

Selective Friendship?
**A Study of Immigrant Segregation within
Finnish Classrooms**

Ellen Sahlström
University of Helsinki
Faculty of Social Sciences
Department of Economics
Master's Thesis
May 4, 2022

Abstract

Faculty: Faculty of Social Sciences

Degree programme: Master's Programme in Economics

Study track: Research Track

Author: Ellen Johanna Sahlström

Title: Selective Friendship? A Study of Immigrant Segregation within Finnish Classrooms

Level: Master's Thesis

Month and year: May 2022

Number of pages: 48

Keywords: Individual segregation, immigrants, academic performance, Finnish education

Supervisor: Roope Uusitalo

Where deposited: Helsinki University Library

Additional information:

Abstract: This thesis investigates the within-school segregation of Finnish students using survey data on friendships. The difference in academic performance between students with and without immigrant backgrounds is large, and in order to make it smaller, the study environment for students with immigrant backgrounds must be understood. Identifying the extent to which students with immigrant background experience a different social environment is one step in that direction.

Individual segregation is the extent to which the social network of an individual is composed of individuals similar to each other regarding some specific trait. This study investigates the existence of individual segregation in Finnish schools using information on who the fifth grade students participating in the study are friends with. The individual segregation level is calculated based on the background of friends, dividing students into two groups: students with or without immigrant backgrounds. This gives an indication of the possible segregation at an individual level, created through friend choice. Additionally, the correlation between individual level segregation and age at arrival to Finland and academic skills respectively is studied.

Clear evidence of individual level segregation among immigrants is found. Students with immigrant backgrounds are more likely to have friends with immigrant backgrounds and more likely to be lonely, as in have no friends. However, neither correlation between individual segregation and age at arrival nor correlation between individual segregation and academic skills can be found. This could be explained by problems with the data, but can also indicate that peer effects in class are smaller than what was expected based on previous research. It seems that also segregation patterns differ from what has been found in similar American studies. More research need to be done, but this thesis shows that students with immigrant backgrounds experience a different social environment when it comes to friends than students with non-immigrant backgrounds do, as the share of friends with immigrant backgrounds is significantly higher for students who themselves have immigrant backgrounds.

Contents

1	Introduction	1
2	Theory and Related Literature	5
2.1	Segregation Indexes	5
2.2	Previous Research	7
3	Data and Methods	13
3.1	Institutional Background	13
3.2	Data Collection	15
3.3	Data Modification	18
3.4	Descriptive Data	18
3.4.1	Immigrant Background and Sample Size	19
3.4.2	Social Networks	20
3.4.3	Age at Arrival to Finland	21
3.4.4	Academic Skills	22
3.5	Methods	24
3.5.1	Measuring Segregation	25
3.5.2	The Regression Model	26
4	Results	27
4.1	Segregation Levels	28
4.2	Segregation in Relation to Other Factors	35
4.2.1	Correlation with Age at Arrival	36
4.2.2	Correlation with Academic Skills	38
5	Conclusion	42
	References	44

1 Introduction

There is a significant difference in academic performance between Finnish students with immigrant background and Finnish students without immigrant backgrounds (OECD, 2019). In fact, the difference is the largest in the European Union (CEDEFOP, 2020). Little research is conducted on the reasons behind this difference (Bernelius and Huilla, 2021), but it is known that segregation can affect academic performance (Charles, Dinwiddie and Massey, 2004). As the Finnish school system is among the least segregated in the group of OECD countries (Gutiérrez, Jerrim and Torres, 2020), segregation between schools can not provide a direct explanation. However, evidence exists that also segregation on an individual level might affect academic performance (Echenique and Fryer Jr., 2007). Therefore, this thesis aims to investigate the differences in experienced social environments of individual students in Finnish schools through two research questions:

1. Are immigrants within Finnish schools segregated on an individual level?
2. How does the within-school segregation of immigrants relate to their academic performance and spent time in Finland?

Segregation is defined as the systematic and unequal division of people into groups based on specific features, often race (James and Taeuber, 1985). Historically, segregation has been based on laws, but the segregation we see today is mostly not statutory but maintained through social norms. Even if Finnish history does not include racial segregation in law, Finnish society is far from equal and systematic differences between people with and without immigrant background are not only found in school, but also in the labor market (Ahmad, 2020; OECD, 2017).

With the answers to the research questions, this thesis provides new knowledge on segregation in the Finnish education system, focusing solely on individual-level segregation and its correlation with academic performance and age at arrival to Finland. This is primarily

achieved by estimating the level of segregation created by choices of friendship, using self-reported data describing the social networks of fifth grade students (age 11). This kind of an estimation has not before been computed using Finnish data and will provide an idea of the level of segregation. Secondly, correlations between segregation and other factors will be investigated. Using a regression model, this thesis will explore the correlations between age at arrival to Finland and segregation, and similarly between verbal and mathematical skills, and segregation. This provides answers to the research questions exploring the segregation of immigrants within schools and the relation of this segregation to other measurable factors.

Finnish immigration is low in an European context (OECD, 2017), but increasing (Statistics Finland, 2021). Given this increasing amount of foreign-born students and students with immigrant backgrounds, the subject of school segregation in respect to immigrant background is becoming increasingly relevant. The Finnish school system has a good reputation, mostly due to high results in the PISA test in the early 2000's. This has been explained by several factors, among which a relatively socially homogeneous country and an equal and free comprehensive school system (Kupiainen, 2009). As mentioned, the Finnish school system is one of the least segregated among the OECD countries (Gutiérrez, Jerrim and Torres, 2020). Despite this, the academic results gap between immigrants and non-immigrants is high, also in comparison to other countries. Previous studies, mostly in the educational sciences, have shown that even if school choice in theory is restricted, ways to choose school through specialized classes exists and is frequently used especially among parents with high education and in well-off families (Berisha and Seppänen, 2017; Kosunen et al., 2020). The between-school segregation might be low, but the fact that some students are more likely to attend specific classes contribute to segregation inside the schools, a contribution that I find can be seen also on the level of individual friendships.

