrant background	
		Born in Denmark	Born abroad
Mother's age when child is born			
Under 18 years	0.004	0.020	0.033
	(0.065)	(0.140)	(0.178)
18-22 years	0.149	0.274	0.259
	(0.356)	(0.446)	(0.438)
23-27 years	0.387	0.313	0.311
	(0.487)	(0.464)	(0.463)
28-32 years	0.313	0.224	0.192
	(0.464)	(0.417)	(0.394)
33-37 years	0.115	0.116	0.094
	(0.319)	(0.320)	(0.292)
38-42 years	0.024	0.037	0.028
	(0.153)	(0.188)	(0.164)
Over 42 years	0.002	0.005	0.009
	(0.044)	(0.069)	(0.094)
Missing	0.006	0.013	0.074
	(0.079)	(0.112)	(0.261)
Father's age when child is born			
Under 18 years	0.001	0.002	0.003
	(0.028)	(0.041)	(0.059)
18-22 years	0.052	0.111	0.066
	(0.221)	(0.315)	(0.248)
23-27 years	0.274	0.251	0.218
	(0.446)	(0.433)	(0.413)
28-32 years	0.361	0.243	0.241
	(0.480)	(0.429)	(0.428)
33-37 years	0.191	0.193	0.150
	(0.393)	(0.395)	(0.357)
38-42 years	0.066	0.098	0.072
	(0.248)	(0.297)	(0.258)
Over 42 years	0.025	0.052	0.044
	(0.156)	(0.222)	(0.206)
Missing	0.031	0.050	0.206
	(0.174)	(0.219)	(0.404)
Number of siblings			
No siblings	0.423	0.241	0.221
	(0.494)	(0.427)	(0.415)
1 sibling	0.395	0.296	0.273
	(0.489)	(0.456)	(0.446)
2 siblings	0.144	0.241	0.203
	(0.351)	(0.428)	(0.402)
3 siblings	0.031	0.138	0.132
	(0.173)	(0.344)	(0.338)
4 or more siblings	0.007	0.084	0.141
	(0.083)	(0.277)	(0.348)
Missing	0.001	0.001	0.030
	(0.029)	(0.027)	(0.172)
If parents are divorced or not			
Divorced	0.182	0.114	0.095
	(0.386)	(0.318)	(0.294)
Missing	0.001	0.001	0.030
	(0.029)	(0.027)	(0.172)
Child's age at immigration			
0 year			0.025
			(0.157)
1 year			0.060
			(0.238)
2 year			0.051
			(0.221)
3 year			0.048
			(0.215)
4 year			0.057
			(0.232)

5 year			0.060
			(0.237)
6 year			0.060
			(0.238)
7 year			0.064
			(0.245)
8 year			0.065
			(0.246)
9 year			0.066
			(0.248)
10 year			0.069
			(0.254)
11 year			0.072
			(0.259)
12 year			0.071
			(0.257)
13 year			0.068
			(0.251)
14 year			0.061
			(0.239)
15 year			0.052
			(0.222)
Missing			0.051
			(0.219)
Mother's years since migration			
0-4 years			0.159
			(0.366)
5-9 years			0.325
			(0.469)
10-14 years			0.286
			(0.452)
15-19 years		0.417	0.071
		(0.493)	(0.257)
20-24 years		0.238	0.008
		(0.426)	(0.089)
25-29 years		0.109	0.003
		(0.312)	(0.055)
30+ years		0.030	0.001
		(0.171)	(0.036)
Missing		0.205	0.145
		(0.404)	(0.352)
Father's years since migration			
0-4 years		0.012	0.101
		(0.109)	(0.301)
5-9 years		0.004	0.223
		(0.064)	(0.416)
10-14 years		0.006	0.207
		(0.080)	(0.405)
15-19 years		0.214	0.074
		(0.410)	(0.262)
20-24 years		0.213	0.027
		(0.410)	(0.163)
25-29 years		0.130	0.013
		(0.336)	(0.112)
30+ years		0.046	0.003
		(0.209)	(0.053)
Missing		0.374	0.352
		(0.484)	(0.478)
Number of observations	88,915	24,005	31,450

NOTE: Standard deviations in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2 in paper). [1]For 168 mothers of children born in Denmark their immigration date showed that they arrived after the child's birth. The immigration date must therefore not reflect their actual first immigration date and *mother's years since migration* is therefore changed to being missing for these 168 mothers.

C. Full estimation results for table 8 (OLS estimation results after controlling for different groups of control variables)

Dependent variable: Completion of upper secondary education within five years of compulsory school

	(1)	(2)	(3)	(4)	(5)	(6)
Variables of interest						
Immigrant background	-0.169***	-0.138***	-0.133***	0.009**	-0.078***	0.009**
	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Immigrant background · born abroad	-0.163***	-0.190***	-0.166***	-0.133***	-0.183***	-0.147***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Immigrant background · trend	0.005***	0.004***	0.006***	0.003***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend	0.006***	0.004***	0.004***	0.003***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.104***	0.103***	0.105***	0.106***	0.105***	0.106***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Graduating cohort						
(Cohort 1990 as reference category)						
Cohort 1991	-0.012	-0.010	-0.011	-0.010	-0.013*	-0.011
	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)
Cohort 1992	0.004	0.005	0.002	0.002	-0.003	-0.001
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1993	0.028***	0.030***	0.026***	0.030***	0.022***	0.026***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1994	0.014*	0.017**	0.010	0.015**	0.000	0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
Cohort 1995	0.010	0.013	0.005	0.022***	-0.006	0.014*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1996	0.006	0.003	-0.003	0.017**	-0.017**	0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1997	0.003	-0.001	-0.009	0.015*	-0.025***	0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1998	0.006	0.002	-0.009	0.016**	-0.028***	0.000
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1999	0.003	-0.002	-0.015*	0.004	-0.039***	-0.014*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2000	0.009	0.003	-0.002	0.016**	-0.028***	-0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2001	0.005	0.002	-0.011	0.012	-0.041***	-0.011
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2002	0.019**	0.017**	0.001	0.022***	-0.029***	-0.001
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2003	0.014*	0.013*	-0.004	0.022***	-0.037***	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2004	0.002	0.003	-0.014*	0.017**	-0.049***	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2005	0.025***	0.030***	0.008	0.038***	-0.031***	0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
Cohort 2006	0.032***	0.034***	0.011	0.042***	-0.027***	0.011
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2007	0.052***	0.055***	0.030***	0.061***	-0.010	0.028***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Source countries						
(Born in Denmark as reference category)						
Turkey		-0.244***	-0.230***	-0.076***	-0.109***	-0.036***
		(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
Pakistan		-0.124***	-0.137***	0.021*	-0.067***	0.024**
		(0.011)	(0.011)	(0.011)	(0.011)	(0.011)

Lebanon	-0.229***	-0.206***	-0.021	-0.167***	-0.036***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
former Yugoslavia	-0.182***	-0.168***	-0.025**	-0.121***	-0.032***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Bosnia & Herzegovina	0.058***	0.027**	0.178***	0.031**	0.145***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Iraq	-0.167***	-0.180***	0.006	-0.137***	-0.008
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Vietnam	-0.028**	-0.031**	0.104***	0.033**	0.110***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)
Iran	-0.004	0.006	0.136***	0.024*	0.116***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Sri Lanka	0.019	-0.020	0.112***	0.016	0.100***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
Morocco	-0.200***	-0.190***	-0.015	-0.093***	0.006
	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Somalia	-0.281***	-0.207***	-0.053***	-0.170***	-0.065***
	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Afghanistan	-0.062***	-0.059***	0.116***	-0.016	0.103***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
South Korea	0.065***	-0.001	-0.004	-0.012	-0.014
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Other countries	-0.116***	-0.103***	-0.005	-0.084***	-0.020*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)

Mother's age when child is born
(under 18 years as reference category)

18-22 years		0.020*	0.024**	0.013	0.019
		(0.012)	(0.011)	(0.012)	(0.011)
23-27 years		0.085***	0.069***	0.058***	0.057***
		(0.012)	(0.012)	(0.012)	(0.012)
28-32 years		0.133***	0.098***	0.092***	0.082***
		(0.012)	(0.012)	(0.012)	(0.012)
33-37 years		0.143***	0.104***	0.103***	0.090***
		(0.013)	(0.012)	(0.012)	(0.012)
38-42 years		0.128***	0.097***	0.095***	0.087***
		(0.014)	(0.014)	(0.014)	(0.014)
Over 42 years		-0.009	-0.010	-0.016	-0.009
		(0.023)	(0.023)	(0.023)	(0.023)
Missing		-0.019	0.004	-0.006	-0.006
		(0.014)	(0.019)	(0.015)	(0.019)

Father's age when child is born
(under 18 years as reference category)

18-22 years		0.031	0.018	0.029	0.019
		(0.032)	(0.032)	(0.032)	(0.032)
23-27 years		0.089***	0.066**	0.074**	0.062**
		(0.032)	(0.032)	(0.031)	(0.032)
28-32 years		0.118***	0.091***	0.102***	0.088***
		(0.032)	(0.032)	(0.032)	(0.032)
33-37 years		0.108***	0.088***	0.100***	0.089***
		(0.032)	(0.032)	(0.032)	(0.032)
38-42 years		0.090***	0.091***	0.096***	0.097***
		(0.032)	(0.032)	(0.032)	(0.032)
Over 42 years		0.052	0.095***	0.080**	0.108***
		(0.033)	(0.033)	(0.032)	(0.032)
Missing		0.031	0.060*	0.069**	0.063*
		(0.032)	(0.033)	(0.032)	(0.033)

Number of siblings

(No siblings as reference category)

1 sibling	0.046***	0.032***	0.034***	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)
2 siblings	0.035***	0.031***	0.026***	0.027***
	(0.004)	(0.004)	(0.004)	(0.004)
3 siblings	-0.002	0.011*	-0.003	0.009
	(0.006)	(0.006)	(0.006)	(0.006)
4 or more siblings	-0.053***	-0.024***	-0.046***	-0.022***
	(0.007)	(0.007)	(0.007)	(0.007)
Missing	-0.028*	-0.037**	0.006	-0.028
	(0.016)	(0.018)	(0.016)	(0.018)
If parents are divorced or not				
Divorced	-0.131***	-0.102***	-0.118***	-0.099***
	(0.004)	(0.004)	(0.004)	(0.004)
Mother's socioeconomic status				
(Outside labour force as reference category)				
Managers and CEOs		0.169***		0.121***
		(0.015)		(0.015)
Employees on highest level		0.227***		0.170***
		(0.005)		(0.005)
Employees on medium level		0.179***		0.134***
		(0.005)		(0.005)
Employees on lowest level		0.110***		0.091***
		(0.005)		(0.005)
Unskilled worker		0.067***		0.069***
		(0.007)		(0.007)
Other employees		0.062***		0.059***
		(0.005)		(0.005)
Self-employed and assisting spouses		0.116***		0.097***
		(0.007)		(0.007)
Unemployed		0.028***		0.023***
		(0.006)		(0.005)
Enrolled in education		0.025***		0.022**
		(0.010)		(0.009)
Missing		0.013		0.032*
		(0.016)		(0.017)
Father's socioeconomic status				
(Outside labour force as reference category)				
Managers and CEOs		0.216***		0.181***
		(0.008)		(0.008)
Employees on highest level		0.222***		0.181***
		(0.005)		(0.005)
Employees on medium level		0.173***		0.144***
		(0.006)		(0.006)
Employees on lowest level		0.105***		0.088***
		(0.005)		(0.005)
Unskilled worker		0.055***		0.073***
		(0.008)		(0.008)
Other employees		0.067***		0.065***
		(0.005)		(0.005)
Self-employed and assisting spouses		0.138***		0.124***
		(0.005)		(0.005)
Unemployed		0.026***		0.024***
		(0.006)		(0.006)
Enrolled in education		0.001		-0.003
		(0.014)		(0.014)
Missing		0.037***		0.056***
		(0.010)		(0.011)
Mother's education				

