An analysis of the relationship between immigrant generation and math performance in the Finnish school system

Magdalena Planting

Department of Sociology, Demography Unit (SUDA)

Master's Thesis 30 HE credits

Subject: Demography

Multidisciplinary Master's Programme in Demography (120 credits)

Spring term 2022

Supervisor: Eleonora Mussino

Stockholms universitet

An analysis of the relationship between immigrant generation and math performance in the Finnish school system

Magdalena Planting

Abstract

Finland is internationally recognized for its high-performing and equity promoting school system. However, students with an immigrant background perform worse than native student in the Finnish school system. According to previous research, the performance gap between the immigrant- and native-generation students is among the highest of the OECD countries. By building on to what is already known, this thesis aims to provide a more comprehensive understanding of the school performance of students with an immigrant background. This thesis descriptively analyzes the relationship between the immigrant generation and the math performance by performing OLS regressions in Stata using the OECD's PISA data. The relationship is studied over six PISA cycles, 2003 to 2018, compositional and background factors are controlled for and a detailed definition for immigrant generation is used, by including the 1,25-, 1,5-, 1,75-, 2-, 2,5- and native-generation. The results indicate that the immigrant-generation students perform worse than the native-generation students in all cycles and a positive performance trend is present among the immigrant generations, as the higher generations perform better. The results do not indicate that the performance gap between the native and immigrant generations is improving over time, rather fluctuating.

Keywords

Immigrant generation, Finland, school performance, PISA, age of arrival, immigration

Contents

List of graphs

List of tables

List of appendix figures

Introduction

A major driver of migration during the 21st century is the inequality between countries, as people from developing countries seek for better opportunities in the developed countries, resulting in a rise in the international migrant stock in many high-income countries (IOM, 2020; United Nations, 2019). In Finland, the share of the population with immigrant background is growing and the composition is changing over time, which raises the need for understanding the effect that immigration has on the receiving country. The less that immigrants lag behind natives, the more the receiving country can benefit from immigration. In most developed countries, including Finland, immigrants perform worse in school than natives, which has gained great attention in the research field (Behr and Fugger, 2020). The low school performance of immigrants can be costly for the society in the long run, as school performance is highly correlated with later life outcomes (Heckman, Jagelka & Krautz, 2021). The school system has the potential to lower the inequality between immigrants and natives, by investing in the equity in teaching that targets immigrants, to improve their social mobility. During the 21st century, Finland has become internationally recognized for its high-quality and equity promoting school system (Sahlberg, 2012). This thesis examines the relationship between immigrant generation and PISA math performance over time in the high-performing Finnish school system. In this thesis, the school system and school performance refer to the primary school level, unless otherwise stated.

As immigration grows, it becomes more important to develop the school system to serve students with a growing variety of cultural and linguistic backgrounds, in addition to social and economic. For children with an immigrant background, the school is the main environment to gain the feeling of belongingness in a society (Fandrem, Jahnsen, Nergaard & Tveitereid, 2021). In our multi-cultural world, a high-quality school system is a system that is inclusive toward immigrants, which is reflected in the performance of immigrant students (Chircop & Claros, 2019). Finland is reported to have the highest equity in the school system in the world, because of the low performance difference between the highest and lowest performing students (Schwab, 2017) and the low effect of socioeconomic background on students' performance in the PISA test (Sahlberg, 2012). But less focus is given to the school preformance of students with an immigrant background in the Finnish school system. The existing literature mainly studies the immigrants' school performance for a specific year or

two, see Kirjavainen and Pulkkinen (2017). The first contribution of the thesis is to examine the relationship between immigrant generation and PISA math score over time, from 2003 to 2018.

Researchers agree on that the school performance of students with an immigrant background is related to factors that they inherit from their parents and their country of origin. It is found that the socio-economic status and language spoken at home are the two major factors affecting immigrants' school performance (Kirjavainen and Pulkkinen, 2017). The second contribution of this thesis is to study how and if the relationship between immigrant generation and PISA math score changes when controlling for compositional and background factors, including: gender, language spoken at home, social, economic and cultural status.

A comprehensive understanding of how the different immigrant generations are related to school performance is beneficial for policy makers to improve the school system to serve multiculturalism. It is found that age of arrival is inversely related to school performance, as those students that arrive at a younger ager perform better (Heath and Kilpi-Jakonen, 2012). The existing studies commonly use simple classifications of the immigrant generations, such as the one provided by PISA, which includes only first- and second-generation immigrants, see Schleicher (2006). The third contribution of this thesis is to use a more detailed definition of immigrant generation by including the 1,25-, 1,5-, 1,75-, 2-, 2,5- and native-generation. The first generations are produced based on Rumbaut's (2014) age of arrival distribution and the second generations based on whether they have a native parent.

This thesis aims to answer two research questions (RQs): RQ1: How do the different immigrant-generation students perform on the PISA math test in Finland over time in relation to the native-generation students? RQ2: Is the relationship between the PISA math score and immigrant generation explained by the different compositional factors?

To study immigrants' school performance, this study uses the Program for International Student Assessment (PISA) data provided by the Organisation for Economic Co-operation and Development (OECD) to perform regressions in the statistical software Stata. The PISA test measures 15-year-old students' performance in reading, science and math, using internationally standardized tests. This study uses the PISA math score to account for school performance, instead of reading or science, to control for the large effect that language spoken

at home is found to have on school performance, as language proficiency is less important in math.

This thesis first covers some fundamental information about the school system and the immigration history in Finland in the Background section. It then explains the theory behind the school performance gap between natives and immigrants, gathers previous research on immigrants' school performance in Finland and presents the hypotheses in the Theory section. Then the thesis moves on to present the data and variables in the Data section, followed by the methods used, along with ethical considerations in the Methods section. Whereafter, the results are presented in the Results section answering the research questions and then explained and linked to theory based on the hypotheses in the Discussion section. Finally, the main findings of this thesis are highlighted, limitations are listed and suggestions about further research about the subject is given in the Conclusion section.

Background

The Finnish school system

The Finnish school system gained international recognition through its high performance in the first Program for International Student Assessment (PISA) test in 2000 and further by its high promotion of equity reflected in the low effect that the socio-economic status of the student has their school performance in the Finnish school system (Sahlberg, 2021). The high-performance and equity has motivated scientists to study the school system structure and determine which aspects of the system contribute to the success story of the Finnish school system.

In Finland today compulsory education starts at age six and ends latest at age 18 and is free for all students. To ensure that children are equally prepared for school when attending first class at age seven, pre-primary school became mandatory in 2015 in Finland (OKM, 2015). Children first attend pre-primary school for one year, whereafter they begin primary school, which lasts for nine years (OPS, 2022). Until fall 2021, compulsory education ended after the completion of primary school around age 15 or latest 10 years after the beginning of compulsory education, around age 17. Compulsory education was extended to end at age 18 or after completion of upper secondary education, as the Ministry of Education and Culture presumes that an upper secondary qualification is required to succeed in the advancing job market in the future (OKM, 2021). Due to the recent extension, students still have to attend, after the completion of primary school, either upper secondary school (gymnasium) or vocational education, which last, on average, three years (OPS, 2022). The effect, potentially positive, of the recent changes in the age of compulsory education that occurred in 2015 and 2021 could be seen in the near future but will not yet have an effect on this study, as the years examined are 2003 to 2018.

Among others, Pasi Sahlberg, a Finnish professor of education, has put great effort in answering the question about the Finnish school system's success story in the third edition of his award-winning book, "Finnish Lessons 3.0: What Can the World Learn from Educational Change in Finland?" (Sahlberg, 2021). In his book, Sahlberg stresses that behind the successful school system is a complex supporting network of educational, social, cultural and

political elements (p. 9). He highlights teacher professionalism, investment in equity and the small quality difference between schools. The teacher professionalism does not only mean that teachers are highly educated and sufficiently paid, but that their profession is highly respected in the society. Sahlberg explains that teachers in Finland are given great independency and authority in teaching, which is believed to produce the highest possible quality of teaching and student performance (p. 129). Sahlberg mentions that the Finnish school system differs from many other countries in that they do not use national standardized tests to measure the performance of students. Frequent standardized testing is believed in Finland to interrupt the students' learning process and bias teaching. Hence, all schools are independently responsible for determining the level of students' performance. (p. 56, 76).

The Finnish education policies have a central focus on equity, with the aim of not only ensuring school access for all, but more importantly, minimizing the effect that a student's background, mainly socioeconomic, has on their school performance. Every student in Finland is granted to achieve academic success no matter where they live, the level of education or income that their parents earn (Sahlberg, 2021, p. 73). The Finnish school system is unique in that the vast majority of the schools are public, which are also the best schools in the country (p. 173). As Finland has not followed the neoliberal education policies, the Global Educational Reform Movement, that have emerged along with globalization and adopted by majority of the developed countries and lately also by developing countries, with the aim to privatize the school system to reach higher quality and efficiency through the competition between schools (p. 176). Sahlberg stresses that the Finnish education policies alone are not able to produce the equity in the school system. The Finnish education policies are supported by multiple interrelated social policies in the welfare state, such as generous parental leave and free health care (p. 80).

The equity in the Finnish school system has been recognized by the OECD in the PISA test, where Finland has had the lowest performance difference between schools over the years, which also reflects the small quality difference between schools (p. 75). The Finnish school system has also been ranked to have the highest quality of primary education by the World Economic Forum in their annual Global Competitiveness Report. The quality of the primary education is based on an opinion survey produced by the World Economic Forum and is available annually from year 2008 to 2017 and the Finnish school system has ranked number one in all years (Schwab, 2008-2017). Samuelsson and Lindblad (2015) and Varjo, Lundström and Kalalahti (2018), both compare the Finnish and Swedish school systems,

which is motivated by the higher performing system in Finland. Both papers agree on that the divergence of the quality of the school systems has to do with the marketization of the Swedish school system, whereas Finland has followed the professional logic. The authors argue that the high performance of Finnish students on international tests is related to the system's strong emphasis on teacher professionalism, specifically, highly educated, respected and independent teachers.

Despite the high-quality school system, students with an immigrant background were found to lag significantly much behind the native students based on the PISA 2012 results (Harju-Luukkainen, Nissinen, Sulkunen, Suni & Vettenranta, 2014). According to the Ministry of Education and Culture (2016) the rise in immigration during the 21st century has challenged the school system and measures have been taken to meet the needs of immigrant students. Immigrant children living permanently in Finland are entitled to attended compulsory education just like native students. To support immigrant students' learning and participation in teaching, preparatory education is provided. The preparatory teaching includes 900 hours for children aged 10 and below and 1 000 hours for students older than 10. Students then attend school beginning from the grade that aligns with their skill level. As a substitute for the education of the native language, the Finnish school system offers teaching in the native language as a second language for immigrant students. The Ministry of Education and Culture finds it central to learn the native language for long term integration into the society but education in the mother tongue of immigrant students is also provided, which is found to support school performance and general behavior of the students. Immigrant students are also provided with remedial education to additionally ensure that they keep up with the learning pace. The Finnish school system constantly adapts to support immigrant students but the Ministry of Education and Culture stresses that the shortage of teachers is a serious obstacle, which becomes more severe as immigration grows and the immigrant population becomes more diverse. Additionally, the Finnish economy is not strong enough to increase the financial support for educational policies, which leads to that finances have to be redirected within the educational sector to improve the immigrant students' situation (Ministry of Education and Culture, 2016).

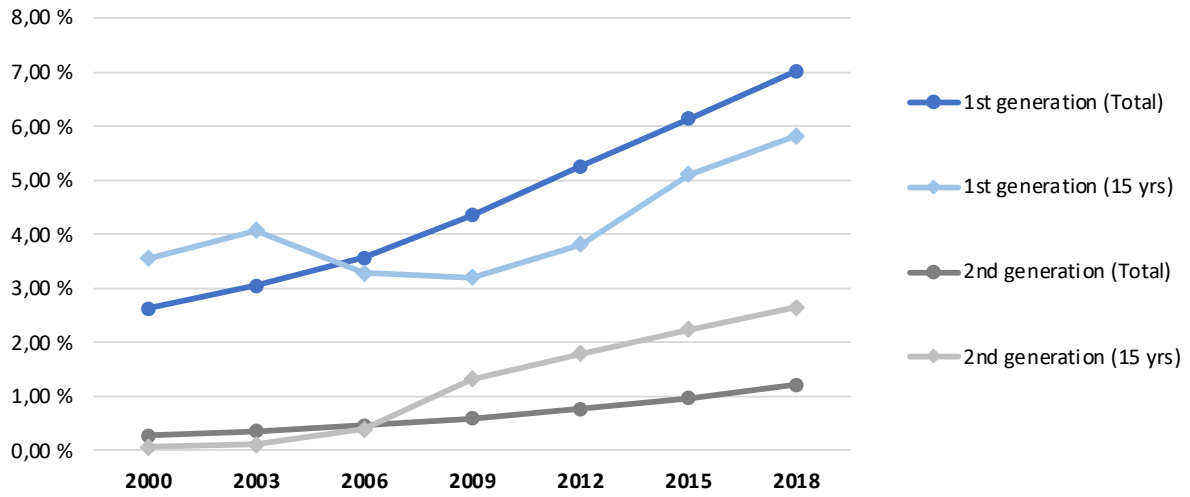
Immigration and integration of immigrants in Finland

The Finnish population was relatively homogenous during the 20th due to low levels of migration and negative net migration. After the second world war majority of the Finnish emigrants moved to Sweden in the search for better jobs due to the stronger economy in Sweden. The Finnish economy had recovered from the war and caught up with the Swedish economy in the 1980's, which slowed down the emigration from Finland. During the 1980's net migration turned positive and has increased with some fluctuations since year 2000. The initial rise in immigration was driven by the fall of the Soviet Union, the civil war in Somalia, the unstable situation in Iran and Iraq and the war in Yugoslavia (Korkisaari & Söderling, 2003; Tilastokeskus, 2004; Tilastokeskus, 2021b).