It is easy to find evidence of active class choice through classes with a subject specialization, and in some extent school choice through the availability of these classes. According to the studies mentioned above, specific groups of students, often with high-educated parents or

high family income, are more likely to actively choose a specific class than students with lower socioeconomic status. The heterogeneity in active class choice creates differences in the student distribution between classes and schools and, following that, raises the question of segregation within the school. Despite this, it is difficult to find research results regarding individual segregation within schools in Finland.

Research on school segregation has instead focused on between-school segregation. The conclusion of Söderström and Uusitalo (2010) stating that active school choice increases segregation is a common conclusion in the field. Results like this creates incentives to assume that active class choice in turn affects segregation within the school. In the few cases where this kind of within-school segregation is studied, the focus lies on differences between classrooms. Segregation in Finnish schools has yet not been studied at an individual level. A reason for this may be the shortage of relevant data for the matter. Finnish administrative data covers the population and contains information on a large number of subjects, but does not include data on networks of individuals necessary for studying individual level segregation. This study will use recent data collected on friendship networks of fifth grade students. This allows exploring the matter of individual-level segregation, referring to the extent to which an individual interacts with people with similar backgrounds.

We know that education has a positive effect on future labor market outcomes. A well-functioning educational system could thus help reduce the mentioned labor market inequalities. This makes it a reasonable goal to provide an education of the highest possible quality to all children attending. The measurement of this goal is difficult, but several findings supporting the idea that social ties have great value for later life outcomes (Mele, 2020) give clear evidence that segregation should be taken into account and prevented in the education system. The fact that exposure to students of different backgrounds has positive impact on the attitude towards people different from oneself (Bursztyn et al., 2021) further validates this. The significant gap in the PISA test results between immigrants and non-immigrants gives reason to question how well integration works in the Finnish schools. Studying the

social networks of students in classes with a significant number of immigrants can help us understand the reasons why immigrants and student with foreign background perform worse than students with non-immigrant background.

In the following sections, the details of how the research will be conducted, the methods and the results are presented and discussed. The second and following chapter gives an overview of what is previously known on this subject, including a presentation of indexes usually used for estimating segregation and an review of literature discussing segregation in schools. The third chapter presents the data and the methods used, including background on the Finnish schooling system and Finnish immigration. The results are presented in the fourth chapter, including also discussion on potential problems, policy consequences and further research. The sixth chapter concludes.

2 Theory and Related Literature

The research question of this thesis involves two parts: empirically estimating the possible existence and level of individual segregation, and the correlation of segregation with specified student characteristics. Both aspects require knowledge of some background properties for an effective treating. The matter of segregation indexes has been extensively discussed in previous literature. It would be exaggerating to claim that consensus exists, but some points are agreed on. The same holds for the theme of segregation. Extensive research has been done on the effects of segregated societies and also more specifically on segregated schools. We know that certain factors affect the probabilities for a smooth integration of immigrants into their new societies and these will be examined more carefully in the second subsection of this chapter. The first subsection will discuss the theme of measuring segregation through segregation indexes.

2.1 Segregation Indexes

Measuring segregation relies on the fact that segregation is about two groups' entanglement or distance from one another, the groups defined by some specific feature. A segregated society completely separates two groups, whereas a society without segregation has no separation between groups. As long as segregation has been an issue, studies have tried to develop proper indexes for its measurement. No perfect measure has yet been found. This is further complicated by the fact that "segregation" in everyday use can refer to different things. This thesis will therefore study the individual within-school segregation using a simple measure of segregation based on a defined group's share of a student's friends. This measure is intuitive and fairly easy to understand and calculate. The calculating of the chosen measure is presented in detail in Chapter 3.

Every way to measure segregation will at some level be imperfect, due to the complicated nature of segregation. All indexes measuring segregation relies on the idea that people can

be divided into (two) mutually exclusive groups based on some trait in the person. In the case of this thesis, and many other studies, the trait used for division is immigrant background. This view and division is of course highly simplifying and will leave out many details regarding segregation and the reasons behind it. An imperfect measure is however mostly better than no measure, and segregation indexes provide different ways to identify and quantify segregation.

In many fields, including those of education, sociology, economics and statistics, there has been discussion on the properties of good indexes for segregation. Several articles are published presenting new indexes, also fairly recently (Echenique and Fryer Jr., 2007; Ballester and Vorsatz, 2014). In addition, articles discussing and comparing indexes and their properties are quite common, including newer as well as older papers (Allen and Vignoles, 2007; Duncan and Duncan, 1955; Philipson, 1993; Yao et al., 2019). An extensive review of different indexes for segregation is found in the article by Bojanowski and Corten (2014). They define two properties through which they examine different indexes measuring segregation. The properties in this case are symmetry and "insensitivity to adding isolates". Notable is that many of the indexes studied do not comply with both of these properties, giving an indication of the difficulty of creating a good index.