	(1)	(2)	(3)	(4)	(5)	(6)
(0-9 years as reference category)						
10-12 years					0.091***	0.066***
					(0.004)	(0.004)
13+ years					0.170***	0.109***
					(0.004)	(0.004)
Missing					0.035***	0.026***
					(0.005)	(0.005)
Father's education						
(0-9 years as reference category)						
10-12 years					0.090***	0.068***
					(0.004)	(0.004)
13+ years					0.126***	0.084***
					(0.004)	(0.004)
Missing					0.035***	0.026***
					(0.005)	(0.006)
Constant	0.589***	0.590***	0.407***	0.183***	0.267***	0.120***
	(0.006)	(0.006)	(0.033)	(0.033)	(0.033)	(0.033)
Observations	144,370	144,370	144,370	144,370	144,370	144,370
R-squared	0.048	0.063	0.09	0.131	0.121	0.141

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2 in paper). The regressions use the linear probability model. In specification (2) to (6), the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

D. OLS estimation results behind figure 5 to 8

Dependent variable: Completion of upper secondary education within five years of compulsory school

	Socioeconomic status		Years of education	
	Natives	Immigrants	Natives	Immigrants
Variables of interest				
Immigrant background · born abroad		-0.133***		-0.170***
		(0.011)		(0.011)
Immigrant background · born abroad · trend		0.003***		0.004***
		(0.001)		(0.001)
Female	0.085***	0.144***	0.083***	0.144***
	(0.003)	(0.004)	(0.003)	(0.004)
Graduating cohort				
(Cohort 1990 as reference category)				
Cohort 1991	-0.007	-0.016	-0.012	-0.017
	(0.008)	(0.018)	(0.008)	(0.018)
Cohort 1992	0.006	-0.007	-0.001	-0.011
	(0.008)	(0.018)	(0.008)	(0.018)
Cohort 1993	0.036***	0.019	0.029***	0.009
	(0.008)	(0.017)	(0.009)	(0.017)
Cohort 1994	0.017**	0.017	0.001	0.006
	(0.008)	(0.017)	(0.008)	(0.017)
Cohort 1995	0.031***	0.011	0.000	-0.008
	(0.009)	(0.017)	(0.009)	(0.017)
Cohort 1996	0.020**	0.023	-0.017**	0.003
	(0.009)	(0.017)	(0.009)	(0.016)
Cohort 1997	0.019**	0.021	-0.030***	0.002
	(0.009)	(0.017)	(0.009)	(0.016)
Cohort 1998	0.028***	0.012	-0.026***	-0.008
	(0.009)	(0.017)	(0.009)	(0.016)
Cohort 1999	0.001	0.026	-0.052***	0.007
	(0.009)	(0.017)	(0.009)	(0.017)
Cohort 2000	0.018*	0.036**	-0.038***	0.021
	(0.009)	(0.017)	(0.009)	(0.017)
Cohort 2001	0.010	0.035**	-0.057***	0.018
	(0.009)	(0.017)	(0.009)	(0.017)
Cohort 2002	0.023**	0.048***	-0.043***	0.032*
	(0.009)	(0.017)	(0.009)	(0.017)
Cohort 2003	0.021**	0.052***	-0.052***	0.033**
	(0.009)	(0.018)	(0.009)	(0.017)
Cohort 2004	0.012	0.054***	-0.070***	0.032*
	(0.009)	(0.018)	(0.009)	(0.017)
Cohort 2005	0.031***	0.078***	-0.051***	0.054***
	(0.009)	(0.018)	(0.009)	(0.017)
Cohort 2006	0.042***	0.077***	-0.041***	0.056***
	(0.009)	(0.018)	(0.009)	(0.017)
Cohort 2007	0.060***	0.102***	-0.027***	0.082***
	(0.009)	(0.018)	(0.009)	(0.017)
Source countries				
(Turkey as reference category)				
Pakistan		0.120***		0.093***
		(0.009)		(0.009)
Lebanon		0.058***		-0.012
		(0.009)		(0.009)
former Yugoslavia		0.043***		0.012
		(0.009)		(0.009)
Bosnia & Herzegovina		0.249***		0.194***
		(0.010)		(0.011)
Iraq		0.092***		0.026**
		(0.011)		(0.011)
Vietnam		0.201***		0.183***
		(0.010)		(0.011)
Iran		0.223***		0.179***
		(0.011)		(0.011)
Sri Lanka		0.210***		0.188***
		(0.011)		(0.012)
Morocco		0.072***		0.051***

	(1)	(2)	(3)	(4)
			(0.013)	(0.013)
Somalia			0.039***	-0.009
			(0.013)	(0.013)
Afghanistan			0.211***	0.155***
			(0.014)	(0.014)
South Korea			0.161***	0.201***
			(0.014)	(0.013)
Other countries			0.010***	0.080***
			(0.007)	(0.008)
Mother's age when child is born				
(under 18 years as reference category)				
18-22 years	0.114***	0.026**	0.095***	0.023*
	(0.025)	(0.013)	(0.024)	(0.013)
23-27 years	0.196***	0.040***	0.180***	0.034**
	(0.025)	(0.013)	(0.025)	(0.013)
28-32 years	0.241***	0.033**	0.233***	0.026*
	(0.025)	(0.014)	(0.025)	(0.014)
33-37 years	0.262***	0.017	0.261***	0.008
	(0.026)	(0.015)	(0.025)	(0.015)
38-42 years	0.269***	0.000	0.271***	-0.011
	(0.027)	(0.018)	(0.027)	(0.018)
Over 42 years	0.271***	-0.130***	0.263***	-0.137***
	(0.044)	(0.027)	(0.043)	(0.026)
Missing	0.151***	0.006	0.168***	-0.059***
	(0.037)	(0.025)	(0.034)	(0.016)
Father's age when child is born				
(under 18 years as reference category)				
18-22 years	0.001	0.019	0.011	0.031
	(0.059)	(0.038)	(0.056)	(0.038)
23-27 years	0.040	0.063*	0.052	0.071*
	(0.059)	(0.038)	(0.056)	(0.038)
28-32 years	0.059	0.086**	0.077	0.089**
	(0.059)	(0.038)	(0.056)	(0.038)
33-37 years	0.055	0.084**	0.078	0.084**
	(0.060)	(0.038)	(0.056)	(0.038)
38-42 years	0.064	0.081**	0.089	0.071*
	(0.060)	(0.038)	(0.056)	(0.038)
Over 42 years	0.076	0.087**	0.074	0.063
	(0.060)	(0.039)	(0.057)	(0.039)
Missing	0.044	0.045	0.060	0.038
	(0.061)	(0.040)	(0.057)	(0.038)
Number of siblings				
(No siblings as reference category)				
1 sibling	0.046***	0.011*	0.049***	0.009
	(0.004)	(0.006)	(0.004)	(0.006)
2 siblings	0.046***	-0.001	0.048***	-0.013**
	(0.005)	(0.007)	(0.005)	(0.007)
3 siblings	0.014	-0.026***	0.016*	-0.048***
	(0.009)	(0.008)	(0.009)	(0.008)
4 or more siblings	-0.007	-0.075***	-0.009	-0.105***
	(0.019)	(0.009)	(0.019)	(0.008)
Missing	0.004	-0.003	0.035	-0.024
	(0.060)	(0.020)	(0.059)	(0.018)
If parents are divorced or not				
Divorced	-0.107***	-0.064***	-0.128***	-0.072***
	(0.004)	(0.007)	(0.004)	(0.007)
Mother's socioeconomic status				
(Outside labour force as reference category)				
Managers and CEOs	0.190***	0.116**		
	(0.016)	(0.051)		
Employees on highest level	0.247***	0.158***		
	(0.007)	(0.014)		
Employees on medium level	0.202***	0.116***		
	(0.006)	(0.011)		
Employees on lowest level	0.130***	0.098***		

	(1)	(2)	(3)	(4)
	(0.006)	(0.008)		
Unskilled worker	0.098***	0.025*		
	(0.009)	(0.015)		
Other employees	0.088***	0.038***		
	(0.007)	(0.007)		
Self-employed and assisting spouses	0.141***	0.070***		
	(0.009)	(0.012)		
Unemployed	0.037***	0.024***		
	(0.009)	(0.007)		
Enrolled in education	0.006	0.020*		
	(0.025)	(0.010)		
Missing	0.084***	-0.075***		
	(0.022)	(0.024)		
Father's socioeconomic status				
(Outside labour force as reference category)				
Managers and CEOs	0.249***	0.062**		
	(0.010)	(0.027)		
Employees on highest level	0.247***	0.150***		
	(0.008)	(0.013)		
Employees on medium level	0.200***	0.120***		
	(0.008)	(0.013)		
Employees on lowest level	0.126***	0.104***		
	(0.007)	(0.008)		
Unskilled worker	0.070***	0.078***		
	(0.010)	(0.013)		
Other employees	0.084***	0.065***		
	(0.008)	(0.007)		
Self-employed and assisting spouses	0.178***	0.079***		
	(0.008)	(0.008)		
Unemployed	0.029***	0.029***		
	(0.011)	(0.008)		
Enrolled in education	0.063	-0.013		
	(0.046)	(0.015)		
Missing	0.059***	0.011		
	(0.014)	(0.016)		
Mother's education				
(0-9 years as reference category)				
10-12 years			0.123***	0.052***
			(0.005)	(0.006)
13+ years			0.213***	0.071***
			(0.005)	(0.006)
Missing			0.037***	0.000
			(0.014)	(0.006)
Father's education				
(0-9 years as reference category)				
10-12 years			0.103***	0.042***
			(0.005)	(0.007)
13+ years			0.138***	0.071***
			(0.004)	(0.006)
Missing			0.004	0.014**
			(0.010)	(0.007)
Constant	0.041	0.182***	0.131**	0.264***
	(0.060)	(0.042)	(0.056)	(0.042)
Observations	88,915	55,455	88,915	55,455
R-squared	0.115	0.11	0.109	0.102

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2 in paper). The regressions use the linear probability model. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

E. Full estimation results for table 9 (OLS estimation results after allowing for immigrant interaction with parental background)

Dependent variable: Completion of upper secondary education within five years of compulsory school