Graph 1 below shows immigrants' share of the total population in Finland for the first- and second-generation separately, from year 2000 to 2018, which covers the years analyzed in this thesis. The graph shows both the total immigrant-generation populations' share of the total Finnish population, shown with the darker colors and round markers, and the 15-year-old immigrant populations' share of the total 15-year-old Finnish population, shown with the lighter colors and diamond shaped markers, which is the true population studied in this thesis. The first generation includes those who are born abroad and the second generation those who are born in Finland but whose parents are born abroad. The graph shows that the first generation, as the share of the total population in Finland, has been steadily increasing since 2000 and reached 7 percent in 2018, but is still relatively low compared to other neighboring developed countries. According to the United Nation's population division, the share of the population born abroad in 2019 was 14 percent in Estonia, 16 percent in Norway and 20 percent in Sweden (United Nations, 2019). The second-generation population is smaller in relative size and increases slower than the first generation, due to the short history of immigration in Finland. The second generation can eventually in the long run catch up with the first generation size wise, but it depends much on the birth rate of the first-generation population in Finland. The 15-year-old immigrant populations, compared to the total immigrant populations, have developed similarly but with some difference. The first-generation population of the 15-year-olds falls below the total graph in 2006, which means that the first-generation population is underrepresented in that specific age group in relation to the total population. The average age of the first-generation population in Finland was 38

years in 2017 (Tilastokeskus, 2018). The second generation of the 15- year-olds rises above the total graph after 2006, which means that the second-generation population is overrepresented in that specific age group compared to the total population, which is logical in a country with a short history of immigration as the second-generation population still mostly consists of children. The average age of the second-generation population in Finland was 11 years in 2017 (Tilastokeskus, 2018).

Graph 1: The share of immigrants based on generation in the Finnish population (total and 15-year-olds).



Data: Statistics Finland.

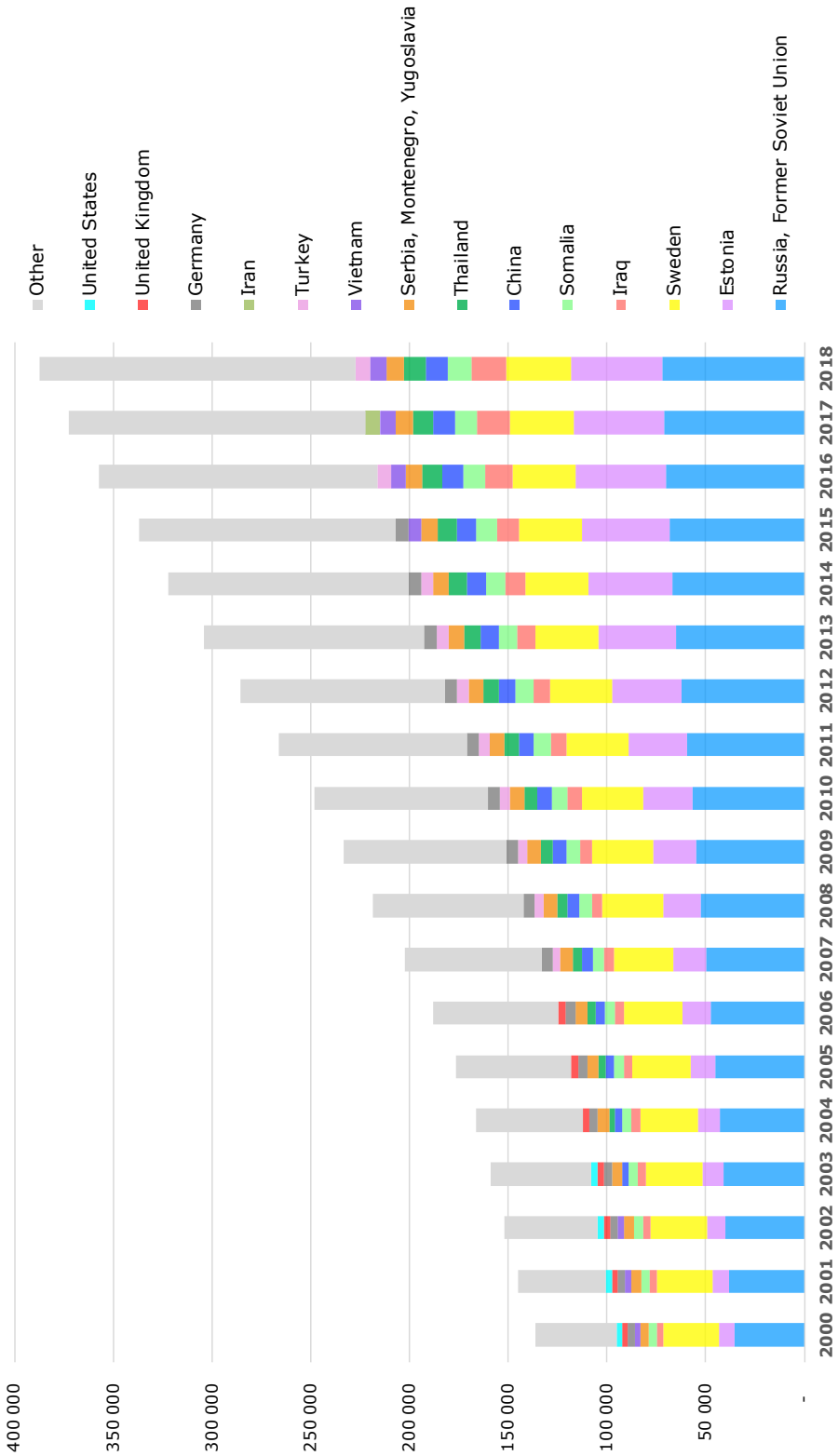
The low number of immigrants in Finland is mainly due to the strict immigration policies that lasted until the 1990's (Ansala, Åslunf & Sarvimäki, 2020), but also because of the geographical northern location and the long border with Russia (Koikkalainen, 2021). **Graph 2** below shows the number of first generation from the top ten countries of birth of immigrants in Finland year 2000 to 2018. The main change visible in the graph is the growth of Russian and Estonian migrants and the arrival of refugee migrants. The graph indicates that Russia (including the Former Soviet Union), Estonia and Sweden, which are all geographically located close to Finland, clearly make up the three largest countries of birth for all years. The immigrants from Sweden include many return labor migrants that moved to Sweden after the second world war or that are of farther Finnish descent. By 2016, about 30 000 of the Russian and Estonian migrants are Ingrian Finns, Lutheran Finns that moved to former Ingria (current area around Saint Petersburg) during the Swedish Empire era, around year 1700. The Ingrian Finns were given the right of return to Finland in 1990 until 2016, which can explain the rise some of the rise of the Russian and Estonian country of birth. The

Somalis also came to Finland during the 1990's and have the longest refugee history in Finland and played an important role in turning net migration positive. Along with the Somalis, Iraqis and Iranians are the largest populations of refugees in 2018. The refugees remain relatively few, as during the refugee crisis in 2015, the Finnish cabinet consisted of populist and conservative parties that were against immigration (Koikkalainen, 2021). Finland granted asylum to 7 745 refugees in year 2016 (Finnish Immigration Service, 2017), which is relatively low in comparison to the neighboring country Sweden, where 67 258 were granted asylum in the same year (Swedish Migration Agency, 2017).

Graph 3 below shows the top ten countries of birth of the mother of the second-generation immigrants in Finland. If the mother's country of birth is not available, the father's country of birth is used. Comparing **Graph 2** and **Graph 3,** it is visible that they differ in composition. The composition of the second generation has not changed over the years as much as the first generation has. The top ten countries for the second generation remained the same until year 2016, when Afghanistan passed Iran. The rise in immigration that has occurred in Finland during the 21st century, including the refugee crisis, is more visible among the first generation as an effect of compositional change. The second generation is a reproduction of the first and the compositional changes will probably be more visible in the near future. The compositions of the two generations also differ in the country ranking. For example, the Somalis make up the second largest country of birth for the second generation but are about the fifth largest in the first generation. Oppositely, the Swedes are ranked at the higher end of the top ten countries of the second generation but as the second and third over the years of the first generation. These ranking differences between the first- and the second-generation populations can be due to that birth rates differ across nations and cultures (Rindfuss & Brewster, 1996), their age of arrival, whether they arrive before or after childbearing age, or their motivation to migrate.

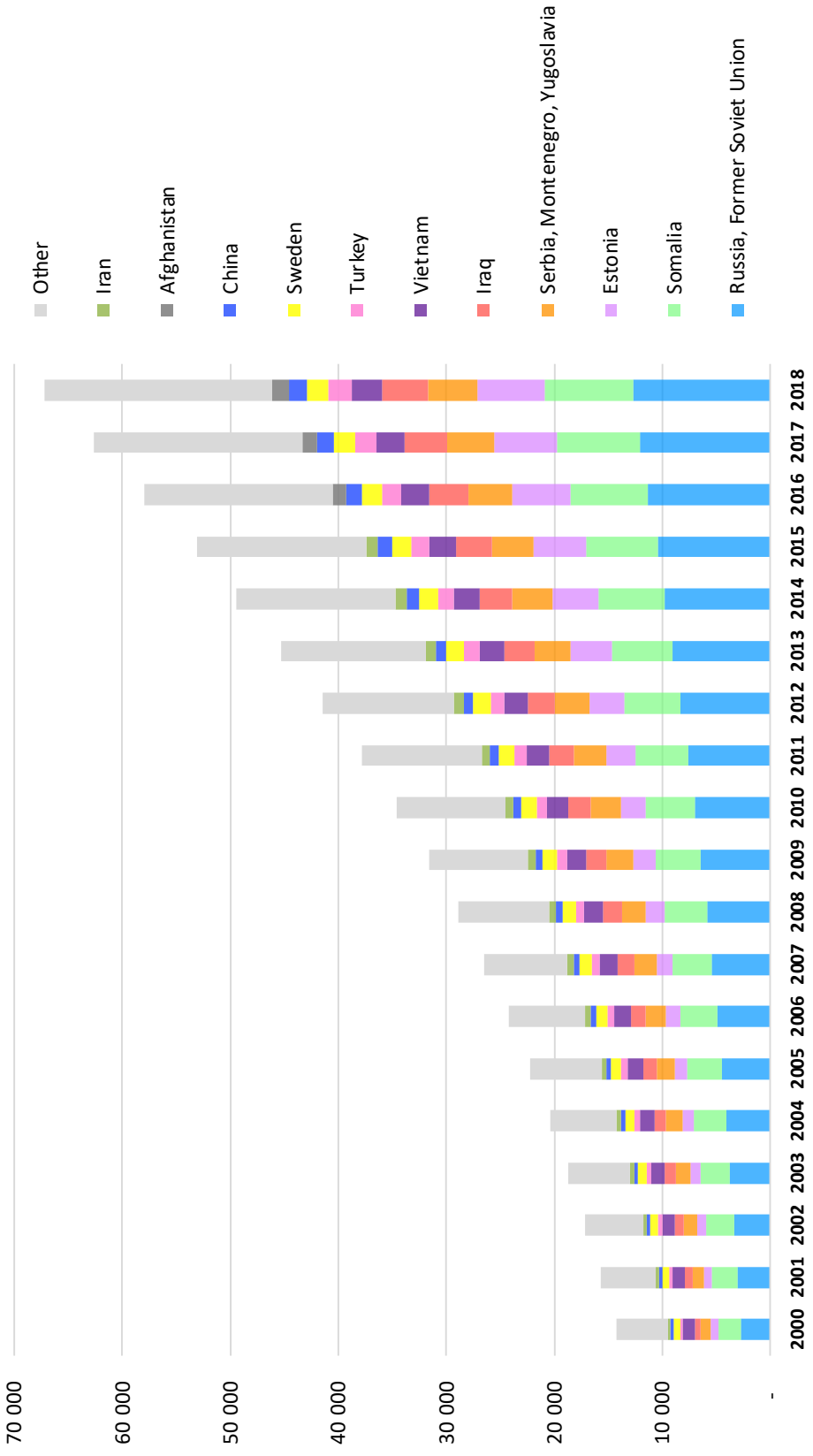
The corresponding graphs showing the top ten countries of birth for the first- and second-generation of the 15-year-olds, the study population of this thesis, are found in the Appendix, **Graph A 1** and **Graph A 2**. The first-generation 15-year-olds have the same top seven countries as the total population but differs slightly in order, the magnitude also fluctuates more over time. The second-generation increases over time and have the same top eight countries as the second-generation total population with slight differences in order.

Graph 2: Top 10 countries of birth of first-generation immigrants in Finland.



Data: Statistics Finland.

Graph 3: Top 3 countries of birth of the mother of the second-generation in Finland.



Data: Statistics Finland

How much immigrants can individually economically succeed and benefit the Finnish economy as a whole, both depends on how well they become integrated into the society, mainly how well they get employed. If immigrants are successfully integrated into the labor market they become valuable tax payers, but if they remain unemployed, they become more of a cost for the society (Väänänen, 2018). Much focus is given to the working age immigrants' integration into the labor market, but the school performance is also important for an individual's long run achievement and to strengthening the economy in Finland. To ensure high employment of people with immigrant background in the long run, their school performance becomes essential as it is strongly correlated with occupational outcomes (Hanushek and Woessmann, 2021). The better immigrants perform in school, the more probably they will attain a higher education and get employed.