There are several indexes that could have been appropriate to use. These include at least the Freeman Segregation Index (FSI) (Freeman, 1978), the Index of Isolation, the Dissimilarity Index (Duncan and Duncan, 1955), Coleman's Homophily Index (Coleman, 1958) and the Spectral Segregation Index (SSI) (Echenique and Fryer Jr., 2007). These indexes have all been used for measuring segregation in order to answer similar questions as in this thesis. Further research on the subject can evaluate whether a more advanced index might provide a more exact measure than the one used in this study.

2.2 Previous Research

Apart from the difficulties in measuring, segregation can occur between any two or more groups defined based on attributes such as gender, economic status, age, race or language skills. Group belonging is not completely straight forward, and in some cases, the group defining attribute can change. Segregation is additionally present in all stages of everyday life. Housing segregation is probably the most discussed form of segregation, and something that also affects segregation elsewhere. If housing is highly segregated, and school choice is determined based only on residence, school segregation will also be high. The Finnish school system consists mainly of municipal schools, to which children are chosen based on where they live. Therefore segregation in housing can have a large effect on between-school segregation. This section will focus on presenting previous research on school segregation and especially within-school segregation between immigrants and non-immigrants, but nevertheless will also segregation on a broader level be discussed. Some interesting results are also found regarding segregation between rich and poor.

The for this thesis most relevant previous results are presented by Echenique and Fryer Jr. (2007) and Bleakley and Chin (2010). The article by Echenique and Fryer Jr. presents the segregation index SSI as a new way to measure segregation, but includes also a study of individual-level segregation in American schools, which is of high relevance for this thesis. Bleakley and Chin do not study social networks, but provides results regarding factors behind segregation, which are of high interest when answering the second part of the research question, namely factors that correlate with, or even create, differences in the individual level of segregation.

Echenique and Fryer Jr. (2007) present an application of their measure of within-school racial segregation and regress several different outcomes on this segregation index. The data used to calculate the index comes from the AddHealth database (The National Longitudinal Study of Adolescent Health), a database commonly used and well suited for the kind of research undertaken, as it contains network data on friendships as described on the

project website (<https://addhealth.cpc.unc.edu/>). The friendship data used is very similar to the data collected for this thesis, with every student listing his or her closest friends. Differently from our survey, the AddHealth friendship data is based on every student listing their five closest male and female friends respectively, providing a maximum total friend count of 10, compared to 3 in the survey conducted as part of this study.

Using this data, Echenique and Fryer Jr. estimate school-level and individual-level segregation using the segregation index SSI. The school-level segregation is based on an average of the SSI of all "connected components" (i.e. loose groups of friends) for all racial groups. This method and the results it provides allow Echenique and Fryer Jr. to show that segregation does not grow linearly as the school share of students in a specific group grows. Instead, they show that especially black students experience almost complete segregation when the share of black students in the school is above 25%. A similar procedure is conducted on class level in this thesis. Echenique and Fryer Jr. also show that black students with a higher individual level of segregation have lower test scores and that qualities usually connected to a specific minority group amplifies for more segregated students. This indicates that segregation would be correlated with academic performance, providing a hypothesis regarding the correlations studied as part of this thesis.

The study by Bleakley and Chin (2010) addresses the question of how language skills, in this case English, as the study is conducted in the U.S., relates to and affect integration, measured through a couple different outcomes, including marriage and area of residence. This paper does not discuss segregation through the segregation indexes discussed above, but uses social variables such as marriage and housing geography to estimate the level of integration into the society. They use the psychological fact that language skills are more easily acquired for children below a certain age and add immigrants from English-speaking countries as a control group to rule out other effects of age at immigration to a new country. Through an instrumental variable approach, Bleakley and Chin show that the probability of marrying a person born in the US grows significantly as English skills grow, and the English

language skills of the spouse are remarkably better. Also income and years of education are positively affected by arriving earlier, which in turn is correlated with higher English language skills. Also this strengthens the hypothesis that a correlation between language skills and individual segregation as well as age at arrival and individual segregation might exist.

Individual level segregation has not previously been measured in Finnish schools. However, Holfve-Sabel (2015) has used the SSI (Echenique and Fryer Jr., 2007) to analyze within-school segregation in the area of Gothenburg, Sweden. Her results are in many ways compatible with a Finnish setting, especially as the data is gathered before the Swedish school system reform that allows free school choice was undertaken. Nowadays, Swedish school data is not necessarily comparable with Finnish data, as the Swedish school system allows free choice in a way that the Finnish system does not. It should also be taken into account that immigration has been and is on very different levels in the two neighboring countries. Holfve-Sabel investigates whether the forming of friendships in the classroom can "neutralize" segregation through the individual's choice of working mates. The study uses 6th grade student's ranked answers on which three classmates they prefer to work with in class, practically very similar to the data used in this study. It is then shown that the segregation level for these working networks does not depend on the segregation level of the school, even if the area is well known as being segregated and significant differences in between-school segregation is detectable. This result presented by Holfve-Sabel (2015) partly contradicts the findings of Echenique and Fryer Jr. (2007), as it seems that individual segregation in Swedish schools is not particularly affected by the school share of students with immigrant backgrounds.

Even if the exact setting used in this study has not been used to study segregation and especially within-school segregation in Finland, previous literature in economics, psychology and educational sciences include studies on school choice (Kosunen et al., 2020; Berisha and Seppänen, 2017) effects of age at arrival (Ansala, Hämäläinen and Sarvimäki, 2019) and

well-being of immigrants compared to non-immigrants (Strohmeier, Kärnä and Salmivalli, 2011; Matikka et al., 2015). The papers directly discussing effects of being an immigrant all conclude that having immigrant status, explicitly or in the society, have negative consequences for a series of life outcomes. Kosunen et al. (2020) show that even though the Finnish school system in theory have very scant possibilities of school choice, some schools are more attractive and attract students from other school areas. This applies also on class level, such that school results vary significantly between classes. These results in turn are connected to certain admission criteria determining which students are admitted to the class (Berisha and Seppänen, 2017). This kind of within-school segregation is not what this thesis mainly focuses on, but sheds light on possible weaknesses of the presumed equal Finnish school system. This gives an indication of possible individual level segregation, as within-school segregation on class-level seems to exist.