	(1)	(2)	(3)	(4)	(5)	(6)	YSM
Variables of interest							
Immigrant background	0.009**	-0.029***	-0.078***	-0.115***	0.009**	-0.029***	0.001
	(0.004)	(0.008)	(0.004)	(0.008)	(0.004)	(0.011)	(0.017)
Immigrant background · born abroad	-0.133***	-0.136***	-0.183***	-0.180***	-0.147***	-0.145***	-0.139***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.013)
Immigrant background · trend	0.003***	0.003***	0.006***	0.007***	0.004***	0.004***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend	0.003***	0.003***	0.005***	0.004***	0.004***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.106***	0.107***	0.105***	0.106***	0.106***	0.107***	0.107***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Graduating cohort							
(Cohort 1990 as reference category)							
Cohort 1991	-0.010	-0.009	-0.013*	-0.014*	-0.011	-0.011	-0.012
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Cohort 1992	0.002	0.003	-0.003	-0.005	-0.001	-0.001	-0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1993	0.030***	0.030***	0.022***	0.020***	0.026***	0.026***	0.027***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1994	0.015**	0.016**	0.000	-0.003	0.008	0.008	0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Cohort 1995	0.022***	0.024***	-0.006	-0.010	0.014*	0.014*	0.015*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1996	0.017**	0.020***	-0.017**	-0.021***	0.007	0.008	0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1997	0.015*	0.020**	-0.025***	-0.030***	0.002	0.003	0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1998	0.016**	0.020**	-0.028***	-0.034***	0.000	0.001	0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1999	0.004	0.009	-0.039***	-0.045***	-0.014*	-0.013	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2000	0.016**	0.021***	-0.028***	-0.035***	-0.003	-0.003	-0.004
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2001	0.012	0.017**	-0.041***	-0.048***	-0.011	-0.011	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2002	0.022***	0.029***	-0.029***	-0.038***	-0.001	-0.001	-0.001
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2003	0.022***	0.028***	-0.037***	-0.045***	-0.005	-0.004	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2004	0.017**	0.024***	-0.049***	-0.058***	-0.012	-0.012	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2005	0.038***	0.044***	-0.031***	-0.041***	0.007	0.006	0.005
	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Cohort 2006	0.042***	0.047***	-0.027***	-0.039***	0.011	0.008	0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2007	0.061***	0.067***	-0.010	-0.022***	0.028***	0.025***	0.025***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Source countries							
(Born in Denmark as reference category)							
Turkey	-0.076***	-0.013	-0.109***	-0.030***	-0.036***	0.045***	0.136***
	(0.010)	(0.014)	(0.011)	(0.012)	(0.011)	(0.015)	(0.029)
Pakistan	0.021*	0.090***	-0.067***	0.034***	0.024**	0.123***	0.218***
	(0.011)	(0.014)	(0.011)	(0.013)	(0.011)	(0.016)	(0.029)

Lebanon	-0.021	0.043***	-0.167***	-0.056***	-0.036***	0.064***	0.148***
	(0.013)	(0.016)	(0.013)	(0.015)	(0.013)	(0.017)	(0.031)
former Yugoslavia	-0.025**	0.037**	-0.121***	-0.011	-0.032***	0.066***	0.172***
	(0.012)	(0.015)	(0.012)	(0.013)	(0.012)	(0.016)	(0.030)
Bosnia & Herzegovina	0.178***	0.239***	0.031**	0.166***	0.145***	0.252***	0.376***
	(0.013)	(0.016)	(0.013)	(0.015)	(0.013)	(0.017)	(0.031)
Iraq	0.006	0.072***	-0.137***	-0.024	-0.008	0.094***	0.211***
	(0.014)	(0.016)	(0.014)	(0.015)	(0.014)	(0.017)	(0.031)
Vietnam	0.104***	0.177***	0.033**	0.136***	0.110***	0.210***	0.297***
	(0.013)	(0.016)	(0.014)	(0.015)	(0.014)	(0.017)	(0.031)
Iran	0.136***	0.214***	0.024*	0.148***	0.116***	0.233***	0.327***
	(0.014)	(0.017)	(0.014)	(0.016)	(0.014)	(0.018)	(0.032)
Sri Lanka	0.112***	0.183***	0.016	0.134***	0.100***	0.205***	0.300***
	(0.015)	(0.017)	(0.015)	(0.016)	(0.015)	(0.018)	(0.032)
Morocco	-0.015	0.045***	-0.093***	-0.004	0.006	0.092***	0.197***
	(0.015)	(0.017)	(0.015)	(0.016)	(0.015)	(0.018)	(0.031)
Somalia	-0.053***	0.017	-0.170***	-0.063***	-0.065***	0.040**	0.158***
	(0.015)	(0.018)	(0.016)	(0.017)	(0.016)	(0.019)	(0.032)
Afghanistan	0.116***	0.183***	-0.016	0.095***	0.103***	0.205***	0.340***
	(0.017)	(0.019)	(0.017)	(0.018)	(0.017)	(0.020)	(0.033)
South Korea	-0.004	0.121***	-0.012	0.130***	-0.014	0.139***	0.304***
	(0.015)	(0.018)	(0.015)	(0.016)	(0.015)	(0.019)	(0.033)
Other countries	-0.005	0.081***	-0.084***	0.036***	-0.020*	0.101***	0.218***
	(0.011)	(0.014)	(0.011)	(0.013)	(0.011)	(0.016)	(0.030)

Mother's age when child is born
(under 18 years as reference category)

18-22 years	0.024**	0.027**	0.013	0.016	0.019	0.022*	0.022**
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)
23-27 years	0.069***	0.070***	0.058***	0.057***	0.057***	0.058***	0.060***
	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
28-32 years	0.098***	0.097***	0.092***	0.089***	0.082***	0.081***	0.084***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
33-37 years	0.104***	0.104***	0.103***	0.101***	0.090***	0.091***	0.096***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
38-42 years	0.097***	0.097***	0.095***	0.094***	0.087***	0.087***	0.093***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Over 42 years	-0.010	-0.012	-0.016	-0.019	-0.009	-0.011	0.004
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Missing	0.004	0.032	-0.006	-0.006	-0.006	0.024	0.030
	(0.019)	(0.020)	(0.015)	(0.015)	(0.019)	(0.020)	(0.020)

Father's age when child is born
(under 18 years as reference category)

18-22 years	0.018	0.020	0.029	0.028	0.019	0.020	0.013
	(0.032)	(0.032)	(0.032)	(0.031)	(0.032)	(0.032)	(0.032)
23-27 years	0.066**	0.066**	0.074**	0.072**	0.062**	0.060*	0.055*
	(0.032)	(0.032)	(0.031)	(0.031)	(0.032)	(0.031)	(0.031)
28-32 years	0.091***	0.089***	0.102***	0.099***	0.088***	0.085***	0.080**
	(0.032)	(0.032)	(0.032)	(0.031)	(0.032)	(0.031)	(0.031)
33-37 years	0.088***	0.087***	0.100***	0.098***	0.089***	0.086***	0.082***
	(0.032)	(0.032)	(0.032)	(0.031)	(0.032)	(0.032)	(0.032)
38-42 years	0.091***	0.089***	0.096***	0.096***	0.097***	0.094***	0.091***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Over 42 years	0.095***	0.093***	0.080**	0.076**	0.108***	0.103***	0.101***
	(0.033)	(0.033)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Missing	0.060*	0.059*	0.069**	0.065**	0.063*	0.060*	0.056*
	(0.033)	(0.033)	(0.032)	(0.032)	(0.033)	(0.033)	(0.033)

Number of siblings
(No siblings as reference category)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 sibling	0.032***	0.031***	0.034***	0.032***	0.028***	0.026***	0.026***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
2 siblings	0.031***	0.027***	0.026***	0.023***	0.027***	0.023***	0.021***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
3 siblings	0.011*	0.005	-0.003	-0.006	0.009	0.003	0.000
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
4 or more siblings	-0.024***	-0.033***	-0.046***	-0.052***	-0.022***	-0.032***	-0.032***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Missing	-0.037**	-0.001	0.006	-0.014	-0.028	0.003	0.001
	(0.018)	(0.019)	(0.016)	(0.016)	(0.018)	(0.019)	(0.019)

If parents are divorced or not

Divorced	-0.102***	-0.096***	-0.118***	-0.114***	-0.099***	-0.092***	-0.094***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)

Mother's socioeconomic status
(Outside labour force as reference category)

Managers and CEOs	0.169***	0.194***			0.121***	0.122***	0.122***
	(0.015)	(0.016)			(0.015)	(0.016)	(0.016)
Employees on highest level	0.227***	0.255***			0.170***	0.171***	0.171***
	(0.005)	(0.007)			(0.005)	(0.007)	(0.007)
Employees on medium level	0.179***	0.205***			0.134***	0.139***	0.138***
	(0.005)	(0.006)			(0.005)	(0.006)	(0.006)
Employees on lowest level	0.110***	0.129***			0.091***	0.095***	0.095***
	(0.005)	(0.006)			(0.005)	(0.006)	(0.006)
Unskilled worker	0.067***	0.097***			0.069***	0.093***	0.094***
	(0.007)	(0.009)			(0.007)	(0.008)	(0.008)
Other employees	0.062***	0.088***			0.059***	0.076***	0.076***
	(0.005)	(0.007)			(0.005)	(0.007)	(0.007)
Self-employed and assisting spouses	0.116***	0.147***			0.097***	0.110***	0.109***
	(0.007)	(0.009)			(0.007)	(0.009)	(0.009)
Unemployed	0.028***	0.033***			0.023***	0.016*	0.016*
	(0.006)	(0.009)			(0.005)	(0.009)	(0.009)
Enrolled in education	0.025***	0.001			0.022**	-0.022	-0.021
	(0.010)	(0.025)			(0.009)	(0.025)	(0.025)
Missing	0.013	0.081***			0.032*	0.117***	0.115***
	(0.016)	(0.019)			(0.017)	(0.024)	(0.024)

Father's socioeconomic status
(Outside labour force as reference category)

Managers and CEOs	0.216***	0.258***			0.181***	0.205***	0.205***
	(0.008)	(0.010)			(0.008)	(0.010)	(0.010)
Employees on highest level	0.222***	0.256***			0.181***	0.197***	0.196***
	(0.005)	(0.008)			(0.005)	(0.008)	(0.008)
Employees on medium level	0.173***	0.205***			0.144***	0.159***	0.159***
	(0.006)	(0.008)			(0.006)	(0.008)	(0.008)
Employees on lowest level	0.105***	0.127***			0.088***	0.095***	0.095***
	(0.005)	(0.007)			(0.005)	(0.007)	(0.007)
Unskilled worker	0.055***	0.070***			0.073***	0.088***	0.088***
	(0.008)	(0.010)			(0.008)	(0.010)	(0.010)
Other employees	0.067***	0.085***			0.065***	0.076***	0.076***
	(0.005)	(0.008)			(0.005)	(0.008)	(0.008)
Self-employed and assisting spouses	0.138***	0.185***			0.124***	0.155***	0.155***
	(0.005)	(0.008)			(0.005)	(0.008)	(0.008)
Unemployed	0.026***	0.027**			0.024***	0.018*	0.018*
	(0.006)	(0.011)			(0.006)	(0.011)	(0.011)
Enrolled in education	0.001	0.059			-0.003	0.051	0.052
	(0.014)	(0.046)			(0.014)	(0.044)	(0.044)
Missing	0.037***	0.071***			0.056***	0.122***	0.121***
	(0.010)	(0.013)			(0.011)	(0.018)	(0.018)

Mother's socioeconomic status · immigrant background

(Outside labour force as reference category)

	(1)	(2)	(3)	(4)	(5)
Managers and CEOs · immigrant background	-0.087			-0.033	-0.005
	(0.053)			(0.053)	(0.053)
Employees on highest level · immigrant background	-0.097***			-0.034**	-0.028*
	(0.015)			(0.015)	(0.015)
Employees on medium level · immigrant background	-0.085***			-0.038***	-0.037***
	(0.013)			(0.013)	(0.013)
Employees on lowest level · immigrant background	-0.021**			0.004	0.002
	(0.010)			(0.010)	(0.010)
Unskilled worker · immigrant background	-0.059***			-0.059***	-0.067***
	(0.017)			(0.017)	(0.017)
Other employees · immigrant background	-0.041***			-0.030***	-0.034***
	(0.010)			(0.010)	(0.010)
Self-employed and assisting spouses · immigrant background	-0.073***			-0.037**	-0.042***
	(0.015)			(0.015)	(0.015)
Unemployed · immigrant background	-0.004			0.012	0.006
	(0.011)			(0.011)	(0.011)
Enrolled in education · immigrant background	0.022			0.045*	0.050*
	(0.027)			(0.027)	(0.027)
Missing · immigrant background	-0.137***			-0.154***	-0.126***
	(0.021)			(0.026)	(0.027)