How well immigrants succeed in and become integrated into the destination country are also affected by the reason behind migration and the distance between the country of origin and destination. The reason to migrate can be voluntary or forced. The country of origin and destination can be geographically, culturally, socio-economically, demographically or politically distant. To voluntarily move to a neighboring country with a similar language and culture is usually a relatively uncomplicated process and the probability to socioeconomically succeed is high (Bhugra, 2004). In Finland, the first-generation population is clearly dominated by the three neighboring nations, Russia, Estonia and Sweden, seen in **Graph 3** above. When a person immigrates from a more culturally or geographically distant country, the move becomes more challenging. The migration process involves a series of steps to accomplish which can be mentally exhausting, such as leaving the country of origin, traveling to and entering the country of destination, and finally being legally accepted as a resident. If additionally, the migration is forced, the socio-cultural adjustment in the destination country can be very burdensome (Bhugra, 2004).

Finland has a short history of immigration and a difficult language to learn, which can make it relatively difficult for immigrants to become integrated into the society and compete with natives. The Finnish population is adapted to the homogenous population and is only slowly adjusting to the growing multiculturalism, which has led to some hostility and negative attitudes toward people with foreign ethnicities, making it harder for immigrants to socially integrate with the native population (Ervasti, 2004). To feel belongingness as an immigrant in the relatively culturally homogenous Finland can be more difficult compared to, for example, moving the United States, which is a melting pot of cultures and ethnicities. Finnish is one of

the most difficult languages in the world, which requires immigrants to invest more time and energy in learning the language (Selmer & Lauring, 2011). Having proficiency in the native language is specifically important in education, which is mainly taught in Finnish, where as it is easier to succeed in the job market with proficiency in English. The Finnish school system is the most important institution where immigrant students are able to improve their language proficiency and to gain the feeling of social belongingness, which the students can carry along throughout their lives.

Theory and previous research

In majority of the developed countries and especially the non-Anglo-Saxon countries, including Finland, immigrants perform worse than natives in school (Murat & Frederic, 2015; OECD, 2020b). As migration increases, the low school performance of immigrants is a growing concern, which researchers are trying to actively solve. The OECD's data about the results from the PISA test, which was first taken in 2000, have contributed greatly to the research on school performance world-wide, as it includes detailed information about students' background and the school system that can affect their performance. Immigrants perform worse than natives in Germany, but the gap has decreased over the PISA cycles, however, it remains unclear whether policies are effective (Teltemann & Rauch, 2018). Similarly in Switzerland, the immigrants' school performance improved from 2000 to 2009 and Cattanea and Wolter (2015) found that majority of the change could be explained by a radical migration policy change that allowed more qualified migrants into the country. In Spain, Zinovyeva, Felgueroso and Vazquez (2014) found that immigrants that have arrived earlier perform better than those that have arrived later, but no matter how early immigrants have arrived in Spain, they still perform on average worse than the native students.

The literature on immigrants' school performance in Finland is relatively scarce due to the short history of immigration. Since the beginning of the PISA test cycle in 2000 research about the Finnish school system has boomed due to the high performance of the Finnish students. Only in the more recent test cycles, in 2012 when immigrants where over sampled, the alarming gap between natives' and immigrants' school performance has gained more attention in the research field. The PISA test results from 2012 indicated, according to the OECD's classification, that the performance gap between immigrant and native students corresponded to two years of learning in school. Immigrant students' performance was similar in all three fields, mathematics, science and reading, and the first-generation performed slightly worse than the second (Harju-Luukkainen & McElvany, 2018).

Kirjavainen and Pulkkinen (2017) also studied immigrants' PISA results from 2012 in Finland, specifically how the country of origin is associated with the school performance of students with an immigrant background. They found that there are performance differences between immigrant students that originate from different countries. Students from Somalia

they found to perform the worst. Also, students from Turkey, Iraq and former Yugoslavia preformed significantly worse than native students. Students originating from countries closer to Finland, such as Estonia, Sweden and Russia, had the smallest performance gap to the native students. The performance difference between country of origin remained more or less the same for first- and second-generation students and in all three subjects, mathematics, science and reading. The first generation performed worse than the second generation, and in relation to the native students, out of the three PISA subjects, they performed worst in reading, while the second generation performed worst in mathematics in relation to the native students, which indicates the benefit for the second generation of having proficiency in the native language. Kirjavainen and Pulkkinen controlled for background factors, including the student's gender, age, age of arrival in Finland, language spoken at home, language of the test and factors indicating the parents' socioeconomic status. Controlling for the background factors lowered the effect that country of origin had on students' test performance and made the effect of most countries non-significant. The largest lowering effect that emerged after controlling for background factors, was seen in countries that have the furthest distance to Finland. Kirjavainen and Pulkkinen suggest that the performance difference between students with different origins is due the difference between the school system in Finland and the country of origin, as well as, the difference between the mother tongue and the language of instruction. The performance differences that are linked to the country of origin, found in Kirjavainen's and Pulkkinen's study, are in line with the theory that immigrants' success is related to the distance between the country of origin and destination discussed in the background section above (Bhugra, 2004).

Kilpi-Jakonen (2012) studied immigrants' school performance in Finland based on their graduating grades received from compulsory school in years 2000 to 2004. Kilpi-Jakonen examined what effect parental resources have on the school performance of immigrant versus native students. The study is motivated by the theory that a large fraction of the low school performance of immigrants can be attributed to their parents' low socioeconomic status. Kilpi-Jakonen aimed to determine whether the disadvantage that immigrant students face is removed after controlling for parental resources. The parental resources variable consists of parents' income, labor force participation, education, household composition and parents' socioeconomic status. Kilpi-Jakonen found that controlling for parental resources significantly lowers, but does not remove completely the disadvantage effect that being an immigrant has on school performance. Her main finding is that the effect of parents'

education was smaller and the effect of parents' income was larger for immigrant compared to native students. Kilpi-Jakonen explains that the varying effect of parental resources can be due to that those parents with a high education cannot possibly support their children well without economic resources (low income) or that having a low or no income can cause stress in parents that can then weaken their ability or motivation to help their children in their school work. Similar to Kirjavainen's and Pulkkinen's findings (2012), Kilpi-Jakonen (2012) found that the neighboring countries perform best and students that originate from countries from where immigrants are mostly refugees perform worst. Kilpi-Jakonen suggests that the low performance of students with refugee backgrounds can be due to traumatic experiences, which is in line with the theory that the reason to migrate affects the performance in the destination country discussed in the background section above (Bhugra, 2004). Kilpi-Jakonen's (2012) findings differ to some extent from other studies, specifically that the effect of country of origin is relatively small and the performance difference between second and first generation disappear after controlling for parental resources. Kilpi-Jakonen explains that the unusual findings are most probably due to that the Finnish grading policy allows to adjust grades according to language skills and it is found that immigrants' grades are slightly higher than what their true test performance indicates. Using the PISA test results thus avoids the bias that can emerge in school grading by teachers, as the PISA test results are internationally standardized and based on the true test performance of each student.

Support for that parents' socio-economic status and the language spoken at home are significant factors explaining students' school performance, is also found in international comparative studies. Schleicher (2006) compared immigrants' school performance, based on the PISA 2003 results, in 17 OECD countries with large immigrant population, excluding Finland. He found that immigrants' school performance is related their parents' education and socio-economic status, but controlling for the parental factors does not remove the performance gap between immigrant and native students. Schleicher also found that the more dissimilar the language spoken at home is from the language of instruction, the worse the students perform. Entorf and Minoiu (2005) studied the effect of immigrant background and social status on students' school performance based on the PISA 2000 results in nine countries, including Finland. They found that the effect of parents' socioeconomic status on students' school performance differs significantly across the countries studied and is lowest in Finland, which is in line with the strong emphasis on equity in the Finnish school system highlighted by Sahlberg (2021). Entorf and Minoiu (2005) also found that the benefit of

speaking the national language at home is second highest in Finland, among the nine countries studied. The relatively high effect of language proficiency on school performance in Finland follows the logic of that Finnish language is one of the most difficult languages in the world (Selmer & Lauring, 2011). Entorf and Minoiu (2005) suggest that improving immigrant students' proficiency in the native language is the most important tool to close the performance gap to native students.

Heath and Kilpi-Jakonen (2012) studied how the age of arrival affects immigrants' reading performance based on PISA pooled data from cycle 2003, 2006 and 2009 in about 40 countries, including Finland. They compared immigrant students' reading performance between the early, mid and late arrivers, which correspond to the 1,75-, 1,5- and 1,25-generations respectively analyzed in this study, where instead the math score is used as a measure of school perfromance. Health and Kilpi-Jakonen found that immigrant students in the OECD countries face a "late-arrival penalty" (p.10), specifically, that the penalty rises progressively with age, which varies across countries. Their findings show that, in Finland, the penalty faced by students that arrive late compared to those that arrive early is close to the average of the OECD countries, but the penalty for those that arrive early compared to the native students is higher than the OECD average, and fifth highest of the countries studied. Heath and Kilpi-Jakonen assert that female perform in general better than male students but the that late-arrival penalty does not vary across gender. They suggest that the variation in the size of the late-arrival penalty across countries is due to the distance between country of origin and destination, mainly the difference in language and educational standards between the two countries.

The literature on immigrants' school performance also includes studies that, as an addition, studied the characteristics of the school system that are related to immigrants' school performance. Borgna and Contini (2014) compared the second-generation' school performance, based on the 2006 and 2009 PISA test results, across 17 Western countries, including Finland. They aimed to identify the specific penalty that the second generation faces in their school performance and factors of the school system that affect their school performance. Borgna and Contini found that the size of the migrant penalty is negatively correlated with the size of the socioeconomic penalty that students face in their school performance. Consistent with findings discussed above, Borgna and Contini identified that the socioeconomic penalty in Finland is low, but the second generation students in Finland suffer the highest migrant-specific penalty compared to the other countries included in the study.

Borgna and Contini suggest that factors of the school system also affect immigrants' school performance as the migrant penalty is not removed after controlling for individual background factors. The school system factor that they studied is the age when students enter the school system and find that early entry is related to lower migrant penalty. They highlight that Finland has the latest entry into the school system in addition to the highest migrant penalty. Finland made pre-primary school mandatory in 2015, which hopefully will lower the migrant penalty associated with late entry, but the effects will not be discoverable in the data used in this thesis yet.

Behr and Fugger (2020) tested how efficiently the school system in 20 industrialized countries integrate immigrants, by using the results from the PISA 2015 cycle. They use the Data Envelopment Analysis model to generate efficiency scores for native and immigrant students separately, that indicate how well a student performs in relation to the highest performing student within the same socio-economic status. They found that the Finnish school system was among the least efficient, and in some tests the least efficient, in integrating immigrants. Behr and Fugger highlight the high quality of the Finnish school system in contrast with the low school performance of immigrants in Finland. They also stress that the wide performance gap between native and immigrant students in Finland has gained surprisingly little attention in the research field.

The high migrant penalty that immigrants face in the Finnish school system compared to other nations is alarming. Previous studies have proved that when accounting for background factors, such as the socioeconomic status of parents or language spoken at home, the effect of having an immigrant background on school performance is lowered and, in some countries, it is even removed completely, but this is not the case in Finland (Borgna & Contini, 2014). There is a shortage of literature on the migrant penalty that immigrants face in their school performance in Finland, specifically studies that analyze the immigrant generation in more detail based on the age of arrival and track the trend over a longer period.

Given the previous literature, this thesis aims to provide a more comprehensive understanding of the relationship between immigrant generation and school performance over time. Specifically, this thesis contributes to the literature in three ways. First, the relationship between the PISA math score and immigrant generation is analyzed over time by studying the trend from 2003 to 2018. Second, this thesis studies whether the relationship can be explained by student compositional and background factors. Third, students with an immigrant

background will be divided into five generations, by using Rumbaut's (2014) age of arrival classification to generate the first generations and whether the student has a native parent to generate the second generations.

Based on the theories and previous research presented above, four hypotheses will be tested in this study: Hypothesis 1 (H1): Despite that the Finnish school system is ranked among the best of the OECD countries and has a strong emphasis on equity, we expect that students with an immigrant background perform worse than natives on the PISA math test.

Hypothesis 2 (H2): Students with an immigrant background will perform in relation the native students worse over time.

Hypothesis 3 (H3): Once the model controls for the compositional factors, the performance difference between the immigrant-generation and the native-generation students will decrease, but some difference will still persist.

Hypothesis 4 (H4): Students with an immigrant background that are raised or born in Finland will perform better than those that have arrived at older ages.

Data and variables

Data: PISA

To analyze the relationship between immigrant generation and the math score in the Finnish school system, the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA) data is used. The data is available at and downloaded straight from PISA's website. The PISA test is an assessment test that analyzes the knowledge of fifteen-year-old students, that the OECD considers to be important for succeeding in the contemporary society. The test is taken on a three-year-interval and the first test was taken in 2000. It assesses three subjects, mathematics, reading and science, and is standardized across countries but given, in most cases, in the native language. Additionally, the PISA data set includes information about the student's background, acquired through questionaries assigned to the school councils, teachers, and parents. The questionaries comprise information about the school, teaching and personal informational about the student, such as age, gender, school motivation, and information about the student's family, such as cultural, social and economic status. The comprehensive results from the PISA data are utilized by policy makers world-wide to generate policies to improve their school system (OECD, 2019a). This thesis analyzes the math performance in six test cycles, 2003 to 2018.