Looking outside Finland, the literature on the field of segregation, school and age at arrival is extensive. Articles on age at arrival include Gonzalez (2003) studying the benefits of arriving prior first grade, Hermansen (2017) using a sibling's fixed effects model and finding significant benefits of early arrival in a Norwegian setting and Lee and Edmonston (2011) measuring the effect of age at arrival among Asian immigrants to North America. Lee and Edmonston also conclude that age at arrival is one of the single most important predictors for life outcomes of immigrants, such as education level, income, and wealth. This result is in no way controversial and the field agrees highly on the fact that age at arrival matters. This increases the likeliness that this study will show a correlation between arrival age and student segregation.

Another study that in many ways reminds of what this thesis aims to do is done by Mele (2020). Mele uses the SSI developed by Echenique and Fryer Jr. (2007) and an equilibrium model to study how attempts to desegregate by moving students between schools affect individual segregation. They find that American students clearly prefer homophily and make friends with students from the same racial groups, creating segregation on an individual

level. Different groups show slightly different tendencies on this point. Based on his results, he points out that schools with equal shares of different groups will be the best compromise what comes to desegregating schooling, even if it might not have much impact. The findings are similar to those of Echenique and Fryer Jr. (2007), but contradict the Swedish findings of Holfve-Sabel (2015).

We know from Alan, Duysak, Kubilay and Mumcu (2021) that there are significant differences in how the networks of refugee students and host students are formed. Refugee students in the Turkish schools studied have fewer social ties than non-refugee students in the same schools, and these are highly affected by the teacher's prejudice. Alan has earlier showed that a certain program aiming to assist the integration of refugees in Turkish schools has a remarkable effect on several outcomes, among them lowering segregation (Alan, Baysan, Gumren and Kubilay, 2021). This study shows that within-school segregation is a phenomena that can be tackled and supports the idea that within-school segregation should be studied in order to develop policies preventing it. The findings made in the Turkish setting support the findings of Currarini, Jackson and Pin (2009).

Currarini, Jackson and Pin (2009) form a framework for analyzing friendships and their formations. As many of the other similar studies they use data from the AddHealth project, providing information of friendships in U.S. high schools. Their analysis shows that students who belong to a majority in the school form more friendships in absolute numbers than majority students, and a larger share of these students' friendships are with students from the same group. This holds also for minority group students in the sense that also their friendships tend to be with students from the same group. Thus, also this study shows that the social networks formed by students in school tend to be segregated. The data used is the same as in Echenique and Fryer Jr. (2007) and the result is therefore not surprising. As the methods used are slightly different this gives further evidence that this phenomena is real.

As seen above, the previous research on the exact subject of within-school segregation is limited, especially in Finland. It seems students from the U.S. experience a high degree of segregated networks. These results are however not replicated in Sweden, where students to some extent neutralize between-segregation by forming friendships across groups (Holfve-Sabel, 2015). In addition to the within-school segregation, this thesis will also study the effect of age at arrival and academic skills on social networks. Bleakley and Chin (2010) among others show that age at arrival matters for life outcomes as language skills and area of residence. No previous studies on within-school segregation using social networks has been done in Finland. The literature on which factors affect the forming of social networks is sparse across the world. As the importance of friendships overall and friendships across groups is clear (see i.e. Bursztyn et al. (2021)), this is a field that require further research.

3 Data and Methods

This study relies on detailed survey data gathered in Finnish schools. To be able to interpret results correctly, an understanding of the data and the methods is a prerequisite. The Finnish school system is in many ways unique, as is the survey data used. The methods, on the contrary, are simple and commonly used in similar studies. This chapter presents the practical background needed, including the Finnish school system, data gathering and methods used.

3.1 Institutional Background

Finnish children usually begin their education path in early childhood education and care at a young age, of Finnish 2-year old children 70 % participate in early childhood education and care (Säkkinen and Kuoppala, 2021). The compulsory education starts the year the child turns 6 with one year of pre-primary education. This first year is followed by 9 years of comprehensive schooling. The children thus begin school at an age of 7, and the fifth grade students participating in this study are around 11 years old. Starting with students graduating from comprehensive school in 2021, post-comprehensive schooling is mandatory until the age of 18. Participation in post-comprehensive education happens in a general or vocational upper secondary school. All compulsory schooling is free of charge and children are allocated to a publicly funded school in their area of residence (Varjo, Lundström and Kalalahti, 2018). Students or parents who wish to choose another school can apply for it, but are not guaranteed a place in a school outside their area. There are some possibilities to choose school, directly in form of a few private schools, or indirectly through applying for one of the classes focusing on special subjects, where music and language specialization classes are among the most common. These classes usually accept students based on different forms of ability tests. At the age of 16, after 9 years of comprehensive schooling, students apply to a post-comprehensive school and are accepted based on grading or in some cases a combination of grades and an entrance exam. This is the first active school choice, when not taking into account school changes following moving to another area or

school choice through specialization classes.

Figure 1: Background country of 11 year old children with foreign background in Finland.