Father's socioeconomic status · immigrant background

(Outside labour force as reference category)

	(1)	(2)	(3)	(4)	(5)
Managers and CEOs · immigrant background	-0.204***			-0.168***	-0.139***
	(0.029)			(0.029)	(0.029)
Employees on highest level · immigrant background	-0.106***			-0.068***	-0.063***
	(0.015)			(0.015)	(0.015)
Employees on medium level · immigrant background	-0.083***			-0.049***	-0.048***
	(0.015)			(0.015)	(0.015)
Employees on lowest level · immigrant background	-0.015			0.013	0.008
	(0.010)			(0.010)	(0.010)
Unskilled worker · immigrant background	0.016			0.002	-0.006
	(0.016)			(0.016)	(0.016)
Other employees · immigrant background	-0.012			-0.006	-0.010
	(0.011)			(0.011)	(0.011)
Self-employed and assisting spouses · immigrant background	-0.098***			-0.073***	-0.078***
	(0.011)			(0.011)	(0.011)
Unemployed · immigrant background	0.008			0.016	0.009
	(0.013)			(0.013)	(0.013)
Enrolled in education · immigrant background	-0.070			-0.063	-0.045
	(0.048)			(0.047)	(0.047)
Missing · immigrant background	-0.051***			-0.086***	-0.063***
	(0.013)			(0.019)	(0.020)

Mother's education

(0-9 years as reference category)

	(1)	(2)	(3)	(4)	(5)
10-12 years	0.091***	0.119***	0.066***	0.086***	0.086***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
13+ years	0.170***	0.220***	0.109***	0.145***	0.145***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Missing	0.035***	0.052***	0.026***	0.014	0.015
	(0.005)	(0.012)	(0.005)	(0.016)	(0.016)

Father's education

(0-9 years as reference category)

	(1)	(2)	(3)	(4)	(5)
10-12 years	0.090***	0.105***	0.068***	0.076***	0.077***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
13+ years	0.126***	0.143***	0.084***	0.093***	0.093***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Missing	0.035***	0.013	0.026***	-0.011	-0.011

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(0.005)	(0.009)	(0.006)	(0.014)	(0.014)		
Mother's education · immigrant background							
(0-9 years as reference category)							
10-12 years · immigrant background				-0.064***		-0.041***	-0.045***
				(0.008)		(0.008)	(0.008)
13+ years · immigrant background				-0.151***		-0.095***	-0.098***
				(0.007)		(0.008)	(0.008)
Missing · immigrant background				-0.060***		-0.015	-0.020
				(0.013)		(0.017)	(0.017)
Father's education · immigrant background							
(0-9 years as reference category)							
10-12 years · immigrant background				-0.055***		-0.035***	-0.037***
				(0.009)		(0.009)	(0.009)
13+ years · immigrant background				-0.067***		-0.033***	-0.035***
				(0.008)		(0.008)	(0.008)
Missing · immigrant background				0.005		0.027*	0.018
				(0.010)		(0.015)	(0.015)
Mother's years since migration							
(30 or more years as reference category)							
0-4 years							-0.070***
							(0.021)
5-9 years							-0.047**
							(0.021)
10-14 years							0.007
							(0.020)
15-19 years							0.008
							(0.018)
20-24 years							-0.025
							(0.019)
25-29 years							-0.011
							(0.020)
Missing							-0.068***
							(0.019)
Father's years since migration							
(30 or more years as reference category)							
0-4 years							-0.073***
							(0.019)
5-9 years							-0.045**
							(0.018)
10-14 years							-0.026
							(0.017)
15-19 years							-0.018
							(0.016)
20-24 years							-0.021
							(0.016)
25-29 years							-0.011
							(0.016)
Missing							-0.067***
							(0.016)
Constant	0.183***	0.131***	0.267***	0.231***	0.120***	0.073**	0.076**
	-0.0334	(0.034)	-0.0327	(0.033)	-0.0331	(0.034)	(0.034)
Observations	144,370	144,370	144,370	144,370	144,370	144,370	144,370
R-squared	0.131	0.133	0.121	0.125	0.141	0.144	0.145

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2 in paper). The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, in specification (2), (4), (6), and YSM the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

F. Full estimation results for table 10 (OLS estimation results after allowing for gender interaction)

Dependent variable: Completion of upper secondary education within five years of compulsory school

	(1)	(2)	(3)	(4)
Variables of interest				
Immigrant background	-0.138***	-0.167***	-0.029***	-0.039***
	(0.004)	(0.005)	(0.011)	(0.014)
Immigrant background · female		0.058***		0.021
		(0.007)		(0.018)
Immigrant background · born abroad	-0.190***	-0.156***	-0.145***	-0.113***
	(0.011)	(0.015)	(0.011)	(0.016)
Immigrant background · born abroad · female		-0.069***		-0.066***
		(0.022)		(0.023)
Immigrant background · trend	0.004***	0.004***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · trend · female		0.002		0.003*
		(0.001)		(0.001)
Immigrant background · born abroad · trend	0.004***	0.002*	0.003***	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Immigrant background · born abroad · trend · female		0.004**		0.003*
		(0.002)		(0.002)
Female	0.103***	0.079***	0.107***	0.084***
	(0.003)	(0.003)	(0.002)	(0.003)
Graduating cohort				
(Cohort 1990 as reference category)				
Cohort 1991	-0.010	-0.010	-0.011	-0.011
	(0.008)	(0.008)	(0.007)	(0.007)
Cohort 1992	0.005	0.005	-0.001	-0.002
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1993	0.030***	0.031***	0.026***	0.027***
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1994	0.017**	0.017**	0.008	0.008
	(0.008)	(0.008)	(0.007)	(0.007)
Cohort 1995	0.013	0.013*	0.014*	0.014*
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1996	0.003	0.003	0.008	0.008
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1997	-0.001	-0.002	0.003	0.003
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1998	0.002	0.001	0.001	0.001
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 1999	-0.002	-0.002	-0.013	-0.013
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2000	0.003	0.003	-0.003	-0.003
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2001	0.002	0.002	-0.011	-0.011
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2002	0.017**	0.017**	-0.001	0.000
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2003	0.013*	0.014*	-0.004	-0.004
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2004	0.003	0.003	-0.012	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2005	0.030***	0.030***	0.006	0.006
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2006	0.034***	0.034***	0.008	0.009
	(0.008)	(0.008)	(0.008)	(0.008)
Cohort 2007	0.055***	0.056***	0.025***	0.026***
	(0.008)	(0.008)	(0.008)	(0.008)
Source countries				
(Born in Denmark as reference category)				
Turkey	-0.244***	-0.289***	0.045***	0.011
	(0.010)	(0.014)	(0.015)	(0.019)
Pakistan	-0.124***	-0.157***	0.123***	0.101***
	(0.011)	(0.015)	(0.016)	(0.020)
Lebanon	-0.229***	-0.236***	0.064***	0.062***

	(1)	(2)	(3)	(4)
	(0.013)	(0.017)	(0.017)	(0.021)
former Yugoslavia	-0.182***	-0.212***	0.066***	0.040**
	(0.012)	(0.016)	(0.016)	(0.021)
Bosnia & Herzegovina	0.058***	0.028	0.252***	0.221***
	(0.013)	(0.018)	(0.017)	(0.022)
Iraq	-0.167***	-0.191***	0.094***	0.077***
	(0.014)	(0.019)	(0.017)	(0.022)
Vietnam	-0.028**	-0.026	0.210***	0.217***
	(0.013)	(0.018)	(0.017)	(0.023)
Iran	-0.004	-0.021	0.233***	0.219***
	(0.014)	(0.019)	(0.018)	(0.023)
Sri Lanka	0.019	-0.019	0.205***	0.175***
	(0.014)	(0.020)	(0.018)	(0.024)
Morocco	-0.200***	-0.270***	0.092***	0.035
	(0.014)	(0.020)	(0.018)	(0.023)
Somalia	-0.281***	-0.311***	0.040**	0.021
	(0.015)	(0.020)	(0.019)	(0.024)
Afghanistan	-0.062***	-0.077***	0.205***	0.200***
	(0.017)	(0.023)	(0.020)	(0.026)
South Korea	0.065***	0.036	0.139***	0.102***
	(0.015)	(0.024)	(0.019)	(0.028)
Other countries	-0.116***	-0.137***	0.101***	0.084***
	(0.011)	(0.015)	(0.016)	(0.020)

Source countries · female
(Born in Denmark as reference category)

	(1)	(2)	(3)	(4)
Turkey · female			0.091***	0.069***
			(0.019)	(0.024)
Pakistan · female			0.067***	0.044*
			(0.020)	(0.025)
Lebanon · female			0.012	0.006
			(0.024)	(0.028)
former Yugoslavia · female			0.061***	0.050*
			(0.022)	(0.027)
Bosnia & Herzegovina · female			0.062**	0.062**
			(0.025)	(0.029)
Iraq · female			0.048*	0.036
			(0.027)	(0.030)
Vietnam · female			-0.007	-0.018
			(0.026)	(0.030)
Iran · female			0.031	0.023
			(0.027)	(0.031)
Sri Lanka · female			0.078***	0.060*
			(0.028)	(0.032)
Morocco · female			0.139***	0.115***
			(0.028)	(0.031)
Somalia · female			0.061**	0.041
			(0.029)	(0.033)
Afghanistan · female			0.028	0.013
			(0.033)	(0.036)
South Korea · female			0.051*	0.069**
			(0.030)	(0.035)
Other countries · female			0.041**	0.034
			(0.020)	(0.025)

Mother's age when child is born
(under 18 years as reference category)

	(1)	(2)	(3)	(4)
18-22 years			0.022*	0.021*
			(0.011)	(0.011)
23-27 years			0.058***	0.057***
			(0.011)	(0.011)
28-32 years			0.081***	0.080***
			(0.012)	(0.012)
33-37 years			0.091***	0.089***
			(0.012)	(0.012)
38-42 years			0.087***	0.087***

	(0.014)	(0.014)
Over 42 years	-0.011	-0.013
	(0.023)	(0.023)
Missing	0.024	0.023
	(0.020)	(0.020)

Father's age when child is born
(under 18 years as reference category)

18-22 years	0.020	0.019
	(0.032)	(0.031)
23-27 years	0.060*	0.059*
	(0.031)	(0.031)
28-32 years	0.085***	0.083***
	(0.031)	(0.031)
33-37 years	0.086***	0.084***
	(0.032)	(0.031)
38-42 years	0.094***	0.092***
	(0.032)	(0.032)
Over 42 years	0.103***	0.100***
	(0.032)	(0.032)
Missing	0.060*	0.058*
	(0.033)	(0.033)

Number of siblings
(No siblings as reference category)

1 sibling	0.026***	0.026***
	(0.003)	(0.003)
2 siblings	0.023***	0.021***
	(0.004)	(0.004)
3 siblings	0.003	0.000
	(0.006)	(0.006)
4 or more siblings	-0.032***	-0.037***
	(0.007)	(0.007)
Missing	0.003	0.000
	(0.019)	(0.019)