The sample used in this study includes randomly selected students from Finland which serve as the observations in the regressions. PISA uses a two-step sampling process to gather a student sample that is representative of the true population, by first selecting the schools and then the students. First, the schools are selected based on the probability proportionate to size sampling method (PPS), which means that the larger the school is the higher probability it has to be selected. The PPS method is used for school choice, rather than random sampling, as the test is specifically interested in information about the student population and not schools. Students are then randomly selected within each school, typically 35 students per school. Weights are included in the data in order for the student sample to represent the whole population (OECD, 2009b), and used in the descriptive analysis of the variables and the regressions in this study, which are discussed in more detail in the Method section below. The identity of the student is protected, as the PISA results are analyzed anonymously. As a student is chosen to participate in the test, they receive an identification code to which their

data is then connected and no names are used throughout the test process (Ministry of Education, Children and Youth of the Grand Duchy of Luxembourg, 2018).

Variables

Dependent variable: PISA math score

The dependent variable used in this study is the PISA math test score. The math score is used, instead of the reading or science score, to control for the positive effect that the proficiency in the native language can have on school performance (Entorf and Minoiu, 2005; Schelicher, 2006), which can be more prominent in the reading and science tests. The math score is calculated by PISA based on the performance on the math test, which includes a wide variety of questions to get a comprehensive result on the student's mathematical literacy. On the national level, the average math scores vary from around 350 to 550 points across countries (OECD, 2022).

In the PISA data, each observation's (student) test score is reported along with multiple possible scores, called plausible values. Until and including cycle 2012, five plausible values are given, whereafter, for cycle 2015 and 2018, ten plausible values are given. The plausible values are generated in order for the performance of the student to better represent the true population. Each student's plausible values are taken at random from the posterior distribution around the student's reported score. The plausible values represent the additional possible test scores that the student is realistically adept to achieve, both higher and lower than the reported score (OECD, 2009a). Two factors determine the posterior distribution from where the plausible values are drawn: 1) the student's test responses, scaled according to the item response theory and 2) a set of multiple variables measuring different student background factors, based on the latent regression model (OECD, 2017).

Independent variable: Immigrant generation

The main independent variable used in this study is immigrant generation. The immigrant generation variable includes the 1,25-, 1,5-, 1,75-, 2-, 2,5- and native-generation. The three first generations include students that are born abroad and whose one parent or both parents are born abroad, but differ on the student's age of arrival in Finland and are categorized based on Rumbaut's (2014) age of arrival concept. The 1,75-generation include immigrant students that have arrived before age 6. The 1,5-generation include immigrant students that have

arrived after turning 6 but before turning 13. The 1,25-generation include immigrant students that have arrived after turning 13. The two second generations include students that are born in Finland but differ on whether their parents are born in Finland. The 2-generation include all those students that are born in Finland but both of their parents are born abroad. The 2,5-generation include all those students that are born in Finland and one of their parents is born in Finland and the other one is born abroad. The native generation includes all students whose both parents are born in Finland, no matter where the student is born. If both of the student's parents are born in Finland, even though the student is not born in Finland, we treat them as a native. We expect a student with two native parents to have similar prospects to perform in school as natives, as they supposedly have adapted to the Finnish culture through their native parents. The description of the immigrant generation variable is summarized in **Table 1**. Some observations have missing values on the country of birth of the student or parents, which means that the immigration variable is not able to be generated for those observations and they had to be dropped from this study. The dropped observations are displayed across the cycles in **Table A1** in the Appendix.

Table 1: Description of the independent variable immigrant generation.

Variable subgroup	Condition
Native-generation	Both parents are born in Finland.
2,5-generation	Student is born in Finland and one of the parents is born in Finland and the other abroad.
2-generation	Student is born in Finland and both parents are born abroad
1,75-generation	Student is born abroad and has arrived in Finland before turning 6.
1,5-generation	Student is born abroad and has arrived in Finland after turning 6 and before turning 13.
1,25-generation	Student is born abroad and has arrived in Finland after turning 13.

The immigrant generation variable used in this study is more comprehensive than the one provided by PISA, which is not applied in this study but commonly used in previous studies. A cross tabulation of how the observations of the 2012 cycle are distributed according to the immigrant generation variable used in this study and PISA's variable is displayed in **Table A2** in the Appendix. The PISA immigrant generation variable categorizes a student with only one parent born in the destination country as a native student, whereas the immigrant

generation variable used in this study requires both parents to be born in the destination country in order to be categorized as a native. As a result, the immigrant generation variable used in this study categorizes a larger fraction of the sample as students with an immigrant background. More importantly, the PISA immigrant generation variable only includes categories for the first, second and native generation. It is found that the behavior of people with an immigrant background is associated with their age of arrival and whether they have a native parent, both which the immigrant generation variable used in this study takes into account. For example, Levels, Dronkers and Kraaykamp (2008) studied the 2003 PISA math performance of immigrant students in 13 Western countries and found that the 2,5-generation performed better than the 2-generation immigrant students. Mussino (2021) studied the childbearing behavior across three immigrant generations (1-, 1,5- and 2-generation) and found that the probability of having a second child and the effect that parental leave uptake has on that probability in Sweden is related to the age of arrival of the first-generation. It is common that studies, out of the first-generation, only analyze the 1- and 1,5-generations, immigrants that have arrived after versus before age 18 respectively. Rumbaut (2004), however, advices to distinguish between the 1-, 1,25-, 1,5- and 1,75-generations in scientific studies. (The 1-generation are adult immigrants who have arrived after age 17 and are outside the age range of the population studied in this thesis and hence not included). Rumbaut argues that age of arrival is significantly related to immigrant's performance, as "there are fundamental differences in the pace and mode of adaptation" (2004, p.1166). But instead of studying how the age of arrival as a discrete variable is related to immigrant children's behavior, Rumbaut argues that individual age groups are related to behavior, specifically, life stages in relation to primary education. He classifies the first-generation children in three generations based on what life stage they are in when they arrive in the country – early childhood (1,75-generation), middle childhood (1,5-generation) and teens (1,25-generation). The 1,75-generation (age 0-5) are children who arrive before pre-school and have normally not yet began learning to read or write in their country of origin and hence resemble the native children. The 1,5-generation (age 6-12), have potentially started learning in their in their mother tongue but finishes school in the country of destination. The 1,25-generation (age 13-17) have attended most grades of primary-school in their country of origin and performs more closely to the immigrants that have arrived as adults, the 1-generation. Rumabut also supports the division of the second generation into those who have one versus two native parents (2- vs 2.5-generation), as those with two native parents perform more closely to the native population.

Table 2 below shows the unweighted number of observations across generations and cycles. The number of immigrant-generation observation increases across the cycles, along with the true immigrant population in Finland. This study does not show results where the sample includes 10 or less observations, which is the case for the 2-generation in cycle 2003 and 2006, and for the 1,25-generation in cycle 2006 and 2009. Despite that the immigrant sample size is growing over time, the samples in the early years are still big enough to include in the analysis of this study. The data on the test taken in 2000 does not include the immigrant student's age of arrival for Finland, which is why year 2000 has to be excluded from the analysis as the immigrant generation groups cannot be generated.

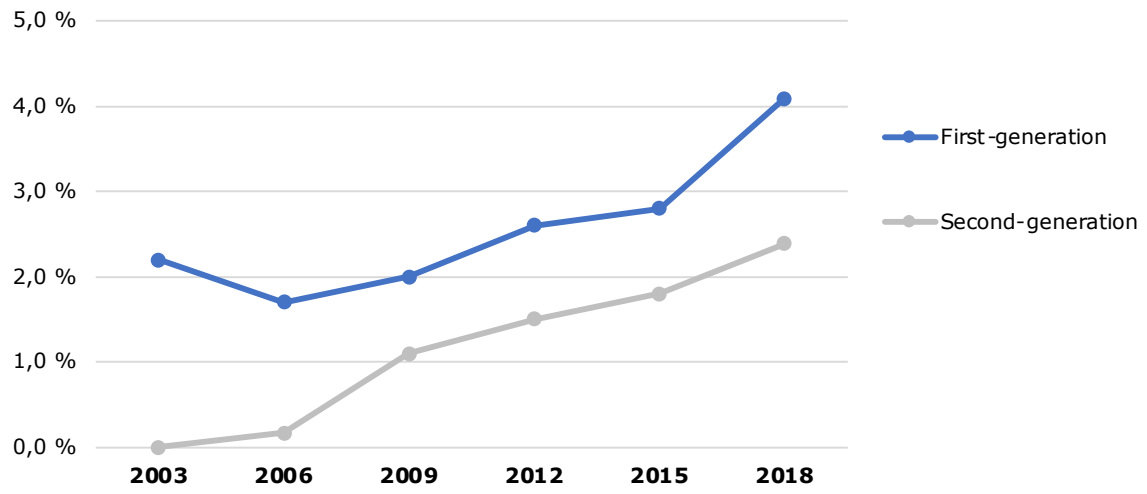
Table 2: Sample size of the independent variable immigrant generation after controls.

Variable subgroup	2003 (#)	2003 (%)	2006 (#)	2006 (%)	2009 (#)	2009 (%)
Native-generation	5 343	95,22 %	4 389	95,14 %	5 270	92,60 %
2,5-generation	133	2,37 %	137	2,97 %	256	4,50 %
2-generation	3	0,05 %	9	0,20 %	58	1,02 %
1,75-generation	60	1,07 %	41	0,89 %	56	0,98 %
1,5-generation	55	0,98 %	31	0,67 %	41	0,72 %
1,25-generation	17	0,30 %	6	0,13 %	10	0,18 %
Total	**5611**	**100 %**	**4613**	**100 %**	**5691**	**100 %**

Variable subgroup	2012 (#)	2012 (%)	2015 (#)	2015 (%)	2018 (#)	2018 (%)
Native-generation	6 353	74,71 %	5 128	89,43 %	4 797	87,79 %
2,5-generation	777	9,14 %	345	6,02 %	318	5,82 %
2-generation	563	6,62 %	102	1,78 %	129	2,36 %
1,75-generation	364	4,28 %	61	1,06 %	97	1,78 %
1,5-generation	359	4,22 %	76	1,33 %	97	1,78 %
1,25-generation	87	1,02 %	22	0,38 %	26	0,48 %
Total	**8503**	**100 %**	**5734**	**100 %**	**5464**	**100 %**

The weighted size of the first and second-generations immigrants in relation to the whole sample is visible in **Graph 4** below. The second generation include the 2-generation students only and the first-generation includes 1,25-, 1,5- and 1,75-generation students. The 2,5-generation is not displayed in the **Graph 4** in order for it to be comparable to **Graph 1**, which displays the true population. The weighted immigrant generation study sample represents the true population well. The second-generation sample is close to identical to the true population in magnitude and trend over time. The first-generation sample has a similar trend but slightly lower in magnitude than the true population.

Graph 4: First- and second-generation immigrants as a share of the total sample.



Control variables

Some observations have missing values on the control variables, specifically the ESCS and Language variable and are dropped from the regressions. The dropped observations are displayed across the cycles and generations in **Table A2** in the Appendix.

ESCS

The first control variable used in this study is the PISA index of economic, social and cultural status, ESCS. The socio-economic background is found to be a significant factor affecting school performance, specifically that higher socio-economic status is related to higher performance (Schleicher, 2014). The ESCS variable is generated and released by PISA by combining three family background PISA variables: the highest occupational status of the parents, the highest education level of the parents and household possessions. The highest

occupational status of the parents-variable is generated by taking the higher International Socio-Economic Index of occupational status of the two parents, which is based on the International Standard Classification of Occupations codes (ISCO) ascribed based the answers found in the PISA questionary. The highest education level of the parents-variable is generated by taking the higher education level of the two parents, which is classified based on the seven level International Standard Classification of Education (ISCED) from 1997. The household possessions variable is generated based on three questions from the PISA questionary. The first question asks for whether the student has specific things at home, such as a desk, own room and computer. The second question asks about the quantity of specific goods, such as number of TVs, cars and electronic devices. The third question asks for the number of books at home. The scale of the ESCS variable is based on the OECD average, which is given 0 and the values vary from around minus to plus 4 in the sample of the students from Finland (OECD, 2019b).

The ESCS variable accounts for economic, social and cultural status in one variable and is considered to comprehensively account for the socio-economic of the student (OECD, 2020a). The ESCS variable is used in studies where the interest is in the effect of the socio-economic status and less so in knowing the individual effect of economic, social or cultural status. See for example, Borgna and Contini (2014) who study the school performance of second-generation immigrants in Western Europe and Behr and Fuegger (2020) who study the effect of immigration policies and school systems on immigrants' school performance, both studies using the ESCS variable. The average weighted ESCS value for each cycle and generation of this study's sample is displayed in **Table A 4** in the Appendix. The general trend of the average ESCS values is that they decease when going from the native-generation toward the 1,25-generation, excluding the 2,5-generation's values that are close to the native-generation's values. Over the years the ESCS take a wider variety of values, excluding 2012 when immigrants are overrepresented. The growing diversity of the ESCS values can be due to the rise of the immigrant population in Finland, specifically that the composition has become more diverse, including cultural, social and economic background. The ESCS value also fluctuates over the years within the same generation, which could be explained by that the compositions differ across the years.

Language

The second control variable used in this study is the language spoken at home. The language variable is taken straight from the PISA data and distinguishes between whether the student speaks the language of the test at home or not. The language variable is constructed as a dummy variable, where the language of the test is indicated with number 1 and foreign language with number 2. Proficiency in the language of instruction is beneficial for school performance. The Finnish language is one of the world's most difficult languages to learn and hence has the proficiency in the Finnish language been found to have a relatively large effect on immigrants' school performance (Entorf and Minoiu, 2005), which has been taken into account in this study by analyzing the PISA math score, instead of reading or science, where language proficiency is potentially more important. **Graph A 3** in the Appendix displays the weighted share of the sample that does not speak the language of the test at home. The share that does not speak the language of the test at home increases over the cycle in line with the general immigration trend in Finland.