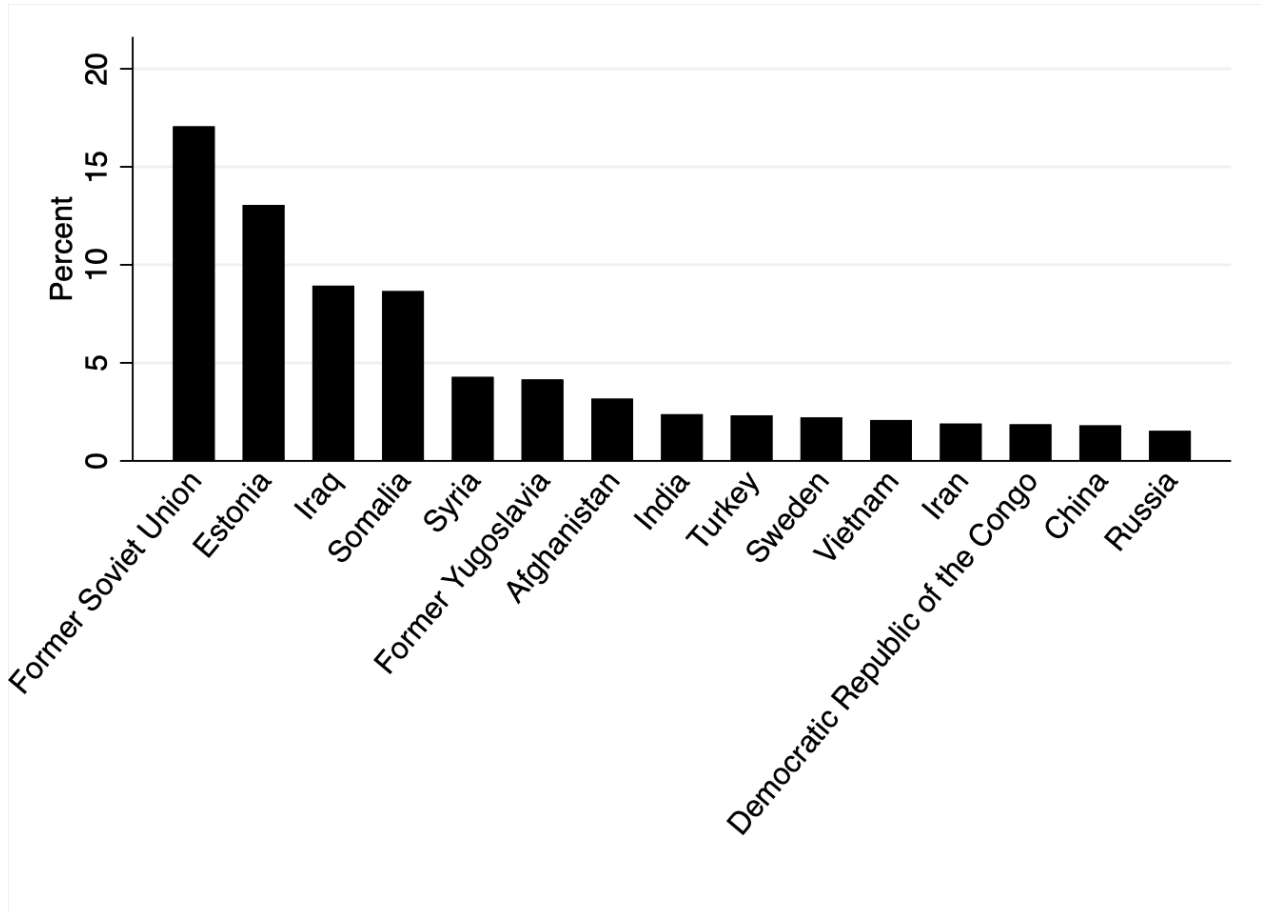


Figure Note: Data from Statistics Finland, situation year 2020: “11vy – Population according to origin and background country, age and sex, 1990-2020.” Only the 15 most common background countries are showed. (n = 5668)

Immigration to Finland has been lower than to most other European countries and the share of the population with foreign background is low, still if growing. Finnish net migration has been negative as late as in the 90s. The total population of Finland is approximately 5 million people. Of these, less than 450 000 have foreign background (2020). Among persons with foreign background, the most common is a background in the former Soviet Union. The second greatest foreign background group is from Estonia, still significantly greater than the following groups. These two are followed by Iraq, Somalia, Former Yugoslavia and China. When restricting the analysis to children of age 11, Syria and Afghanistan are added to the

list of most common background countries, while China falls of the list. This reflects the total population from which the sample used in this study is chosen. (Statistics Finland, 2021)

Figure 1 shows the background country of Finnish 11 year old children, in the case the background country is not Finland. Figure 2 shows a similar but simplified graph over the background countries of students answering the survey this thesis is based on. To protect the integrity of the respondents, students with backgrounds in Sweden, Norway and Denmark are combined into a group called Scandinavia. Students answering "Other" or "I do not know" are not showed in the graph, which of that reason does not sum to 100 %. Notable is that *Somalia* was not an answering option in the survey. A comparison of the two graphs permits the conclusion that the survey sample is representative for the age group, even if the composition is not identical.