If parents are divorced or not

Divorced	-0.092***	-0.092***
	(0.004)	0.004

Mother's socioeconomic status
(Outside labour force as reference category)

Managers and CEOs	0.122***	0.122***
	(0.016)	(0.016)
Employees on highest level	0.171***	0.171***
	(0.007)	(0.007)
Employees on medium level	0.139***	0.138***
	(0.006)	(0.006)
Employees on lowest level	0.095***	0.094***
	(0.006)	(0.006)
Unskilled worker	0.093***	0.093***
	(0.008)	(0.008)
Other employees	0.076***	0.076***
	(0.007)	(0.007)
Self-employed and assisting spouses	0.110***	0.109***
	(0.009)	(0.009)
Unemployed	0.016*	0.016*
	(0.009)	(0.009)
Enrolled in education	-0.022	-0.022
	(0.025)	(0.025)
Missing	0.117***	0.118***
	(0.024)	(0.024)

Father's socioeconomic status

(Outside labour force as reference category)

Managers and CEOs	0.205***	0.205***
	(0.010)	(0.010)
Employees on highest level	0.197***	0.196***
	(0.008)	(0.008)
Employees on medium level	0.159***	0.159***
	(0.008)	(0.008)
Employees on lowest level	0.095***	0.094***
	(0.007)	(0.007)
Unskilled worker	0.088***	0.087***
	(0.010)	(0.010)
Other employees	0.076***	0.075***
	(0.008)	(0.008)
Self-employed and assisting spouses	0.155***	0.155***
	(0.008)	(0.008)
Unemployed	0.018*	0.018*
	(0.011)	(0.011)
Enrolled in education	0.051	0.051
	(0.044)	(0.045)
Missing	0.122***	0.123***
	(0.018)	(0.018)

Mother's socioeconomic status · immigrant background

(Outside labour force as reference category)

Managers and CEOs · immigrant background	-0.033	0.047
	(0.053)	(0.079)
Employees on highest level · immigrant background	-0.034**	-0.019
	(0.015)	(0.021)
Employees on medium level · immigrant background	-0.038***	-0.034*
	(0.013)	(0.018)
Employees on lowest level · immigrant background	0.004	0.014
	(0.010)	(0.013)
Unskilled worker · immigrant background	-0.059***	-0.084***
	(0.017)	(0.021)
Other employees · immigrant background	-0.030***	-0.035***
	(0.010)	(0.012)
Self-employed and assisting spouses · immigrant background	-0.037**	-0.036*
	(0.015)	(0.019)
Unemployed · immigrant background	0.012	0.001
	(0.011)	(0.013)
Enrolled in education · immigrant background	0.045*	0.050*
	(0.027)	(0.029)
Missing · immigrant background	-0.154***	-0.141***
	(0.026)	(0.027)

Father's socioeconomic status · immigrant background

(Outside labour force as reference category)

Managers and CEOs · immigrant background	-0.168***	-0.137***
	(0.029)	(0.044)
Employees on highest level · immigrant background	-0.068***	-0.025
	(0.015)	(0.020)
Employees on medium level · immigrant background	-0.049***	-0.059***
	(0.015)	(0.021)
Employees on lowest level · immigrant background	0.013	0.019
	(0.010)	(0.013)
Unskilled worker · immigrant background	0.002	0.019
	(0.016)	(0.020)
Other employees · immigrant background	-0.006	-0.007
	(0.011)	(0.013)

Self-employed and assisting spouses · immigrant background	-0.073***	-0.087***
	(0.011)	(0.014)
Unemployed · immigrant background	0.016	0.000
	(0.013)	(0.015)
Enrolled in education · immigrant background	-0.063	-0.065
	(0.047)	(0.049)
Missing · immigrant background	-0.086***	-0.089***
	(0.019)	(0.021)
Mother's socioeconomic status · immigrant background · female		
(Outside labour force as reference category)		
Managers and CEOs · immigrant background · female		-0.140
		(0.102)
Employees on highest level · immigrant background · female		-0.031
		(0.027)
Employees on medium level · immigrant background · female		-0.006
		(0.022)
Employees on lowest level · immigrant background · female		-0.021
		(0.016)
Unskilled worker · immigrant background · female		0.053*
		(0.029)
Other employees · immigrant background · female		0.007
		(0.014)
Self-employed and assisting spouses · immigrant background · female		0.000
		(0.024)
Unemployed · immigrant background · female		0.025*
		(0.014)
Enrolled in education · immigrant background · female		-0.010
		(0.021)
Missing · immigrant background · female		-0.031
		(0.021)
Father's socioeconomic status · immigrant background · female		
(Outside labour force as reference category)		
Managers and CEOs · immigrant background · female		-0.059
		(0.056)
Employees on highest level · immigrant background · female		-0.082***
		(0.025)
Employees on medium level · immigrant background · female		0.016
		(0.026)
Employees on lowest level · immigrant background · female		-0.010
		(0.015)
Unskilled worker · immigrant background · female		-0.033
		(0.026)
Other employees · immigrant background · female		0.005
		(0.014)
Self-employed and assisting spouses · immigrant background · female		0.027*
		(0.016)
Unemployed · immigrant background · female		0.034**
		(0.015)
Enrolled in education · immigrant background · female		0.005
		(0.030)
Missing · immigrant background · female		0.004
		(0.017)
Mother's education		
(0-9 years as reference category)		
10-12 years	0.086***	0.086***
	(0.005)	(0.005)
13+ years	0.145***	0.145***
	(0.005)	(0.005)

	(1)	(2)	(3)	(4)
Missing			0.014	0.014
			(0.016)	(0.016)
Father's education				
(0-9 years as reference category)				
10-12 years			0.076***	0.076***
			(0.005)	(0.005)
13+ years			0.093***	0.093***
			(0.005)	(0.005)
Missing			-0.011	-0.012
			(0.014)	(0.014)
Mother's education · immigrant background				
(0-9 years as reference category)				
10-12 years · immigrant background			-0.041***	-0.045***
			(0.008)	(0.010)
13+ years · immigrant background			-0.095***	-0.096***
			(0.008)	(0.010)
Missing · immigrant background			-0.015	-0.019
			(0.017)	(0.018)
Father's education · immigrant background				
(0-9 years as reference category)				
10-12 years · immigrant background			-0.035***	-0.035***
			(0.009)	(0.011)
13+ years · immigrant background			-0.033***	-0.028***
			(0.008)	(0.010)
Missing · immigrant background			0.027*	0.035**
			(0.015)	(0.017)
Mother's education · immigrant background · female				
(0-9 years as reference category)				
10-12 years · immigrant background · female				0.008
				(0.013)
13+ years · immigrant background · female				0.001
				(0.012)
Missing · immigrant background · female				0.009
				(0.012)
Father's education · immigrant background · female				
(0-9 years as reference category)				
10-12 years · immigrant background · female				0.000
				(0.015)
13+ years · immigrant background · female				-0.012
				(0.012)
Missing · immigrant background · female				-0.017
				(0.014)
Constant	0.590***	0.601***	0.073**	0.088***
	(0.006)	(0.006)	(0.034)	(0.033)
Observations	144,370	144,370	144,370	144,370
R-squared	0.063	0.065	0.144	0.146

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 (Sample selection described in notes to table 2 in paper). The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, in specification (3) and (4) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The coefficient on *immigrant background · female* in specification (2) and (4) is computed as the average of the coefficients on the source countries interacted with gender and weighted by the immigrant female population shares. In addition in specification (4) the coefficient on *immigrant background · female* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native female frequency distribution. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

G. Full estimation results for table 12 (OLS estimation results after controlling for grades)

Dependent variable: Completion of upper secondary education within five years of compulsory school

	(1)	(2)	(3)	(4)	(5)	Natives	Immigrants
Variables of interest							
Immigrant background	-0.062***	0.050***	0.049***	0.0371*	0.068***		
	(0.004)	(0.003)	(0.012)	(0.018)	(0.016)		
Immigrant background · born abroad	-0.113***	-0.031***	-0.030***	-0.078***	-0.024***		-0.018**
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)		(0.007)
Female	0.104***	0.042***	0.042***	0.111***	0.050***	0.022***	0.083***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Grade point average							
(Below 02 as reference category)							
Between 02 and 4		0.178***	0.147***		0.176***	0.121***	0.174***
		(0.008)	(0.018)		(0.008)	(0.018)	(0.009)
Between 4 and 6		0.474***	0.402***		0.456***	0.340***	0.497***
		(0.008)	(0.018)		(0.008)	(0.018)	(0.009)
Between 6 and 8		0.720***	0.664***		0.681***	0.574***	0.718***
		(0.008)	(0.017)		(0.008)	(0.018)	(0.009)
Between 8 and 10		0.820***	0.777***		0.762***	0.665***	0.779***
		(0.008)	(0.017)		(0.008)	(0.018)	(0.010)
Between 10 and 12		0.814***	0.779***		0.744***	0.656***	0.732***
		(0.011)	(0.018)		(0.011)	(0.019)	(0.025)
Grade point average · immigrant background							
(Below 02 as reference category)							
Between 02 and 4 · immigrant background			0.030				
			(0.020)				
Between 4 and 6 · immigrant background			0.105***				
			(0.020)				
Between 6 and 8 · immigrant background			0.070***				
			(0.019)				
Between 8 and 10 · immigrant background			0.029				
			(0.019)				
Between 10 and 12 · immigrant background			-0.019				
			(0.031)				
Graduating cohort							
(Cohort 2002 as reference category)							
Cohort 2003	0.003	0.004	0.004	0.001	0.003	0.003	0.005
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.010)
Cohort 2004	-0.001	0.003	0.003	-0.001	0.003	-0.010	0.017*
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.010)
Cohort 2005	0.023***	0.033***	0.033***	0.015**	0.029***	0.012	0.049***
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.010)
Cohort 2006	0.030***	0.066***	0.066***	0.020***	0.059***	0.041***	0.080***
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.010)
Cohort 2007	0.055***	0.071***	0.071***	0.040***	0.065***	0.039***	0.093***
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)
Source countries							
(Born in Denmark as reference category)							
Turkey	-0.155***	0.064***	0.000	0.202***	0.264***		*(Turkey as reference)*
	(0.007)	(0.007)	(0.019)	(0.020)	(0.018)		
Pakistan	-0.042***	0.055***	-0.012	0.272***	0.243***		-0.013
	(0.012)	(0.010)	(0.021)	(0.023)	(0.020)		(0.012)
Lebanon	-0.172***	0.012	-0.052***	0.178***	0.209***		-0.053***
	(0.012)	(0.010)	(0.020)	(0.023)	(0.020)		(0.013)
former Yugoslavia	-0.115***	0.012	-0.054**	0.183***	0.198***		-0.067***