Gender

The third control variable in this study is gender. The gender variable is taken straight from the PISA data, where it is constructed as a dummy variable. Females are represented by number 1 and males by number 2. The sample used in this study is close to evenly represented by boys and girls and the weighted gender distribution is displayed in **Table A 5** in the Appendix. The largest difference occurs in year 2015, in which out of the 5 734 observations there are 124 more boys. The general finding about how gender is related to school performance is that girls outperform boys (Heath and Kilpi-Jakonen, 2012). Girls' higher performance is strongly present in reading performance (Logan and Johnston, 2010). In math performance the situation is found to be the opposite, as boys tend to perform better than girls (Zhu, 2007).

Method

Method

This thesis descriptively analyzes the relationship between immigrant generation and PISA math score by performing two ordinary least squares (OLS) linear regressions in the statistical software Stata. The two OLS regressions will be performed separately for each cycle from 2003 to 2018, instead of pooling the results together and controlling for the year, as a new test is produced for each PISA cycle and the scores are hence not perfectly comparable. For example, a 500-point math score in 2015 might not indicate the exact same level of knowledge as a 500-point score in 2018. The two OLS regressions also include the same variables for all cycles. The OECD (2019c) suggests that if performance is compared over time, the variables included in the analysis must be present and measured the same way in all cycles. The data and variables used in the regressions are presented in detail in the previous section and the regressions are performed in Stata by using the OECD's repest command. The PISA test aims to generate information about the knowledge level of a country's whole population based on the test sample's performance. The repest command allows Stata to utilize all plausible values of the math score and weights in the regression. The plausible values and weights account for sampling variance and measurement error to produce results that represent the entire population. Using plausible values instead of only the one reported math score, allows each observation to represent a broader and more realistic performance ability. The weights include one final student weight and 80 replicate weights per observation. The final student weight is an indication of how many students the specific observation represents in the true population, as the probability to become selected varies across students, due to several reasons associated with the sampling process. The sum of all final student weights equals the size of the true student population. The 80 replicate weights per observation are used to control for sampling variance. PISA uses the Balanced Repeated Replication (BRR) method, specifically Fay's modification, to compute the sampling variance (OECD, 2009a; OECD, 2009b; OECD, 2009c; OECD, 2017).

The first OLS regression performed in Stata for each cycle, is the univariate relationship between the PISA math score and immigrant generation and aims to answer RQ1. The dependent variable, the PISA math score, will be regressed against the main independent variable, the immigrant generation, where the native generation will remain as the reference

category. This regression allows us to descriptively analyze how the different immigrant generations are related to the math performance in the Finnish school system over time. Specifically, whether the relationship differ across the generations or across the cycles.

The second OLS regression performed in Stata for each cycle, is the multivariate linear regression, which produces a more detailed understanding of the relationship between the PISA math score and immigrant generation in the Finnish school system and aims to answer RQ2. The control variables will be added to the univariate regression to descriptively analyze if the relationship between the math score and immigrant generation can be explained by the compositional and background factors. Specifically, how much controlling for the economic, social and cultural status, language spoken at home and gender, changes the potential association between having immigrant background and math performance, found in the first regression.

The control variables used in the final model are selected based on the theory and previous findings discussed in the theory section above and availability across the cycles. The PISA test has developed over the cycles to become more comprehensive by including more variables and more detailed questionaries. However, in order for this study to analyze the trend over time, identical regressions have to be performed for all cycles, which means that the variable choice is limited to the variables available in the earliest cycle analyzed, 2003. For example, the country of birth for the student or their parents is not given in the first cycles, which is why such analysis is not included in this study. Neither the specific language spoken at home could be used to control for the composition, as the variable was not consistent over the cycles and majority of the foreign language speaking observations were classified as speaking “other” language.

The relationship between immigrant generation and math performance is studied over time to examine the development of the relationship. All available PISA cycles are included in the study, except the first cycle, 2000, which had to be excluded from this study as age of arrival was not given in the data, which is required to produce the immigrant generation variable used in this study. Previous studies have observed that students with an immigrant background lag behind native students based on their school performance for a specific year. This thesis studies the trend over time to get a more comprehensive understanding of the relationship. Studying the relationship over time we can see whether it is present in all cycles and if it changes or potentially improves, in practice, whether the performance gap is

widening or narrowing. As we have split the students into generations, we can potentially observe whether the performance of a specific generation improves or worsens more than another.

Ethical considerations

The motivation for writing this thesis has arisen from previous research that has found that the school performance gap between native and immigrant students in Finland is large compared to other OECD countries. By using a more comprehensive definition of immigrant generation and studying their performance over time in relation to the native generation, this thesis contributes to previous knowledge about the school performance of students with an immigrant background. The findings of this thesis are meant to help students with an immigrant background to catch up with the native students. Specifically, by providing researchers and policy makers with more insight into how age of arrival, for immigrant students born abroad, and having one native parent, for immigrant students born in Finland, is related to their school performance, and whether this relationship changes over time. We believe that closing the school performance gap between the native students and students with an immigrant background, will have long lasting effects on the performance gap in later life outcomes, mainly in the labor market. We are aware of the ongoing political debates about immigration and refugee admissions in Finland and the ethical issues that might follow. This thesis does not take a political stance toward the topic.

This thesis adheres the Swedish Research Council's (2017) good research practice and all steps taken to conduct this research is shared in detail and honestly throughout the thesis. This thesis studies students' school performance based on the OECD's PISA data, where students are randomly picked and remain completely anonymous. The OECD protects the identity of the student by providing them a personal code used throughout the whole testing process and in the released data. The PISA data is accessible and downloaded straight from the OECD's website. The regressions in this study are performed by using the PISA data only, where all variables are generated and released by PISA, except the immigrant generation variable, which we generated by only using variables included in the PISA data set. No results are shown if the number of observations is equal to or less than 10 to avoid underrepresented results. This thesis does not study the specific foreign country of birth nor ethnicity. In the

regressions, this thesis only categorizes the students based on whether they or their parents are born in Finland and the student's age of arrival.

Results

How does the relationship between the PISA math score and immigrant generation change across the generations and over time? (RQ1)

The univariate relationship between the PISA math score and immigrant generation in the Finnish school system is displayed in **Table 3** below, which are the results from the first OLS regression of this study performed for all cycles from 2003 to 2018. The stars indicate the level of significance of the results and the blanks indicate that the sample for those specific generations and cycles include 10 or less observations and not shown in this study. The table also includes the number of observations and the r-squared. The number of observations is the number of students that are analyzed in the regression, which is the original sample excluding observations with missing values on variables included in the two regressions. The first and the second regression include the same observations in order for them to be comparable. The r-squared is relatively low in for all cycles in the univariate regression, but higher in the three last cycles than the thee first cycles. The constant represents the average math score of the reference population, the native-generation students. The coefficients displayed for each generation is the relationship between the immigrant generation and PISA math score in relation to the native generation, i.e., the average math score gap between the immigrant-generation students and the native-generation students. The negative immigrant-generation coefficients indicate that all the immigrant-generation students performed on average worse than the native-generation students in the PISA math test in Finland in all cycles from 2003 to 2018. Majority of the coefficients are highly significant, at the 1 % level, indicated with three stars. The least significant coefficients are found for the 2,5-generation, for which the 2003, 2006 and 2015 coefficient was found to be non-significant, but the rest of the 2,5-generation's coefficient were significant and highly significant.

The results indicate that there is a positive performance trend when moving from the low toward the high immigrant generations, excluding the 2-generation. The 1,25-generation students, those who are born abroad and have arrived in Finland after turning 13, performed on average the worst on the math test out of the immigrant-generation students compared to the native-generation students, excluding cycle 2012 and 2015 when they performed better

than the 1,5-generation students. The 1,25-generation students scored on average between 64 to 114 points worse than the native-generation across the test cycles. The 1,5-generation students, those who are born abroad and have arrived in Finland after turning 6 and before turning 13, performed on average worse than the 1,75-generation students and on average between 54 and 95 points worse than the native-generation students across the cycles. The 1,75-generation students, those who are born abroad and have arrived in Finland before age 6, performed on average better than the 2-generation but worse than the 2,5-generation students in all cycles and scored on average between 35 and 55 points worse than the native-generation students. The 2-generation students, those who are born in Finland and both parents are born aboard, performed on average worse than the 2,5- and the 1,75-generation students, but still better than the 1,5-generation students, in all cycles. The 2-generation students scored on average 44 to 72 points worse than the native-generation students. Out of the immigrant-generations, the 2,5-generation students, those who are born in Finland with one native parent, performed on average the best among the immigrant generations in all cycles, by only scoring on average between 12 and 19 points worse than the native-generation students.

The relationship between the immigrant generation and the PISA math score fluctuates over the cycles but no clear increasing or decreasing trend is visible in the univariate regression results. The immigrant-generation coefficients fluctuate across the cycles differently between the generations and it is not possible to notice any clear trend within any of the generations over the cycles if all cycles are included. For the 1,75- and 1,5-generations the score gap is decreasing if only the last three cycles are considered or if comparing the first cycle with the last. The 2,5-generation's coefficients are constantly fluctuating and when comparing coefficient of the first significant and last cycle the change is minor. The 2- and 1,25-generations do not have large enough samples to analyze for all cycles, but their coefficients not show any specific trend in the given cycles in addition to the fluctuation. The native-generation's performance, indicated by the constant, increased from the first to the second cycle, but decreased constantly thereafter.

Table 3: Regression results of the univariate relationship between PISA math score and immigrant generation.

Regression 1	2003	2006	2009	2012	2015	2018
Native-generation	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
2,5-generation	**-11.75** (7.32)	**-10.72** (7.71)	**-19.35***** (6.96)	**-11.55**** (5.39)	**-3.17** (5.94)	**-15.49***** (5.33)
2-generation			**-43.60***** (12.90)	**-72.02***** (5.33)	**-48.44***** (10.61)	**-56.11***** (10.06)
1,75-generation	**-43.91***** (14.92)	**-53.37***** (15.18)	**-41.02**** (17.58)	**-54.61***** (6.92)	**-36.93**** (17.73)	**-34.60***** (10.28)
1,5-generation	**-78.28***** (16.33)	**-75.20***** (26.77)	**-54.13***** (11.34)	**-95.26***** (11.97)	**-87.56***** (14.26)	**-62.12***** (10.82)
1,25-generation	**-88.52***** (20.69)			**-73.38**** (29.21)	**-63.94**** (28.68)	**-113.68***** (18.47)
Constant	**547.19***** (1.93)	**551.32***** (2.19)	**543.50***** (2.14)	**524.56***** (1.86)	**515.03***** (2.23)	**513.58***** (1.82)
Observations	5 611	4 613	5 691	8 503	5 734	5 464
R-squared	0.016	0.015	0.017	0.032	0.025	0.034

*Notes: *p<0.10, **p<0.05 and ***p<0.01. Standard errors displayed inside the parenthesis. Blank results indicate that the sample includes 10 or less observations. Weights are used.*

Is the relationship between the PISA math score and immigrant generation explained by the different compositional factors? (RQ2)

The second OLS regression in this study is the relationship between the PISA math score and the immigrant generation along with controls that account for the compositional and background factors of the student population. The control variables are: the economic, social and cultural status, gender and the language spoken at home. **Table 4** below displays the results of the second regression, which is performed for all cycles from 2003 to 2018. The table includes coefficient results for the immigrant generation and control variables along with the constant, number of observations and r-squared. The number of observations is the same as in the first regression in order for the two regressions to be comparable. The r-squared rises substantially after adding the controls to the regression and fluctuates somewhat across the cycles. The r-square varies between 0,084 and 0,137, which means that 8,4 to 13,7 percent of the variation in the math performance of the whole sample is explained by the variables in the second regression. The constant is about 10 points lower than and changes similarly to the constant in the first regression, as it increases from the first to the second cycle and decreases thereafter. The reference category in this regression is an average female native-generation student who has a zero ESCS score and speaks the language of the test at home.

By adding in the control variables to the regression the immigrant generation's coefficients fall, but are not removed. The addition of the control variables has the smallest effect on the 2,5-generation's coefficients, which only change by less than 2 points. For the rest of the generations, for all cycles, the coefficients decrease rather randomly by 6 to 40 points. Some of the coefficients also become non-significant, as the p value rises above 0,1, which is the case for the 2,5- and 1,5-generations in 2015 and the 1,75-generation in 2009, 2015 and 2018. The rest of the immigrant generations' coefficients remain significant on the 10 % level or higher and the 2,5-generation's coefficient in 2003 becomes significant.

Despite the general fall in the association between immigrant generation and math score, the positive performance trend remains, when analyzing only the significant results. The performance difference between the immigrant- and native-generation decreases as we move from the lower toward the higher immigrant-generations. The 1,25-generation students still performed on average the worst out of the immigrant-generation across the cycles, by scoring

on average between 38 and 77 points worse than the native-generation. The 1,25-generation students scored better than the 1,5-generation student in 2012 and 2015, which is in line with the findings of the first regression. A change to the first regression is that according to the second regression, the 1,25-generation students performed on average also better than the 2-generation students in 2012, but worse than 1,75-generation students. The 1,5-generation students performed on average worse than the 1,75-generation, by scoring between 31 and 63 points worse than the native-generation students. The 1,75-generation students' performance is difficult to relate to the 2-generation students' performance, based on the results from the second generation, as the coefficients for both generations are only given and significant for cycle 2012. On average across the available cycles, the 1,75-generation students scored between 22 to 34 points worse than the native-generation students, while the 2-generation students scored on average 18 to 46 points worse than the native-generation students. The 2,5-generation students still performed on average better than the rest of the immigrant generation students, but worse than the native-generation students, by scoring on average 10 to 18 points worse than the native-generation students.