3.2 Data Collection

The data for this study is collected as part of an experiment setting. The experiment is conducted as part of the research project *Preventing Social Exclusion* led by a group of researchers from Harvard University and Aalto University, in collaboration with the Finnish Education Evaluation Center (FINEEC), Centre for Learning Analytics at the University of Turku and Walter ry. The main goal of the research project is an evaluation of an intervention done by the organization Walter ry, aiming to prevent exclusion through workshops for students and teachers. The subjects of the study are students in fifth grade in large Finnish cities. The 92 schools chosen to participate all have a substantial number of immigrants (more than 8 per grade) and at least three parallel fifth grade classes. For the evaluation, the participating schools are randomly assigned to one of three groups, to the group receiving Walter workshops on inclusion for the students (15 schools), for the teachers (25 schools), or to the control group. To allow for evaluation of the workshops two surveys are conducted during the school year. This thesis uses data from the baseline survey conducted at the beginning of the school year in the fall. The same students answer to another survey in the spring,

Figure 2: Background country of sampled children with immigrant background.

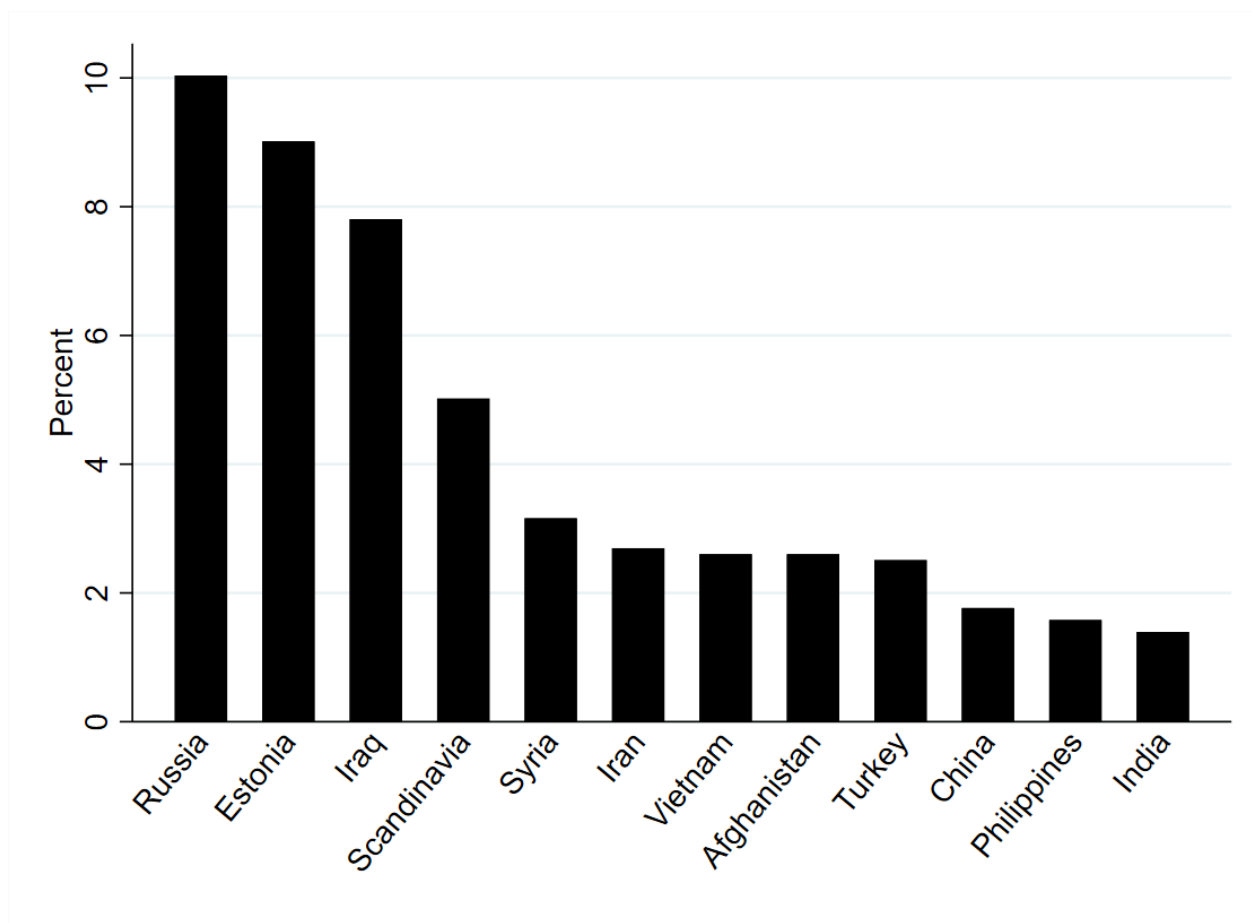


Figure Note: Data from student survey showing the background country of students reporting immigrant backgrounds. All answering options are shown, with the Scandinavian countries presented as one to protect the integrity of the respondents. Students answering "Other country" or "I do not know" are excluded. Note that Somalia was not an answering option, possible respondents from Somalia are thus not showed. (n = 540)

before the end of the school year. The project lasts 3 years and spans over three age cohorts. The data collection relevant for this thesis has been done in September 2021. This data was collected before any of the workshops were held and this thesis does thus not comment on the experiment, but provides a descriptive analysis using data collected by the project.

The survey consisted of three parts. Two of the survey's three parts were answered by the students and one by the teacher's teaching the students participating in the student survey. The students' first part covered bias, bullying, soft skills and social networks and the second part language and mathematical skills. This thesis will use data collected on social networks and academic skills. Information on age at arrival to Finland and immigrant background,

based on birth country of student and parents, is collected through the survey. The data on background will at a later stage in the project be complemented by administrative data on the same subjects, but for the purpose of this thesis, only survey data is accessible.