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.014)	(0.012)	(0.021)	(0.024)	(0.021)	(0.013)
Bosnia & Herzegovina	0.107***	0.088***	0.021	0.344***	0.258***	-0.006
	(0.013)	(0.011)	(0.022)	(0.024)	(0.021)	(0.013)
Iraq	-0.069***	0.071***	0.006	0.239***	0.254***	-0.006
	(0.014)	(0.012)	(0.021)	(0.024)	(0.021)	(0.014)
Vietnam	0.042***	0.026**	-0.040*	0.329***	0.218***	-0.043***
	(0.015)	(0.013)	(0.022)	(0.025)	(0.022)	(0.015)
Iran	0.031*	0.059***	-0.007	0.323***	0.259***	-0.002
	(0.016)	(0.014)	(0.023)	(0.025)	(0.022)	(0.016)
Sri Lanka	0.091***	0.089***	0.022	0.339***	0.261***	0.007
	(0.015)	(0.013)	(0.022)	(0.025)	(0.022)	(0.015)
Morocco	-0.107***	0.046***	-0.020	0.235***	0.240***	-0.016
	(0.019)	(0.017)	(0.025)	(0.027)	(0.024)	(0.018)
Somalia	-0.148***	0.056***	-0.008	0.208***	0.260***	0.003
	(0.018)	(0.016)	(0.024)	(0.027)	(0.024)	(0.018)
Afghanistan	0.053***	0.141***	0.075***	0.364***	0.322***	0.060***
	(0.017)	(0.015)	(0.023)	(0.026)	(0.023)	(0.017)
South Korea	0.082***	0.022	-0.043*	0.240***	0.184***	-0.056***
	(0.019)	(0.017)	(0.025)	(0.029)	(0.026)	(0.020)
Other countries	-0.058***	0.011	-0.055***	0.217***	0.202***	-0.056***
	(0.009)	(0.008)	(0.020)	(0.022)	(0.020)	(0.011)
Mother's age when child is born						
(under 18 years as reference category)						
18-22 years				0.031	0.034*	0.009	0.050**
				(0.021)	(0.019)	(0.067)	(0.020)
23-27 years				0.062***	0.054***	0.070	0.057***
				(0.021)	(0.019)	(0.068)	(0.020)
28-32 years				0.092***	0.073***	0.092	0.073***
				(0.021)	(0.019)	(0.068)	(0.021)
33-37 years				0.103***	0.082***	0.106	0.072***
				(0.022)	(0.020)	(0.068)	(0.022)
38-42 years				0.091***	0.062***	0.096	0.041
				(0.024)	(0.022)	(0.069)	(0.026)
Over 42 years				0.031	0.048	0.060	0.056
				(0.040)	(0.034)	(0.081)	(0.042)
Missing				0.044	0.026	-0.026	0.071
				(0.039)	(0.036)	(0.083)	(0.045)
Father's age when child is born						
(under 18 years as reference category)						
18-22 years				0.059	-0.017	-0.135	0.022
				(0.064)	(0.063)	(0.115)	(0.076)
23-27 years				0.110*	0.015	-0.097	0.049
				(0.064)	(0.063)	(0.115)	(0.076)
28-32 years				0.134**	0.025	-0.085	0.051
				(0.064)	(0.063)	(0.115)	(0.076)
33-37 years				0.140**	0.026	-0.080	0.043
				(0.064)	(0.063)	(0.115)	(0.076)
38-42 years				0.143**	0.021	-0.074	0.031
				(0.065)	(0.063)	(0.115)	(0.077)
Over 42 years				0.153**	0.021	-0.070	0.035
				(0.065)	(0.064)	(0.116)	(0.077)
Missing				0.133**	0.019	-0.092	0.045
				(0.067)	(0.065)	(0.118)	(0.080)
Number of siblings						
(No siblings as reference category)						
1 sibling				0.020***	0.007	0.010*	0.011
				(0.005)	(0.005)	(0.006)	(0.008)
2 siblings				0.030***	0.017***	0.014*	0.023***

		(0.006)	(0.006)	(0.008)	(0.009)
3 siblings		0.007	0.006	-0.016	0.014
		(0.009)	(0.008)	(0.014)	(0.011)
4 or more siblings		-0.012	0.010	0.031	0.006
		(0.012)	(0.010)	(0.029)	(0.012)
Missing		0.016	0.013	-0.060	0.031
		(0.037)	(0.033)	(0.106)	(0.035)

If parents are divorced or not

	(1)	(2)	(3)	(4)
Divorced	-0.074***	-0.054***	-0.061***	-0.044***
	(0.006)	(0.005)	(0.007)	(0.009)

Mother's socioeconomic status
(Outside labour force as reference category)

	(1)	(2)	(3)	(4)
Managers and CEOs	0.122***	0.071***	0.069***	0.070
	(0.022)	(0.020)	(0.020)	(0.054)
Employees on highest level	0.179***	0.093***	0.087***	0.009
	(0.013)	(0.011)	(0.011)	(0.016)
Employees on medium level	0.135***	0.085***	0.083***	0.015
	(0.012)	(0.011)	(0.011)	(0.013)
Employees on lowest level	0.091***	0.073***	0.072***	0.018**
	(0.011)	(0.010)	(0.010)	(0.009)
Other employees	0.080***	0.064***	0.064***	0.009
	(0.013)	(0.012)	(0.012)	(0.008)
Self-employed and assisting spouses	0.105***	0.064***	0.060***	0.007
	(0.017)	(0.015)	(0.015)	(0.016)
Unemployed	-0.004	0.012	0.013	0.013
	(0.020)	(0.018)	(0.018)	(0.010)
Enrolled in education	0.013	0.019	0.017	0.022*
	(0.041)	(0.037)	(0.037)	(0.012)
Missing	0.205***	0.129***	0.176***	-0.056
	(0.053)	(0.048)	(0.051)	(0.044)

Father's socioeconomic status
(Outside labour force as reference category)

	(1)	(2)	(3)	(4)
Managers and CEOs	0.182***	0.094***	0.097***	-0.062*
	(0.016)	(0.014)	(0.014)	(0.033)
Employees on highest level	0.190***	0.088***	0.089***	0.012
	(0.014)	(0.012)	(0.013)	(0.015)
Employees on medium level	0.166***	0.092***	0.096***	0.050***
	(0.014)	(0.013)	(0.013)	(0.017)
Employees on lowest level	0.094***	0.064***	0.069***	0.041***
	(0.013)	(0.012)	(0.012)	(0.009)
Other employees	0.088***	0.055***	0.059***	0.018**
	(0.013)	(0.012)	(0.012)	(0.009)
Self-employed and assisting spouses	0.146***	0.093***	0.095***	0.003
	(0.015)	(0.013)	(0.013)	(0.010)
Unemployed	0.027	0.019	0.020	0.009
	(0.023)	(0.021)	(0.020)	(0.011)
Enrolled in education	0.081	0.029	0.038	0.002
	(0.066)	(0.065)	(0.065)	(0.018)
Missing	0.120***	0.055*	0.063*	-0.011
	(0.035)	(0.032)	(0.034)	(0.027)

Mother's socioeconomic status · immigrant background
(Outside labour force as reference category)

	(1)	(2)	(3)	(4)
Managers and CEOs · immigrant background	-0.025	-0.002		
	(0.062)	(0.056)		
Employees on highest level · immigrant background	-0.074***	-0.084***		
	(0.023)	(0.019)		
Employees on medium level · immigrant background	-0.033*	-0.067***		
	(0.020)	(0.017)		

	(1)	(2)	(3)	(4)
Employees on lowest level · immigrant background	-0.005	-0.048***		
	(0.015)	(0.014)		
Other employees · immigrant background	-0.047***	-0.051***		
	(0.016)	(0.014)		
Self-employed and assisting spouses · immigrant background	-0.061**	-0.056**		
	(0.025)	(0.022)		
Unemployed · immigrant background	0.028	0.004		
	(0.023)	(0.021)		
Enrolled in education · immigrant background	-0.005	0.004		
	(0.043)	(0.039)		
Missing · immigrant background	-0.228***	-0.138***		
	(0.056)	(0.051)		

Father's socioeconomic status · immigrant background
(Outside labour force as reference category)

	(1)	(2)	(3)	(4)
Managers and CEOs · immigrant background	-0.222***	-0.158***		
	(0.042)	(0.036)		
Employees on highest level · immigrant background	-0.084***	-0.073***		
	(0.022)	(0.019)		
Employees on medium level · immigrant background	-0.048**	-0.037*		
	(0.024)	(0.021)		
Employees on lowest level · immigrant background	0.000	-0.017		
	(0.017)	(0.015)		
Other employees · immigrant background	-0.032*	-0.030**		
	(0.017)	(0.015)		
Self-employed and assisting spouses · immigrant background	-0.092***	-0.083***		
	(0.019)	(0.017)		
Unemployed · immigrant background	0.009	-0.005		
	(0.026)	(0.024)		
Enrolled in education · immigrant background	-0.099	-0.027		
	(0.070)	(0.067)		
Missing · immigrant background	-0.114***	-0.057*		
	(0.036)	(0.033)		

Mother's education
(0-9 years as reference category)

	(1)	(2)	(3)	(4)
10-12 years	0.100***	0.047***	0.054***	0.003
	(0.011)	(0.010)	(0.010)	(0.008)
13+ years	0.167***	0.077***	0.084***	0.007
	(0.011)	(0.010)	(0.010)	(0.008)
Missing	-0.041	-0.018	-0.021	-0.001
	(0.041)	(0.036)	(0.036)	(0.009)

Father's education
(0-9 years as reference category)

	(1)	(2)	(3)	(4)
10-12 years	0.093***	0.034***	0.038***	0.002
	(0.010)	(0.009)	(0.009)	(0.010)
13+ years	0.120***	0.052***	0.055***	0.020**
	(0.009)	(0.009)	(0.009)	(0.008)
Missing	0.019	0.011	0.012	-0.001
	(0.028)	(0.026)	(0.026)	(0.010)

Mother's education · immigrant background
(0-9 years as reference category)

	(1)	(2)	(3)	(4)
10-12 years · immigrant background	-0.060***	-0.043***		
	(0.014)	(0.013)		
13+ years · immigrant background	-0.112***	-0.069***		
	(0.014)	(0.013)		
Missing · immigrant background	0.042	0.010		
	(0.042)	(0.037)		

Father's education · immigrant background
(0-9 years as reference category)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
10-12 years · immigrant background				-0.068***	-0.028**		
				(0.015)	(0.013)		
13+ years · immigrant background				-0.059***	-0.030**		
				(0.013)	(0.012)		
Missing · immigrant background				-0.002	-0.012		
				(0.031)	(0.028)		
Constant	0.626***	0.073***	0.127***	-0.014	-0.211***	-0.003	-0.032
	(0.006)	(0.009)	(0.017)	(0.067)	(0.066)	(0.111)	(0.078)
Observations	51,360	51,360	51,360	51,360	51,360	27,192	24,168
R-squared	0.046	0.284	0.285	0.116	0.300	0.269	0.318

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2 in paper). Grade information is available for 93 percent of the native subsample and 89 percent of the immigrant subsample. GPA is based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades (*standpunktskarakterer*) in ninth grade are available, and for these students GPA is measured based on these. The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. Additionally, in specification (3) the coefficient of *immigrant background* is averaged across six grade point average levels plus missing weighted by the native frequency distribution. In addition, in specification (4) and (5) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

H. OLS estimation results behind figure 9

Dependent variable: Completion of upper secondary education within five years of compulsory school

	Natives	Immigrants
Immigrant background · born abroad		-0.028***
		(0.007)
Female	0.008	0.081***
	(0.005)	(0.005)
Grade point average		
(Below 02 as reference category)		
Between 02 and 4	0.149***	0.176***
	(0.018)	(0.009)
Between 4 and 6	0.404***	0.503***
	(0.017)	(0.009)
Between 6 and 8	0.670***	0.729***
	(0.017)	(0.009)
Between 8 and 10	0.784***	0.797***
	(0.017)	(0.009)
Between 10 and 12	0.789***	0.753***
	(0.018)	(0.025)
Graduating cohort		
(Cohort 2002 as reference category)		
Cohort 2003	0.004	0.006
	(0.009)	(0.010)
Cohort 2004	-0.009	0.018*
	(0.009)	(0.010)
Cohort 2005	0.018**	0.051***
	(0.009)	(0.010)
Cohort 2006	0.053***	0.083***
	(0.009)	(0.010)
Cohort 2007	0.050***	0.096***
	(0.008)	(0.009)
Source countries		
(Turkey as reference category)		
Pakistan		-0.011
		(0.012)
Lebanon		-0.053***
		(0.011)
former Yugoslavia		-0.054***
		(0.013)
Bosnia & Herzegovina		0.022*
		(0.012)
Iraq		0.004
		(0.013)
Vietnam		-0.037**
		(0.014)
Iran		-0.005
		(0.015)
Sri Lanka		0.026*
		(0.014)
Morocco		-0.019
		(0.018)
Somalia		-0.009
		(0.017)
Afghanistan		0.073***
		(0.016)
South Korea		-0.037**
		(0.018)