A specific changing trend over time that would be consistent for all immigrant generations was not noticeable in the first regression and is neither so, after adding controls, in the second regression. The coefficients for each generation fluctuate over the cycles, without a clear common pattern. If we compare the first and the last cycle only (2003 and 2018), a decrease is noticeable, as the coefficients fall for the 2,5- and 1,5-generations by 2 and 14 points respectively, but rises by 13 points for the 1,25-generation. The 2-generation is not shown in the 2003 cycle due to the small sample size. If only the last three cycles are examined, the coefficient falls for the 2- and 1,5-generations, but rises for the 2,5- and 1,25-generations.

The economic, social and cultural status of the student, has a positive and highly significant association to the PISA math score in all cycles. The value of the ESCS coefficient indicates how many points the math score changes on average as the economic, social and cultural status of a student increases by one unit, holding other variables constant. The ESCS variable fluctuates slightly over the cycles. The ESCS coefficient is smallest in the 2009 cycle, when one unit increase in the economic, social and cultural status of the student corresponds to a 27,5-point increase in the math score. The ESCS coefficient is largest in the 2015 cycle, when one unit increase in the economic, social and cultural status of the student corresponds to a 35,9-point increase in the math score.

The coefficient of the Other language variable represents the association between speaking another language than the test and math score, i.e., the average score difference between those students that speak the language of the test and those that speak some other language at home, keeping other variables constant. The Other language coefficient is negative and significant in all cycles, which means that if a student does not speak the language of the test at home, they score on average lower than a student who speaks the language of the test at home. The negative effect of not speaking the language of the test at home is largest in 2003 and smallest in 2012, -32,3 and -14,9 points respectively. The association between speaking another language at home and math score has decreased by half from the first to the last cycled studied, indicating that the association might be weakening.

The Male variable represents the association between the student's gender and the PISA math score. The Male coefficient represents the average math score difference between boys and girls, holding other variables constant. The coefficient changes in magnitude and significance level over the cycles. In cycle 2003 and 2006 the gender coefficient is significant and positive. I.e., boys performed on average better than girls on the math test by 9,2 and 12,5 points respectively. The situation is reversed in cycle 2015, as girls performed on average better than boys, by 4,7 points. For the remaining cycles, the gender coefficient is non-significant

Table 4: Regression 2 results: The relationship between PISA math score and immigrant generation with controls.

Regression 2	2003	2006	2009	2012	2015	2018
Native-generation	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
2,5-generation	**-16.19**** (7.09)	**-9.80** (7.60)	**-18.17***** (6.52)	**-9.89*** (5.10)	**-2.48** (5.68)	**-13.81***** (5.27)
2-generation			**-30.41**** (14.09)	**-45.87***** (5.65)	**-17.51*** (10.02)	**-30.36**** (11.95)
1,75-generation	**-21.84*** (13.16)	**-28.57*** (14.69)	**-21.44** (15.89)	**-34.47***** (7.39)	**-30.38** (18.70)	**-17.19** (11.63)
1,5-generation	**-45.09**** (20.44)	**-40.70** (26.00)	**-33.00***** (12.49)	**-63.36***** (11.37)	**-54.94***** (14.99)	**-30.56***** (11.83)
1,25-generation	**-63.93***** (23.85)			**-37.91*** (19.62)	**-51.51*** (29.68)	**-76.55***** (21.90)
ESCS	**33.15***** (1.61)	**31.24***** (1.52)	**27.50***** (1.83)	**31.38***** (1.76)	**35.91***** (2.03)	**33.20***** (1.86)
Language of test	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Other language	**-32.33*** (16.65)	**-24.36**** (11.62)	**-17.03**** (7.10)	**-14.85***** (4.66)	**-17.21**** (7.39)	**-16.37**** (7.69)
Female	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Male	**9.18***** (2.55)	**12.52***** (2.52)	**3.73** (2.49)	**0.46** (2.57)	**-4.73**** (2.23)	**-4.04** (2.59)
Constant	**534.21***** (2.03)	**537.03***** (2.27)	**531.40***** (2.49)	**512.32***** (2.11)	**508.11***** (2.32)	**505.03***** (1.96)
Observations	5 611	4 613	5 691	8 503	5 734	5 464
R-squared	0.127	0.117	0.084	0.114	0.134	0.137

*Notes: *p<0.10, **p<0.05 and ***p<0.01. Standard errors displayed inside the parenthesis. Blank results indicate that the sample includes 10 or less observations. Weights are used.*

Discussion

In line with previous research, this thesis hypothesized that students with an immigrant background perform worse in school than native students in Finland (H1). Based on the results, this thesis indeed finds that the immigrant-generation students perform worse in math than the native-generation students. The performance gap between the immigrant and native generation students does not come as a surprise and is in line with previous research. For example, Harju-Luukkainen and McElvany (2018) studied the immigrant school performance in Finland and found that they performed worse than natives on the PISA test in 2012 in all three subjects. In addition, Finland has gained attention in cross-country comparative studies, as students in Finland face a relatively large migrant performance penalty compared to other countries studied (Borgna & Contini, 2014). Previous research has found that immigrants lag behind the native population due to various integration barriers. In the school system, two major factors that are commonly linked to the performance gap between the immigrant and native generation students are: the native language proficiency of the student (Entorf & Minoiu, 2005) and the socio-economic status of the parents (Kilpi-Jakonen, 2018). The higher proficiency the student has in the native or the instruction language, the easier it is for them to understand and absorb the teachings. Having proficiency in the native language also makes it easier for the student to socialize with other students and supports their social integration. Finnish is considered to be one of the world's most difficult languages to learn (Selmer & Lauring, 2011), which could explain to some extent why the performance gap between the immigrant- and native-generation students is high in relation to other developed countries. On average in the OECD countries in 2017, a larger share of foreign-born working-age people (37%) were highly educated, compared to native born (33%). However, in Finland the situation is reversed, as the highly-educated shares in 2017 were 32% for foreign-born and 39% for native-born (OECD/European Union, 2018). As a student's education is related to their parent's education, some of the school performance gap could be associated to the educational attainment gap among the working-age population in Finland.

The second hypothesis states that, over time the immigrant-generation students will perform worse in relation to the native students. The first regression, which does not control for compositional factors, does not indicate any clear decreasing or increasing performance trend of the immigrant generation students over time. The performance gap between the native and

immigrant generations has fluctuated rather randomly over the cycles and no common change among the immigrant generations is distinguishable. Immigration in Finland has increased during the beginning of the 21st century and the immigrant population composition has become more diverse, which could lead to a wider performance diversity. Before the 21st century, neighboring country descendants have made up majority of the immigration population in Finland, but lately immigrants have increasingly arrived from further distant countries. The rise of migration into developed countries, including Finland, today is to a large extent driven by the inequality between countries, which means that people from less developed countries move to more developed countries. Additionally, the refugee crises have contributed to the rise in the diversity of the immigrant population in Finland. People that migrate from countries that are culturally more distant and at a lower level of socio-economic development, have a harder time to integrate, which can be further worsened by forced migration (Bhugra, 2004). Kirjavainen and Pulkkinen (2017) also stress that the more different the immigrant student's mother tongue is from the language of instruction, the worse they perform. Sahlberg (2021) asserts that as the cultural diversity rises among the immigrant population it can generate a wider variety in learning ability, which can be reflected in test performance. The compositional change that has occurred during the 21st century in the immigrant population in Finland is not reflected in the results of this study as an increase in performance diversity as expected. The regression results can hence not give support to H2. The random fluctuation over the cycles can potentially be due to that the composition of the PISA sample of students with an immigrant background varies across cycles. The compositional fluctuation can be seen in the fluctuation of the average value of the economic, social and cultural variable (ESCS) across the cycles presented in **Table 2**.

The second regression controls for compositional and background factors to test the third hypothesis. H3 hypothesized that the performance gaps between the native- and immigrant-generation students will decrease but remain. The controls substantially lower all significant immigrant-generation coefficient for all cycles. Despite controlling for the socio-economic status (Kilpi-Jakonen, 2012) and the native language proficiency (Kirjavainen & Pulkkinen, 2017), that are the two major factors behind the performance penalty that students with an immigrant background face, a gap between the native- and immigrant-generation students remains, which confirms H3. The student's integration success is based on a wide variety of factors in addition to the two main individual level variables. The performance gap that remains between the native- and immigrant-generation students can to some extent be

explained by institutional level factors and social factors that are not included in this study. Such as the difference between the school system in the student's country of origin and destination (Kilpi-Jakonen, 2012) and specific factors of the destination country's school system, such as at what age children are required to attend school or whether pre-school is mandatory (Borgna & Contini, 2014). Parents' support and a student's social network play an important role in children's school performance. Parents that do not have proficiency in the language of instruction might have difficulty helping with homework or exam preparation. Immigrant students might not have a diverse social network to turn to for help or tuition. Immigrants in Finland face some hostility and negative attitudes by the native population (Ervasti, 2004), and if experienced within the school system it can harm a student's school motivation. Immigrant students are found to be graded slightly above their true school performance in the Finnish school system to compensate for their lack of language proficiency (Kilpi-Jakonen, 2012). The higher grades could be motivating immigrant students to invest more time in studying but it could also, contrarywise, give them a distorted understanding of their true performance level and less incentive to study harder.

According to the fourth and last hypothesis students with an immigrant background that are raised or born in Finland will perform better than those that have at arrived later ages. The first and the second regression show a positive performance trend among the immigrant generations. The score difference between the native and immigrant generations decreases on average when moving from the lower toward the higher generations, excluding the second generation, which gives strong support for H4. The correlation between age of arrival and immigrants' school performance finds support from Heath's and Kilpi-Jakonen's study (2012), in which they study how age of arrival is related to the PISA reading test score in the PISA participating countries and found that the early arrivals perform better than the late arrivals. A logical explanation to positive math performance trend among the immigrant-generations found in this study is that the more grades a student has attended in the Finnish school system compared to a foreign system, the better they are integrated into the system and the less performance barriers they face. However, there are also factors outside the Finnish school system that can explain the positive performance trend. If a student is raised in the country of destination, they have had more time to learn the native language and socially integrate into the society compared to those who arrive at a later age. Vice versa, the more grades a student has spent in a foreign school system the more it takes to adapt to the Finnish system. A student's school performance is also related to their parents' support. If a student

has arrived in Finland a few years before primary school has started, their foreign-born parents have potentially had time to integrate into the society by the time primary school starts, perhaps by getting a job and learning the language, and are hence able to support their children's learning. Oppositely, if a student arrives just a few years before completing primary school, their parents might still be busy settling and unavailable to fully support their children in school. Native parents are conceivably a valuable support for the student's school performance, as a native parent is proficient in the native language and familiar with the school system. If the student has one native parent than non, they perform better, as the 2,5-generation performs significantly better than the 2-generation students. However, the 2-generation performs on average worse than the 1,75-generation, which can possibly be explained by that the 2-generation in this study includes only students whose both parents are born abroad, while about 20 percent of the 1,75-generation students have one parent that is born in Finland. The 1-generations are not separated further based on whether they have one native parent, like the 2-generations, as the sample sizes in the early years would become too small to analyze.

In addition to what is covered by the hypotheses, second regression shows how the control variables are related to the math performance. The socio-economic status and whether the language of instruction is spoken at home are significantly related to the math performance. In relation to other Western countries, the effect of the socio-economic status on school performance in Finland is low, but as long as it is present there are incentives to invest in equity in the school system. There is a significant benefit of having proficiency in the language of instruction on math performance, despite that instruction language proficiency is less important in math than for example in reading. This study does not account for whether the language of instruction is Finnish or Swedish, the second official language spoken in Finland. To determine whether the difficulty of learning the Finnish language is reflected in school performance, future research could compare the school performance of students with an immigrant background by comparing the Finnish and Swedish speaking schools to determine whether there is a benefit difference in having proficiency in the language of instruction between Finnish and Swedish.

Conclusion

As migration into Finland has risen during the 21st century the share of the population with a foreign background is increasing, which raises the need to comprehensively understand the effect that immigration has on the receiving country. The aim of this study is to get a more detailed perception of the relationship between the immigrant generation and the PISA math score in the high-performing Finnish school system. This study has contributed to what is already known about the relationship in three ways. First, to study the relationship over time. Second, to see if the relationship is explained by compositional and background factors. Third, to analyze the relationship based on a more detailed immigrant-generation definition.

There are two main limitations of this study. First, due to the small immigrant population in Finland and the short immigration history, the first cycles produce small samples for certain generations. Despite that some coefficients were not shown, a positive performance trend across the immigrant generations was found. Second, as the aim was to analyze the relationship over time to include as many cycles as possible, the regression had to be identical for all cycles, which led to that the variables had to be limited to those available and identically measured in all cycles. For example, specific countries for the country of birth variables are not given in the 2003 cycle and rather limited in the rest of the cycles, by categorizing a large fraction of the observations' country of birth as "other". The results of this thesis could be challenged by controlling more for the compositional factors, mainly country of birth or language spoken at home, to see if they can explain the performance gap.

This study found that the immigrant generation perform worse than native students in the Finnish school system. The results can confirm that all the immigrant generations perform on average worse than the native generation in each PISA math test from 2003 to 2018 in Finland. The gap between the immigrant generation and native generation students is large and significant in all cycles. This study additionally contributes with the finding that there is a significant positive performance trend among the immigrant generations. Namely, the first generation benefit in math performance from arriving earlier and the second by being born in Finland and having a native parent. The gap decreases when controlling for compositional and background factors, but remains.