Information on social networks is gathered through students listing their peers given certain criteria. For each criteria, the students can choose 0-3 classmates from a provided list. In this thesis, the social networks will be created solely based on the answers to the following statement: Select up to 3 classmates who are important to you. This allows to count the number of friendship nominations each student receive, a calculation that provides the foundation for the segregation analysis. The ethnic background of the student is defined based on questions on birth location of student and parents and a question on age at arrival to Finland, where no difference is made between years before school. All these questions are multiple choice questions, with the most common countries among Finnish immigrants as options for the question on birth location, including the options "Other" and "I do not know". These students have been interpreted as student with immigrant backgrounds.

The survey received negative feedback and publicity during and after the period of data collection. Some teachers and parents presented critique on the survey being racist and unpleasant to answer. The data does however not show signs of great attrition in the survey as whole, and most questions have an answer rate around 85 %, meaning around 5000 respondents. However, the last question on background countries has a significantly lower answering rate, with only around half of the sampled students answering. This has led to some necessary modification of the data. The reason for this can be the reported unpleasant experience, but it may as well be that some students simply have not had the time to finish the survey.

3.3 Data Modification

To allow for analysis despite some missing background data, some modifications has been made. To adapt for the missing data, all students with unknown backgrounds are removed. They are also removed from the friend count, such that a student reporting three friends, one with immigrant background, one with non-immigrant background and one with unknown background, will for the purpose of the segregation measuring only have two friends in the data, half of the friends having immigrant background and half of the friends with non-immigrant background. This removes information, but limits the risk of defining someones background incorrectly, which would create another kind of bias, as nothing points to the group with unknown background being homogeneous in other senses than that they have left the question unanswered.

The total class size includes similarly all students with known background, excluding those who have not answered to the last question. This means that every class is reduced to the students with known background. For example a class with 24 students, of which 8 have reported immigrant background, 8 have reported non-immigrant background and 8 have not answered the relevant question, will receive a class share of immigrants of 50 %. This assumes that the students with unknown background are evenly distributed into the two groups. There is no reason to assume this is the exact case, but it is more likely than the students with unknown background belonging completely to one of the groups. Among several bad solutions to this challenge, this was found to be the least problematic.

3.4 Descriptive Data

The final dataset used for the main study conducted by the research group will contain of data from two resources: the surveys described above and administrative data from Statistics Finland. At the time of writing this thesis, administrative data has not yet been made available to the research group. Therefore, data on background characteristics are insuffi-

cient (immigrant background, age at arrival) or not at all available (gender, family background).

Characteristics interesting for the study of this thesis consists of four parts: immigrant background, age at arrival to Finland for students with immigrant background, academic skills, and persons involved in the immediate social network.

3.4.1 Immigrant Background and Sample Size

Several ways exist to define immigrant background. A common definition, used among others by Strohmeier, Kärnä and Salmivalli (2011), is to create three categories, first-generation immigrants, second-generation immigrants and natives. First-generation immigrants are born abroad with at least one parent born abroad. Second-generation immigrants are born in Finland, but have at least one parent born abroad. Students not identifying in one of these two categories are categorized as natives. This implies that students born abroad with parents born in Finland counts as natives. This is not completely intuitive, but as the group counts for under one percent of the sample, no further focus is put on it and this definition will be used. For the main analysis of this study, this definition will be used in a simplified way, creating one group of everyone with immigrant backgrounds (first- or second-generation) and one group consisting of the students with non-immigrant backgrounds. The sample sizes are seen in Table 1. Notable is, that the response rate on this survey question has been approximately 40 % lower than the response rates on the survey as whole. The full sample number includes also students for whom data on immigrant background is missing. Data on the immigrant background of these remaining students will be added first when administrative data is accessed, and in this thesis, only students reporting their backgrounds and answering the background question will be included in the analysis.

Table 1: Sample characteristics defined by self-reported birth country.

1st generation	Immigrants		Non-immigrants			Sample size
	2nd generation	Total	Born abroad	Born in Finland	Total	
925	214	1139	25	1625	1650	2789

Table Note: Backgrounds of the students participating in the study, based on survey data. Division into groups as defined in Section 3.4.1.

3.4.2 Social Networks

A key feature of this thesis will be the usage of data on social networks. As described above, the data used will rely on answers to the question "*Select up to 3 classmates who are important to you.*". This gives data on two things: which classmates that are the most important to you, and the classmates that think you are the most important to them, in other words, we receive two-dimensional data on social networks. Using this information, networks can be formed and the index of segregation can be calculated.

Table 2 shows summary statistics regarding friend nominations from classmates. The average student is nominated by 1.72 other students, and 0.73 of these have immigrant backgrounds. This is in line with the share of immigrants in the sample. Both the total amount of friendship nominations and the share of nominations made by immigrants differ significantly for students who themselves have immigrant backgrounds. Students that have immigrant backgrounds have less friends overall, but a larger amount of friendship nominations by other students with immigrant backgrounds. In addition to this, 20 % of the students included do not get any nominations at all. Of these, just below half are students with immigrant backgrounds. This is a slight but significant over representation, comparing to the total immigrant share in the survey being around 40 %.

The differences in nominations by immigrant background indicates that the question of segregation is meaningful to study. Most students do get nominated. The survey data is thus consistent with data used in similar studies as discussed in the previous section, and is well suited for the planned analysis.

Table 2: Friendship nominations overall and by students with immigrant background.