Other countries		-0.055***
		(0.010)
Constant	0.150***	0.095***
	(0.018)	(0.011)
Observations	27,192	24,168
R-squared	0.239	0.313

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2). Grade information is available for 93 percent of the native subsample and 89 percent of the immigrant subsample. GPA is based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades (*standpunktskarakterer)* in ninth grade are available, and for these students GPA is measured based on these. The regressions use the linear probability model. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

I. Full estimation results for table 13 (OLS estimation results after controlling for grades and gender interaction)

Dependent variable: Completion of upper secondary education within five years of compulsory school

	(1)	(2)
Variables of interest		
Immigrant background	0.050***	0.015***
	(0.003)	(0.004)
Immigrant background · female		0.071***
		(0.006)
Immigrant background · born abroad	-0.031***	-0.034***
	(0.007)	(0.010)
Immigrant background · born abroad · female		0.006
		(0.014)
Female	0.042***	0.008
	(0.004)	(0.005)
Grade point average		
(Below 02 as reference category)		
Between 02 and 4	0.178***	0.175***
	(0.008)	(0.008)
Between 4 and 6	0.474***	0.469***
	(0.008)	(0.008)
Between 6 and 8	0.720***	0.716***
	(0.008)	(0.008)
Between 8 and 10	0.820***	0.818***
	(0.008)	(0.008)
Between 10 and 12	0.814***	0.813***
	(0.011)	(0.011)
Graduating cohort		
(Cohort 2002 as reference category)		
Cohort 2003	0.004	0.005
	(0.007)	(0.007)
Cohort 2004	0.003	0.003
	(0.007)	(0.007)
Cohort 2005	0.033***	0.033***
	(0.006)	(0.006)
Cohort 2006	0.066***	0.066***
	(0.006)	(0.006)
Cohort 2007	0.071***	0.072***
	(0.006)	(0.006)
Source countries		
(Born in Denmark as reference category)		
Turkey	0.064***	0.005
	(0.007)	(0.009)
Pakistan	0.055***	-0.006
	(0.010)	(0.015)
Lebanon	0.012	-0.012
	(0.010)	(0.015)
former Yugoslavia	0.012	-0.016
	(0.012)	(0.017)
Bosnia & Herzegovina	0.088***	0.074***
	(0.011)	(0.016)
Iraq	0.071***	0.042**
	(0.012)	(0.017)
Vietnam	0.026**	0.023
	(0.013)	(0.019)
Iran	0.059***	0.050***

	(0.014)	(0.019)
Sri Lanka	0.089***	0.059***
	(0.013)	(0.018)
Morocco	0.046***	-0.014
	(0.017)	(0.024)
Somalia	0.056***	-0.004
	(0.016)	(0.021)
Afghanistan	0.141***	0.120***
	(0.015)	(0.020)
South Korea	0.022	0.013
	(0.017)	(0.026)
Other countries	0.011	-0.018
	(0.008)	(0.012)
Source countries · female		
(Born in Denmark as reference category)		
Turkey · female		0.116***
		(0.013)
Pakistan · female		0.118***
		(0.020)
Lebanon · female		0.047**
		(0.020)
former Yugoslavia · female		0.056**
		(0.023)
Bosnia & Herzegovina · female		0.029
		(0.022)
Iraq · female		0.058**
		(0.024)
Vietnam · female		0.006
		(0.026)
Iran · female		0.017
		(0.028)
Sri Lanka · female		0.063**
		(0.025)
Morocco · female		0.120***
		(0.034)
Somalia · female		0.128***
		(0.032)
Afghanistan · female		0.044
		(0.030)
South Korea · female		0.019
		(0.034)
Other countries · female		0.057***
		(0.016)
Constant	0.073***	0.093***
	(0.009)	(0.009)
Observations	51,360	51,360
R-squared	0.284	0.286

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 2002 and 2007 with grade information available (Sample selection described in notes to table 2). Grade information is available for 93 percent of the native subsample and 89 percent of the immigrant. GPA is based on exam grades in ninth grade. For 756 out of the 51,360 students only continuous assessment grades in ninth grade are available, and for these students GPA is measured based on these. The regressions use the linear probability model. In specification (2), the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. The coefficient on *immigrant background · female* is computed as the average of the coefficients on the source countries interacted with gender and weighted by the immigrant female population shares. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

J. Full estimation results for table 14 (OLS estimation results after controlling for age at arrival)

Dependent variable: Completion of upper secondary education within five years of compulsory school

	(1)	(2)	(3)	(4)
Age at arrival				
(Born in Denmark as reference category)				
0 years	-0.039**	-0.005	0.004	0.048*
	(0.020)	(0.028)	(0.029)	(0.025)
1 years	-0.088***	-0.059***	-0.051***	-0.019
	(0.016)	(0.022)	(0.018)	(0.015)
2 years	-0.074***	-0.052**	-0.058***	-0.026
	(0.016)	(0.023)	(0.020)	(0.017)
3 years	-0.071***	-0.055**	-0.062***	-0.045***
	(0.017)	(0.023)	(0.021)	(0.018)
4 years	-0.091***	-0.061***	-0.092***	-0.059***
	(0.016)	(0.022)	(0.018)	(0.015)
5 years	-0.090***	-0.050**	-0.080***	-0.031**
	(0.016)	(0.022)	(0.017)	(0.015)
6 years	-0.118***	-0.094***	-0.133***	-0.063***
	(0.016)	(0.022)	(0.018)	(0.015)
7 years	-0.106***	-0.099***	-0.112***	-0.038**
	(0.016)	(0.021)	(0.018)	(0.015)
8 years	-0.131***	-0.099***	-0.142***	-0.061***
	(0.015)	(0.021)	(0.018)	(0.016)
9 years	-0.156***	-0.137***	-0.156***	-0.042***
	(0.015)	(0.021)	(0.018)	(0.015)
10 years	-0.151***	-0.115***	-0.159***	-0.021
	(0.015)	(0.021)	(0.018)	(0.016)
11 years	-0.183***	-0.155***	-0.200***	-0.026
	(0.015)	(0.020)	(0.019)	(0.017)
12 years	-0.175***	-0.145***	-0.185***	-0.006
	(0.015)	(0.020)	(0.019)	(0.017)
13 years	-0.194***	-0.155***	-0.175***	0.002
	(0.015)	(0.020)	(0.022)	(0.019)
14 years	-0.240***	-0.189***	-0.254***	-0.057***
	(0.015)	(0.020)	(0.025)	(0.022)
15 years	-0.280***	-0.224***	-0.262***	-0.061**
	(0.015)	(0.021)	(0.031)	(0.029)
Age at arrival · female				
(Born in Denmark as reference category)				
0 years · female		-0.071*		
		(0.040)		
1 years · female		-0.058*		
		(0.031)		
2 years · female		-0.047		
		(0.033)		
3 years · female		-0.031		
		(0.033)		
4 years · female		-0.063*		
		(0.032)		
5 years · female		-0.083***		
		(0.032)		
6 years · female		-0.048		
		(0.032)		
7 years · female		-0.014		
		(0.031)		
8 years · female		-0.065**		
		(0.031)		
9 years · female		-0.039		
		(0.030)		
10 years · female		-0.074**		
		(0.030)		

	(1)	(2)	(3)	(4)
11 years · female		-0.058**		
		(0.029)		
12 years · female		-0.060**		
		(0.029)		
13 years · female		-0.079***		
		(0.029)		
14 years · female		-0.107***		
		(0.030)		
15 years · female		-0.117***		
		(0.030)		
Missing · female		-0.081**		
		(0.033)		
Immigrant background	-0.030***	-0.038***	-0.054***	0.052***
	(0.011)	(0.014)	(0.004)	(0.003)
Immigrant background · female		0.019		
		(0.018)		
Immigrant background · trend	0.004***	0.003***		
	(0.001)	(0.001)		
Immigrant background · trend · female		0.003**		
		(0.001)		
Immigrant background · born abroad · trend	0.001	0.000		
	(0.001)	(0.001)		
Immigrant background · born abroad · trend · female		0.003		
		(0.002)		
Female	0.106***	0.084***	0.103***	0.042***
	(0.002)	(0.003)	(0.004)	(0.004)

Graduating cohort

(Cohort 1990 as reference category)   *Cohort 2002 as reference category*

	(1)	(2)	(3)	(4)
Cohort 1991	-0.011	-0.011		
	(0.007)	(0.007)		
Cohort 1992	-0.001	-0.001		
	(0.008)	(0.008)		
Cohort 1993	0.028***	0.029***		
	(0.008)	(0.008)		
Cohort 1994	0.009	0.009		
	(0.007)	(0.007)		
Cohort 1995	0.015*	0.015**		
	(0.008)	(0.008)		
Cohort 1996	0.010	0.010		
	(0.008)	(0.008)		
Cohort 1997	0.006	0.006		
	(0.008)	(0.008)		
Cohort 1998	0.003	0.003		
	(0.008)	(0.008)		
Cohort 1999	-0.012	-0.012		
	(0.008)	(0.008)		
Cohort 2000	-0.002	-0.002		
	(0.008)	(0.008)		
Cohort 2001	-0.011	-0.012		
	(0.008)	(0.008)		
Cohort 2002	0.001	0.001		
	(0.008)	(0.008)		
Cohort 2003	-0.003	-0.002	0.003	0.005
	(0.008)	(0.008)	(0.008)	(0.007)
Cohort 2004	-0.011	-0.011	-0.002	0.003
	(0.008)	(0.008)	(0.008)	(0.007)
Cohort 2005	0.008	0.008	0.021***	0.033***
	(0.008)	(0.008)	(0.007)	(0.006)
Cohort 2006	0.010	0.010	0.027***	0.066***

	(0.008)	(0.008)	(0.007)	(0.006)
Cohort 2007	0.027***	0.028***	0.052***	0.072***
	(0.008)	(0.008)	(0.007)	(0.006)

Source countries

(Born in Denmark as reference category)

Turkey	0.058***	0.024	-0.155***	0.064***
	(0.015)	(0.019)	(0.007)	(0.007)
Pakistan	0.132***	0.110***	-0.046***	0.054***
	(0.016)	(0.020)	(0.012)	(0.010)
Lebanon	0.067***	0.066***	-0.188***	0.011
	(0.017)	(0.021)	(0.012)	(0.010)
former Yugoslavia	0.085***	0.059***	-0.114***	0.016
	(0.016)	(0.021)	(0.014)	(0.012)
Bosnia & Herzegovina	0.303***	0.269***	0.119***	0.102***
	(0.017)	(0.023)	(0.014)	(0.012)
Iraq	0.138***	0.119***	-0.038***	0.071***
	(0.017)	(0.023)	(0.014)	(0.012)
Vietnam	0.223***	0.229***	0.039**	0.027**
	(0.017)	(0.022)	(0.015)	(0.013)
Iran	0.250***	0.237***	0.033**	0.058***
	(0.018)	(0.023)	(0.016)	(0.014)
Sri Lanka	0.222***	0.193***	0.089***	0.089***
	(0.018)	(0.024)	(0.014)	(0.013)
Morocco	0.010***	0.043*	-0.108***	0.045***
	(0.018)	(0.023)	(0.019)	(0.017)
Somalia	0.075***	0.055**	-0.124***	0.064***
	(0.019)	(0.024)	(0.019)	(0.016)
Afghanistan	0.268***	0.257***	0.114***	0.139***
	(0.020)	(0.027)	(0.018)	(0.016)
South Korea	0.158***	0.121***	0.078***	0.022
	(0.019)	(0.028)	(0.019)	(0.017)
Other countries	0.128***	0.111***	-0.044***	0.012
	(0.016)	(0.020)	(0.010)	(0.008)