The less immigrants lag behind the native population, the more of a benefit they can be to the receiving country. The immigrant native school performance gap in Finland is large in comparison to other nations and is in need of action by the policy makers. The results of this study emphasize that the gap is significant in all PISA cycles and dependent on the age of arrival of the first generation and whether the second-generation students have a native parent. Much of the immigrant integration work focuses on immigrants' job performance and as a consequence has immigrants' school performance become relatively deprioritized. School performance is strongly correlated with the occupational performance (Hanushek and Woessmann, 2021), and hence, can lowering the performance gap already at a young age solve performance penalties faced by immigrants in later life stages.

The school system serves as an important, if not the most important, institution to lower the performance gap between immigrants and natives in the long run. The school system also serves as the main environment for many immigrant children to feel belongingness in the society. The school system is the first institution where the population develop skills and where the initial performance differences emerge. The aim is not to abolish all performance differences among students, which is not possible nor efficient, but to reduce the association between school performance and specific background factors, mainly having an immigrant background. The Finnish school system has succeeded in promoting equity in school performance, by reducing the association between a student's socio-economic background and their school performance and the performance difference between students. That being said, does the Finnish school system also have the tools to reduce the association between immigrant generation and school performance? The results of this thesis do not show an increase in the performance gap over time, which H2 stated, despite that the immigrant population has become more diverse over the recent years. The Finnish school system has potentially adapted to the compositional change over time and new policies might have improved the system to be more inclusive toward students with an immigrant background. This thesis does not analyze factors related to the school system nor policies, but aims to motivate such topics in future research. Specifically, future research should examine whether recent school policies are effective in improving the school performance of students with an immigrant background in Finland. For example, educating teachers about multiculturalism and making pre-primary school mandatory.

Acknowledgments

I would like to state my gratitude to Stockholm University Demography Unit (SUDA) and the people that have supported me throughout my writing process. Eleonora Mussino, thank you for supervising me in writing my thesis and for sharing your expertise in migration. Andrea Monti and Ben Wilson, thank you for taking the time to evaluate and discuss my thesis. Livia Olah, thank you for always being available to help and coordinate. Sven Drefahl, thank you for making statistics fun and uncomplicated. I also want to thank my family for their endless support and love. Pappa, thank you for teaching me to never give up, and Mamma, thank you for compensating and trying to teach me to fail. Eemeli, my fiancé, thank you for motivating and believing in me but also for challenging my thesis arguments. Finally, I would like to dedicate all the efforts put into this thesis to my big brother, Conrad, who is courageously fighting a rare cancer.

References

Ansala, L., Åslund, O. and Sarvimäki, M. (2020). Immigration History, Entry Jobs, and the Labor Market Integration of Immigrants. *IZA Discussion Paper Series No.13089.*

Behr, A. and Fugger, G. (2020). PISA Performance of Natives and Immigrants: Selection versus Efficiency. *Open Education Studies,* 2, 9-36.

Bhugra, D. (2004). Migration and mental health. *Acta Psychiatr Scand*, 109(4), 243–258.

Borgna, C. and Contini, D. (2014) Migrant Achievement Penalties in Western Europe: Do Educational Systems Matter? *European Sociological Review* , 30(5), 670-683.

Cattaneo, M. A. & Wolter, S. C. (2015). Better migrants, better PISA results: Findings from a natural experiment. *IZA Journal of Migration,* 4(18).

Chircop, D. & Claros, E. (2019). Inclusion of migrants in formal education. *European Parliament Research Service.*

Entorf, H. and Minoiu, N. (2005) What a Difference Immigration Policy Makes: A Comparison of PISA Scores in Europe and Traditional Countries of Immigration. *German Economic Review*, 6(3) 355-376.

Ervasti, H. (2004). Attitudes Towards Foreign-born Settlers: Finland in a Comparative Perspective. *Yearbook of Population Research in Finland* 40, 25-44.

Fandrem, H., Jahnsen, H., Nergaard, S. E. & Tveitereid, K (2021): Inclusion of immigrant students in schools: the role of introductory classes and other segregated efforts. *International Journal of Inclusive Education.*

Finnish Immigration Service (2017). *Turvapaikkapäätökset 1.1.-31.12.2016.* Available from: https://migri.fi/documents/5202425/6162429/2016+Turvapaikkap%C3%A4%C3%A4t%C3%B6kset+(fi).pdf/2ebecb7b-d821-4c77-889e-eb3c70ea14d4/2016+Turvapaikkap%C3%A4%C3%A4t%C3%B6kset+(fi).pdf?version=1.1&t=1554190816000

Hanushek, E. and Woessmann, L. (2021). Education and economic growth. *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Harju-Luukkainen, H., Nissinen, K., Sulkunen, S., Suni, M., & Vettenranta, J. (2014). *Avaimet osaamiseen ja tulevaisuuteen: Selvitys maahanmuuttajataustaisten nuorten osaamisesta ja siihen liittyvistä taustatekijöistä PISA 2012 -tutkimuksessa.*

Harju-Luukkainen, H. and McElvany, N. (2018). Immigrant Student Achievement and Education Policy in Finland. *Immigrant Student Achievement and Education Policy*, 1(1), 87-102.

Heath, A. and Kilpi-Jakonen, E. (2012). Immigrant Children's Age at Arrival and Assessment Results. *OECD Education Working Papers No. 75.*

Heckman, J. J., Jagelka, T. and Kautz, T. D. (2021). Some contributions of economics to the study of personality. *Handbook of Personality,* 853-892. New York: Guilford Press.

IOM (2020). Migration Factsheet No. 1 – Drivers of migration*. World Migration Report.*

Kirjavainen, T. and Pulkkinen, J. (2017). Miten lähtömaa on yhteydessä maahanmuuttajataustaisten oppilaiden osaamiseen? *Yhteiskuntapolitiikka*, 82(4), 430-439.

Kilpi-Jakonen, E. (2012). Does Finnish Educational Equality Extend to Children of Immigrants? *Nordic Journal of Migration Research,* 2(2), 167-181.

Koikkalainen, S. (2021). Finland's Immigration Policy: State Objectives, Local Solutions. *Local Integration of Migrants Policy: European Experiences and Challenges.*

Krishnakumar, P. & Indumathi, T. (2014). Pull and Push Factors of Migration. *Sona Global Management Review,* 8(4), 8-12.

Levels, M., Dronkers, J. & Kraaykamp, G. (2008). Immigrant Children's Educational Achievement in Western Countries: Origin, Destination, and Community Effects on Mathematical Performance. *American Sociological Review,* 38, 835-853.

Logan, S. & Johnston, R. (2010). Investigating gender differences in reading. *Educational Review*, 62(2), 175-187.

Ministry of Education, Children and Youth of the Grand Duchy of Luxembourg (2018). Information about the PISA study 2018. Available from: https://men.public.lu/dam-assets/catalogue-publications/statistiques-etudes/secondaire/pisa-2018-en.pdf

Mussino, E. (2021). Do leave policies impact fertility? The case of immigrants from low-fertility countries in Sweden(Version1).*Stockholm Research Reports in Demography.*

Murat, M & Frederic, P. (2015). Institutions, culture and background: the school performance of immigrant students. *Education Economics*, 23(5), 612-630.

OECD (2009a). Plausible values. *PISA Data Analysis Manual: SAS, Second Edition,* 93-101. OECD Publishing, Paris*,* https://doi.org/10.1787/9789264056275-en.

OECD (2009b). Sample weights. *PISA Data Analysis Manual: SAS, Second Edition,* 49-58. OECD Publishing, Paris*,* https://doi.org/10.1787/9789264056275-en.

OECD (2009c). Replicate Weights. *Pisa Data Analysis Manual: SAS, Second Edition,* 59-77.

OECD (2017). PISA 2015 Technical Report. *PISA, OECD Publishing*.

OECD (2018). *Working Together: Skills and Labour Market Integration of Immigrants and their Children in Finland.*

OECD (2019a). PISA 2018 Assessment and Analytical Framework. *PISA, OECD Publishing, Paris,* https://doi.org/10.1787/b25efab8-en.

OECD (2019b). Annex A1. Construction of indices. *PISA 2018 Results (Volume III): What School Life Means for Students' Lives.*

OECD (2019c). Comparing reading, mathematics and science performance across PISA cycles. *PISA 2018 Results (Volume I): What Students Know and Can Do, OECD Publishing, Paris.*

OECD (2020a). Students' socio-economic status and performance. *PISA 2018 Results (Volume II): Where All Students Can Succeed*, 49-62.

OECD (2020b). Performance and academic resilience amongst students with an immigrant background. *PISA 2018 Results (Volume II): Where All Students Can Succeed,* 177-194.

OECD (2022). *Mathematics performance (PISA) (indicator)*. Accessed on 07 May 2022, doi: 10.1787/04711c74-en

OECD/European Union (2018). Immigrant skills and labour market integration. *Settling In 2018: Indicators of Immigrant Integration,* 63-99.

OKM (2021). Oppivelvollisuus laajenee syksystä 2021 alkaen. *Ministry of Education and Culture.* Available from: https://julkaisut.valtioneuvosto.fi/handle/10024/162873

OPH (2022) *Finnish education system.* Available from: https://www.oph.fi/en/education-system

Poder, K., Lauri, T. and Veski, A. (2017) Does School Admission by Zoning Affect Educational Inequality? A Study of Family Background Effect in Estonia, Finland, and Sweden. *Scandinavian Journal of Educational Research*, 61(6), 668-688.

Rindfuss, R. R. and Brewster, K. L. (1996). Childbearing and Fertility. *Population and Development Review,* 22, 258-289.

Rumbaut, R. G. (2014). Ages, Life Stages, and Generational Cohorts: Decomposing the Immigrant First and Second Generations in the United States. *International Migration Review,* 38(2), 1160-1205.

Sahlberg, P. (2012). Quality and Equity in Finnish Schools. *School Administrator,* 69(8), 27-30.

Sahlberg, P. (2021). *Finnish Lessons 3.0: What Can the World Learn from Educational Change in Finland?*

Samuelsson, K. and Lindblad, S. (2015). School management, cultures of teaching and student outcomes: Comparing the cases of Finland and Sweden. *Teaching and Teacher Education*, 49, 168-177.

Schleicher, A. (2006) Where immigrant students succeed: a comparative review of performance and engagement in PISA 2003. *Intercultural Education*, 17(5), 507-516.

Schwab, K. (2008). The Global Competitiveness Report 2008–2009. *World Economic Forum.* Available from: https://www.weforum.org/reports/global-competitiveness-report-2008-2009

Schwab, K. (2009). The Global Competitiveness Report 2009–2010. *World Economic Forum.* Available from: https://www.weforum.org/reports/global-competitiveness-report-2009-2010

Schwab, K. (2010). The Global Competitiveness Report 2010–2011. *World Economic Forum.* Available from: https://www.weforum.org/reports/global-competitiveness-report-2010-2011

Schwab, K. (2011). The Global Competitiveness Report 2011–2012. *World Economic Forum.* Available from: http://reports.weforum.org/global-competitiveness-2011-2012/

Schwab, K. (2012). The Global Competitiveness Report 2012–2013. *World Economic Forum.* Available from: http://reports.weforum.org/global-competitiveness-report-2012-2013/

Schwab, K. (2013). The Global Competitiveness Report 2013–2014. *World Economic Forum.* Available from: http://reports.weforum.org/the-global-competitiveness-report-2013-2014/

Schwab, K. (2014). The Global Competitiveness Report 2014–2015. *World Economic Forum.* Available from: https://www.weforum.org/reports/global-competitiveness-report-2014-2015

Schwab, K. (2015). The Global Competitiveness Report 2015-2016. *World Economic Forum.* Available from: https://www.weforum.org/reports/global-competitiveness-report-2015

Schwab, K. (2016). The Global Competitiveness Report 2016–2017. *World Economic Forum.* Available from: https://www.weforum.org/reports/the-global-competitiveness-report-2016-2017-1

Schwab, K. (2017). The Global Competitiveness Report 2017–2018. *World Economic Forum.* Available from: https://www.weforum.org/reports/the-global-competitiveness-report-2017-2018

Statistics Finland (2021). Decrease in birth rate stopped in 2020. Available from: https://www.stat.fi/til/synt/2020/synt_2020_2021-04-23_tie_001_en.html

Swedish Migration Agency (2017). *Asylum decisions, Swedish Migration Agency, 2016.* Available from: https://www.migrationsverket.se/download/18.2d998ffc151ac3871592564/1485556054285/Avgjorda%20asyl%C3%A4renden%202016%20-%20Asylum%20decisions%202016.pdf

Swedish Research Council (2017). *Good research practice.*

Tilastokeskus (2004). *Maahanmuuttajien ikärakenne poikkeaa valtaväestöstä.* Available from: https://www.stat.fi/tup/tietoaika/tilaajat/ta_08_04_nieminen.html

Tilastokeskus (2018). *Toisen polven ulkomaalaistaustaisten keski-ikä 11 vuotta*. Available from: https://tilastokeskus.fi/til/vaerak/2017/02/vaerak_2017_02_2018-06-20_tie_001_fi.html

Tilastokeskus (2021). *Birth rate.* Available from: https://www.stat.fi/tup/maahanmuutto/perheet/syntyvyys_en.html

Teltemann, J. & Rauch, D. P. (2018). Immigrant Student Achievement and Education Policy in Germany. *Immigrant Student Achievement and Education Policy*, 35-52.

United Nations (2019). *International Migrant Stock 2019*. United Nations database, POP/DB/MIG/Stock/Rev.2019

Varjo, J., Lundström, U. & Kalalahti, M. (2018). The governors of school markets? Local education authorities, school choice and equity in Finland and Sweden. *Research in Comparative & International Education,* 13(4) 481-498.

Väänänen, T. (2018). *Can immigration save the ageing economy of Finland? The labor market integration of immigrants in the 21st century and the impacts of immigration on public economics.*

Wang, Z. de Graaff, T. and Nijkamp, P. (2018). Barriers of Culture, Networks, and Language in International Migration: A Review. *REGION, The Journal of ERSA*, 5(1), 73-89.

Wilson, B. (2021). The childbearing of immigrants who arrived as children: Understanding the role of age at arrival for women and men. *Stockholm Research Reports in Demography.*

Zinovyeva, N., Felgueroso, F. & Vazquez, P. (2013). Immigration and student achievement in Spain: evidence from PISA. *SERIEs*, 5, 25–60.

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, 8(2), 187-203.

Data references

Graph 1: Statistics Finland, 11vt -- Population according to citizenship, country of birth, language and origin (overlap of the groups), 1990-2020. Available from: https://pxnet2.stat.fi/PXWeb/pxweb/en/Maahanmuuttajat_ja_kotoutuminen/Maahanmuuttajat_ja_kotoutuminen__Maahanmuuttajat_ja_kotoutuminen/maakoto_pxt_11vt.px/

Graph 2: Statistics Finland, 11vw -- Population according to country of birth, age and sex, 1990-2020. Available from: https://pxnet2.stat.fi/PXWeb/pxweb/en/Maahanmuuttajat_ja_kotoutuminen/Maahanmuuttajat_ja_kotoutuminen__Maahanmuuttajat_ja_kotoutuminen/maakoto_pxt_11vw.px/

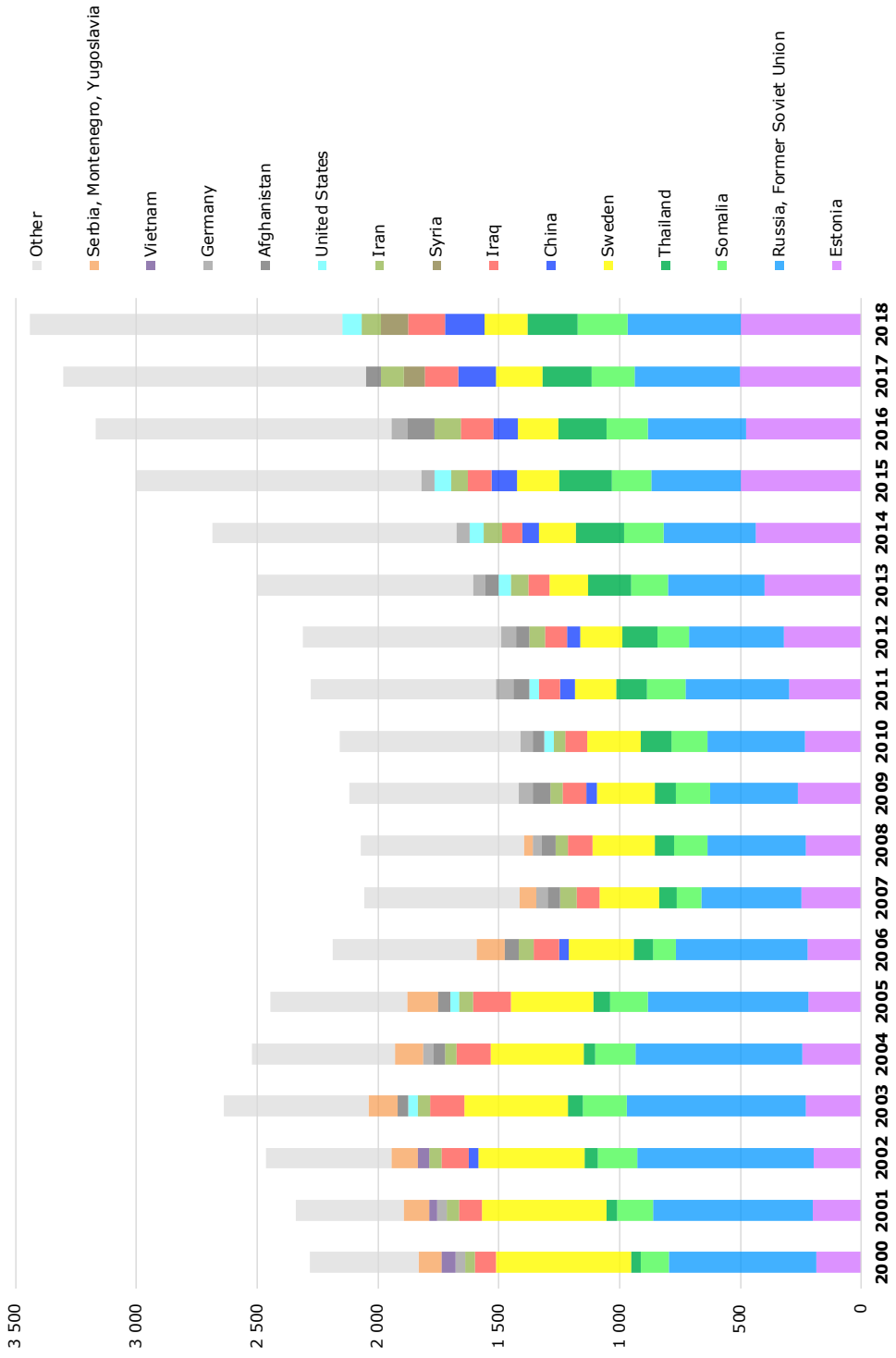
Graph 3: Statistics Finland, 11rs -- Origin and background country by language, age (1-year) and sex, 1990-2021. Available from: https://statfin.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__vaerak/statfin_vaerak_pxt_11rs.px/

Graph A1: Statistics Finland, 11vw -- Population according to country of birth, age and sex, 1990-2020. Available from: https://pxnet2.stat.fi/PXWeb/pxweb/en/Maahanmuuttajat_ja_kotoutuminen/Maahanmuuttajat_ja_kotoutuminen__Maahanmuuttajat_ja_kotoutuminen/maakoto_pxt_11vw.px/

Graph A2: Statistics Finland, 11rs -- Origin and background country by language, age (1-year) and sex, 1990-2021. Available from: https://statfin.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__vaerak/statfin_vaerak_pxt_11rs.px/

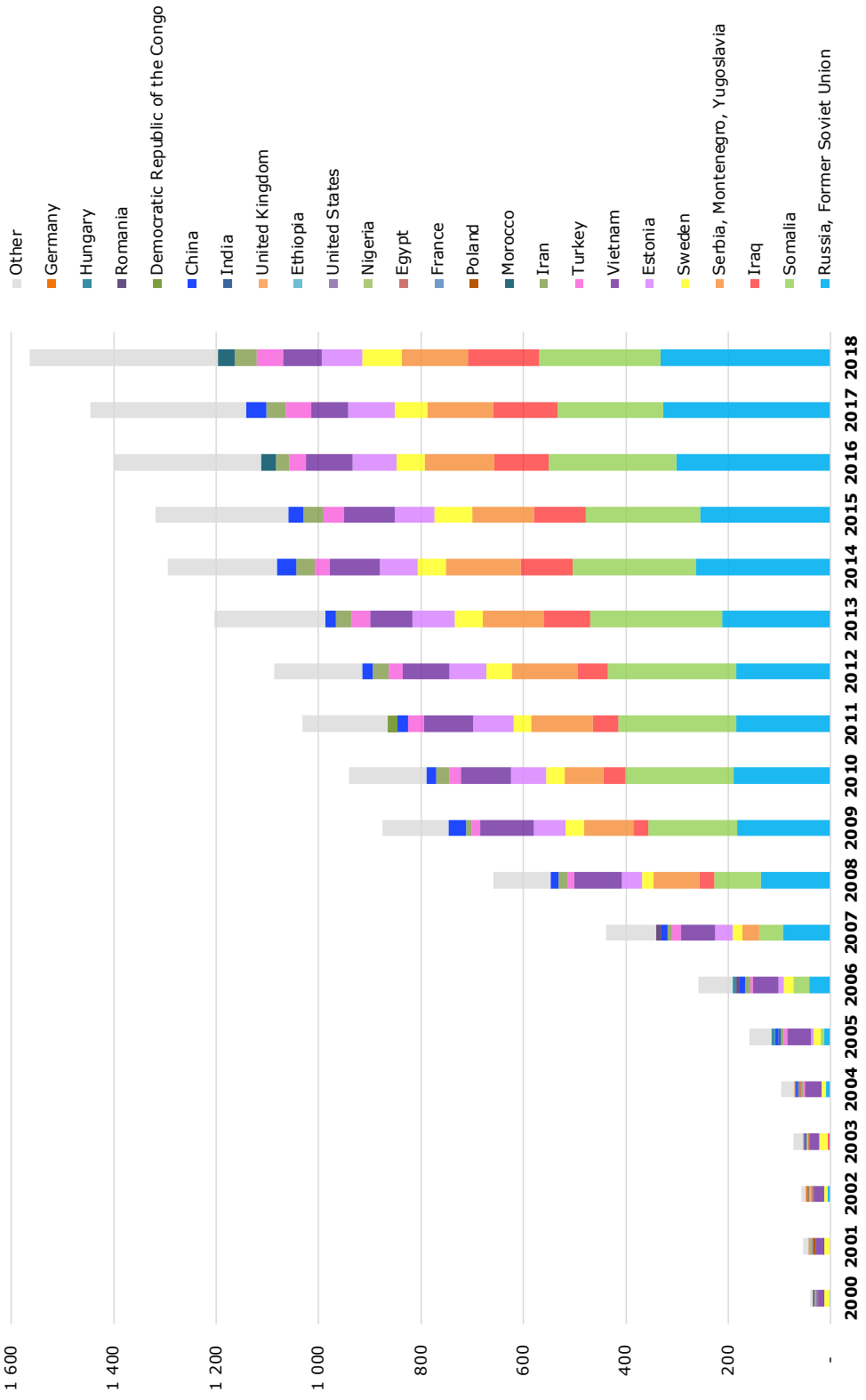
Appendix

Graph A 1: Top 3 country of birth of the mother of the first-generation 15-year-old immigrants in Finland.



Data: Statistics Finland.

Graph A 2: Top 3 country of birth of the mother of the second-generation 15-year-olds in Finland.



Data: Statistics Finland.

Table A 1: Dropped observations due to missing values on country of birth.

Dropped observations	2003	2006	2009	2012	2015	2018
Number of observations	100	53	74	206	128	35
Share of original sample	1,77 %	1,12 %	1,27 %	2,33 %	2,18 %	0,63 %

Table A 2: Dropped observations due to missing values on control variables.

Dropped observations	2003 (#)	2006 (#)	2009 (#)	2012 (#)	2015 (#)	2018 (#)
Native-generation	70	30	40	60	15	20
2,5-generation	12	9	4	14		
2-generation		2		17	1	1
1,75-generation		3	1	14		1
1,5-generation	3	4		9	3	2
1,25-generation				6	1	1
Total	**85**	**48**	**45**	**120**	**20**	**25**

Table A 3: Cross tabulation of the immigrant generation variable used in this study and PISA's.

Generations	Native	Second-Generation	First-Generation	Total
Native-generation	6 413			**6 413**
2,5-generation	791			**791**
2-generation		580		**580**
1,75-generation	103		275	**378**
1,5-generation	57		311	**368**
1,25-generation	7		86	**93**
Total	**7 371**	**580**	**672**	**8 623**

Table A 4: The weighted average values of the ESCS variable.

ESCS	2003	2006	2009	2012	2015	2018
Native-generation	0,253	0,267	0,380	0,389	0,271	0,330
2,5-generation	0,461	0,287	0,371	0,368	0,284	0,338
2-generation	0,249	-0,035	0,281	-0,107	-0,236	-0,147
1,75-generation	0,148	-0,151	0,011	0,021	0,321	0,069
1,5-generation	0,056	-0,176	0,113	-0,256	-0,254	-0,232
1,25-generation	0,252	-0,306	-0,256	-0,380	0,269	-0,309

Graph A 3: The weighted share of the students that use another language than the test language at home.

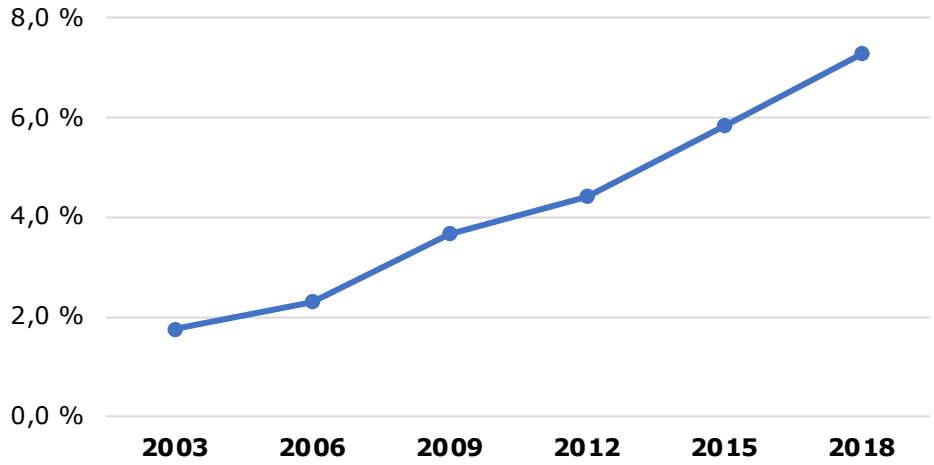


Table A 5: The weighted share of female and male students in the sample.

Gender	2003	2006	2009	2012	2015	2018
Female	50,40 %	50,50 %	50,23 %	49,50 %	48,51 %	49,57 %
Male	49,59 %	49,50 %	49,77 %	50,50 %	51,49 %	50,42 %

Stockholm University

SE-106 91 Stockholm

Telefon/Phone: 08 – 16 20 00

www.su.se

Stockholms universitet