Source countries ·female

(Born in Denmark as reference category)

Turkey ·female	0.069***
	(0.024)
Pakistan ·female	0.043*
	(0.025)
Lebanon ·female	0.003
	(0.028)
former Yugoslavia ·female	0.051*
	(0.027)
Bosnia & Herzegovina ·female	0.067**
	(0.030)
Iraq ·female	0.040
	(0.030)
Vietnam ·female	-0.015
	(0.030)
Iran ·female	0.023
	(0.031)
Sri Lanka ·female	0.059*
	(0.032)
Morocco ·female	0.115***
	(0.031)
Somalia ·female	0.046
	(0.033)
Afghanistan ·female	0.024
	(0.036)
South Korea ·female	0.069**
	(0.035)
Other countries ·female	0.035
	(0.026)

Grade point average			
(Below 02 as reference category)			
Between 02 and 4			0.179***
			(0.008)
Between 4 and 6			0.475***
			(0.008)
Between 6 and 8			0.721***
			(0.008)
Between 8 and 10			0.822***
			(0.008)
Between 10 and 12			0.816***
			(0.011)
Mother's age when child is born			
(under 18 years as reference category)			
18-22 years	0.019*	0.019*	
	(0.011)	(0.011)	
23-27 years	0.054***	0.053***	
	(0.011)	(0.011)	
28-32 years	0.077***	0.076***	
	(0.012)	(0.012)	
33-37 years	0.085***	0.084***	
	(0.012)	(0.012)	
38-42 years	0.081***	0.080***	
	(0.014)	(0.014)	
Over 42 years	-0.008	-0.011	
	(0.023)	(0.023)	
Missing	0.029	0.028	
	(0.020)	(0.020)	
Father's age when child is born			
(under 18 years as reference category)			
18-22 years	0.012	0.011	
	(0.032)	(0.031)	
23-27 years	0.052*	0.051	
	(0.031)	(0.031)	
28-32 years	0.076**	0.075**	
	(0.031)	(0.031)	
33-37 years	0.077**	0.076**	
	(0.032)	(0.031)	
38-42 years	0.085***	0.083***	
	(0.032)	(0.032)	
Over 42 years	0.094***	0.091***	
	(0.032)	(0.032)	
Missing	0.052	0.050	
	(0.033)	(0.033)	
Number of siblings			
(No siblings as reference category)			
1 sibling	0.026***	0.026***	
	(0.003)	(0.003)	
2 siblings	0.022***	0.021***	
	(0.004)	(0.004)	
3 siblings	0.004	0.001	
	(0.006)	(0.006)	
4 or more siblings	-0.027***	-0.032***	
	(0.007)	(0.007)	
Missing	-0.012	-0.016	
	(0.019)	(0.019)	
If parents are divorced or not			
Divorced	-0.095***	-0.095***	

	(0.004)	(0.004)
Mother's socioeconomic status		
(Outside labour force as reference category)		
Managers and CEOs	0.122***	0.122***
	(0.016)	(0.016)
Employees on highest level	0.171***	0.171***
	(0.007)	(0.007)
Employees on medium level	0.139***	0.138***
	(0.006)	(0.006)
Employees on lowest level	0.094***	0.094***
	(0.006)	(0.006)
Unskilled worker	0.093***	0.093***
	(0.008)	(0.008)
Other employees	0.075***	0.075***
	(0.007)	(0.007)
Self-employed and assisting spouses	0.109***	0.109***
	(0.009)	(0.009)
Unemployed	0.016*	0.015*
	(0.009)	(0.009)
Enrolled in education	-0.022	-0.022
	(0.025)	(0.025)
Missing	0.114***	0.115***
	(0.024)	(0.024)
Father's socioeconomic status		
(Outside labour force as reference category)		
Managers and CEOs	0.205***	0.204***
	(0.010)	(0.010)
Employees on highest level	0.196***	0.196***
	(0.008)	(0.008)
Employees on medium level	0.158***	0.158***
	(0.008)	(0.008)
Employees on lowest level	0.094***	0.094***
	(0.007)	(0.007)
Unskilled worker	0.087***	0.087***
	(0.010)	(0.010)
Other employees	0.075***	0.075***
	(0.008)	(0.008)
Self-employed and assisting spouses	0.155***	0.155***
	(0.008)	(0.008)
Unemployed	0.018*	0.017
	(0.011)	(0.011)
Enrolled in education	0.051	0.051
	(0.044)	(0.045)
Missing	0.122***	0.123***
	(0.018)	(0.018)
Mother's socioeconomic status · immigrant background		
(Outside labour force as reference category)		
Managers and CEOs · immigrant background	-0.050	0.033
	(0.053)	(0.078)
Employees on highest level · immigrant background	-0.052***	-0.034
	(0.015)	(0.021)
Employees on medium level · immigrant background	-0.058***	-0.051***
	(0.013)	(0.018)
Employees on lowest level · immigrant background	-0.007	0.003
	(0.010)	(0.013)
Unskilled worker · immigrant background	-0.076***	-0.100***
	(0.017)	(0.021)
Other employees · immigrant background	-0.036***	-0.040***

	(0.010)	(0.012)
Self-employed and assisting spouses · immigrant background	-0.050***	-0.048**
	(0.015)	(0.019)
Unemployed · immigrant background	0.002	-0.008
	(0.011)	(0.013)
Enrolled in education · immigrant background	0.054**	0.058**
	(0.027)	(0.029)
Missing · immigrant background	-0.124***	-0.115***
	(0.026)	(0.027)
Father's socioeconomic status · immigrant background		
(Outside labour force as reference category)		
Managers and CEOs · immigrant background	-0.180***	-0.152***
	(0.029)	(0.044)
Employees on highest level · immigrant background	-0.080***	-0.036*
	(0.015)	(0.020)
Employees on medium level · immigrant background	-0.061***	-0.071***
	(0.015)	(0.021)
Employees on lowest level · immigrant background	0.005	0.013
	(0.010)	(0.013)
Unskilled worker · immigrant background	-0.006	0.012
	(0.016)	(0.020)
Other employees · immigrant background	-0.010	-0.011
	(0.011)	(0.013)
Self-employed and assisting spouses · immigrant background	-0.082***	-0.095***
	(0.011)	(0.014)
Unemployed · immigrant background	0.011	-0.004
	(0.013)	(0.015)
Enrolled in education · immigrant background	-0.034	-0.041
	(0.047)	(0.049)
Missing · immigrant background	-0.068***	-0.072***
	(0.019)	(0.021)
Mother's socioeconomic status · immigrant background · female		
(Outside labour force as reference category)		
Managers and CEOs · immigrant background · female		-0.142
		(0.102)
Employees on highest level · immigrant background · female		-0.035
		(0.027)
Employees on medium level · immigrant background · female		-0.011
		(0.022)
Employees on lowest level · immigrant background · female		-0.022
		(0.016)
Unskilled worker · immigrant background · female		0.051*
		(0.028)
Other employees · immigrant background · female		0.006
		(0.014)
Self-employed and assisting spouses · immigrant background · female		-0.004
		(0.024)
Unemployed · immigrant background · female		0.024*
		(0.014)
Enrolled in education · immigrant background · female		-0.007
		(0.021)
Missing · immigrant background · female		-0.021
		(0.021)
Father's socioeconomic status · immigrant background · female		
(Outside labour force as reference category)		
Managers and CEOs · immigrant background · female		-0.053
		(0.056)
Employees on highest level · immigrant background · female		-0.083***

		(0.025)
Employees on medium level · immigrant background · female		0.016
		(0.026)
Employees on lowest level · immigrant background · female		-0.013
		(0.015)
Unskilled worker · immigrant background · female		-0.034
		(0.025)
Other employees · immigrant background · female		0.003
		(0.014)
Self-employed and assisting spouses · immigrant background · female		0.025
		(0.016)
Unemployed · immigrant background · female		0.030**
		(0.015)
Enrolled in education · immigrant background · female		0.017
		(0.030)
Missing · immigrant background · female		0.007
		(0.017)

Mother's education
(0-9 years as reference category)

10-12 years	0.085***	0.085***
	(0.005)	(0.005)
13+ years	0.145***	0.145***
	(0.005)	(0.005)
Missing	0.014	0.013
	(0.016)	(0.016)

Father's education
(0-9 years as reference category)

10-12 years	0.076***	0.076***
	(0.005)	(0.005)
13+ years	0.093***	0.093***
	(0.005)	(0.005)
Missing	-0.011	-0.012
	(0.014)	(0.014)

Mother's education · immigrant background
(0-9 years as reference category)

10-12 years · immigrant background	-0.044***	-0.048***
	(0.008)	(0.010)
13+ years · immigrant background	-0.097***	-0.099***
	(0.008)	(0.010)
Missing · immigrant background	-0.020	-0.025
	(0.017)	(0.018)

Father's education · immigrant background
(0-9 years as reference category)

10-12 years · immigrant background	-0.034***	-0.035***
	(0.009)	(0.011)
13+ years · immigrant background	-0.033***	-0.028***
	(0.008)	(0.010)
Missing · immigrant background	0.029*	0.036**
	(0.015)	(0.017)

Mother's education · immigrant background · female
(0-9 years as reference category)

10-12 years · immigrant background · female		0.009
		(0.013)
13+ years · immigrant background · female		0.002
		(0.012)
Missing · immigrant background · female		0.012
		(0.012)

Father's education · immigrant background · female

(0-9 years as reference category)

10-12 years · immigrant background · female		0.001		
		(0.015)		
13+ years · immigrant background · female		-0.012		
		(0.012)		
Missing · immigrant background · female		-0.016		
		(0.014)		
Constant	0.086***	0.100***	0.628***	0.071***
	(0.033)	(0.033)	(0.006)	(0.009)
Observations	144,370	144,370	51,360	51,360
R-squared	0.147	0.148	0.049	0.285

NOTE: Robust standard errors are reported in parentheses. Administrative register information from Statistics Denmark for sample of non-Western immigrants and natives graduating ninth grade between 1990 and 2007 in specification (1) and (2) and between 2002 and 2007 with grade information available in specification (3) and (4) (Sample selection described in notes to table 2 in paper). The regressions use the linear probability model. In all specifications, the coefficient on *immigrant background* is computed as the average of the coefficients on the source countries weighted by the immigrant population shares. In addition, in specification (1) and (2) the coefficient on *immigrant background* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native frequency distribution. The coefficient on *immigrant background · female* in specification (2) is computed as the average of the coefficients on the source countries interacted with gender and weighted by the immigrant female population shares. In addition, the coefficient on *immigrant background · female* is averaged across ten socioeconomic status levels and three educational levels plus missing levels weighted by the native female frequency distribution. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.