	Overall mean	SD	Immigrant difference	(s.e.)	Obs
Nominations	1.72	(1.38)	-0.35***	(0.06)	2,684
Nominations by immigrants	0.73	(0.93)	0.26***	(0.04)	2,430

Table Note: Friendship nominations for all students participating in the survey, in total and by immigrant background. Immigrant difference computed using a regression model. Each student was asked to nominate 0-3 friends. See exact description in Section 3.4.2.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

3.4.3 Age at Arrival to Finland

The survey gathers information on first-generation immigrants' age at arrival to Finland, through a question answered by the fifth grade students who also report being born in another country. This characteristic is a part of the administrative data that will be accessed later, and the final research project will use a more exact estimate. At this stage, administrative data is not accessed and this thesis uses self-reported data based on the survey. However, even if the data existing is insufficient, it is consistent with what can be assumed. The answer choices to the question "*When have you moved to Finland?*" were six. The first alternative was "*Before School*" and the other five respectively "*In Grade x*", x ranging from first to fifth grade. A majority (80 %) of the students asked to answer this question report that they have moved to Finland before school. The administrative data will later give more exact information on the arrival year for these students, allowing insight in exact year of arrival also for the years before school. The distribution between grades 1-5 is even, as seen in Figure 3.

Variance in answers to this question is a requirement for the analysis of correlation between segregation and age at arrival. Given this result, analysis can be conducted and possible differences in segregation for children arrived at different ages will be possible to detect, even if the small sample size makes it challenging to identify significant results.

Figure 3: Histogram over year when moved to Finland for first-generation immigrants arriving during or after first grade.

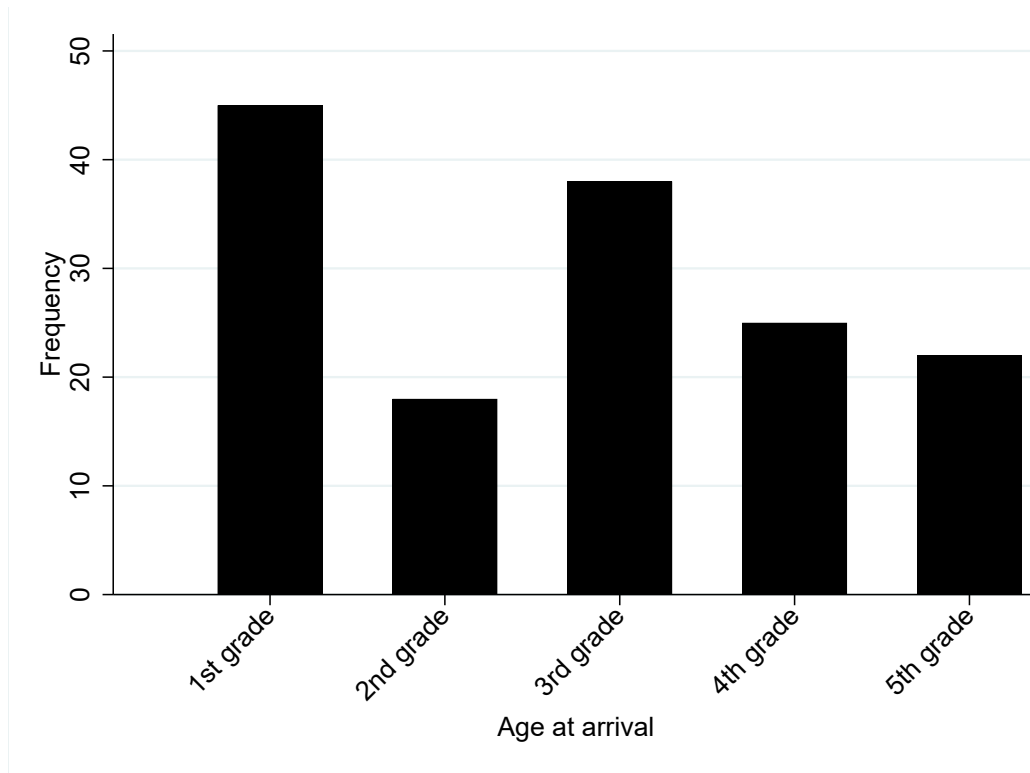


Figure Note: Grade at arrival to Finland relying on self-reported data from survey. Students arriving before school are left out. (n=142)

3.4.4 Academic Skills

The correlation between academic skills and segregation is included in the last part of the research question of this thesis. Data on academic skills has, as the rest of the data presented, been gathered as part of the survey, using a test developed by the Centre for Learning Analytics at the University of Turku in collaboration with the University of Helsinki and Åbo Akademi. The test is newly developed as part of the project FUNA (Functional Numeracy Assessment) and provides an online tool for assessment of functional numeracy such as multiplication tables and verbal skills through reading comprehension and verbal problems (Oppimisanalytiikan Keskus, 2022). Over large groups of students of different age, the test allows us to clearly see increasing results with increasing age, but also increasing variation, indicating that the test measures not only innate abilities, but actually academic skills (Räsä-

nen et al., 2021).

The measures of specific matters, such as vocabulary and word matching in case of the language part, have for this study been combined into two standardized measures, each including all relevant subsections of the test. The first measure gives a normalized value for language skills, the second gives a normalized value for skills in mathematics. In Figure 4, the distribution of the standardized language skills score is showed. Most students perform in a similar way, but the variation required can be identified. Figure 5 shows the standardized score for mathematical skills. The variation is greater than regarding language skills. Note that values more than 4 standard deviations from zero have been deleted. This counted for a small proportion of the students and removing them ensures that the regressions are not driven by outliers. Either way, the measures have enough variation to allow for us to try to answer the research question, which is the most important at time being.

Figure 4: Distribution of the standardized measure of verbal skills.

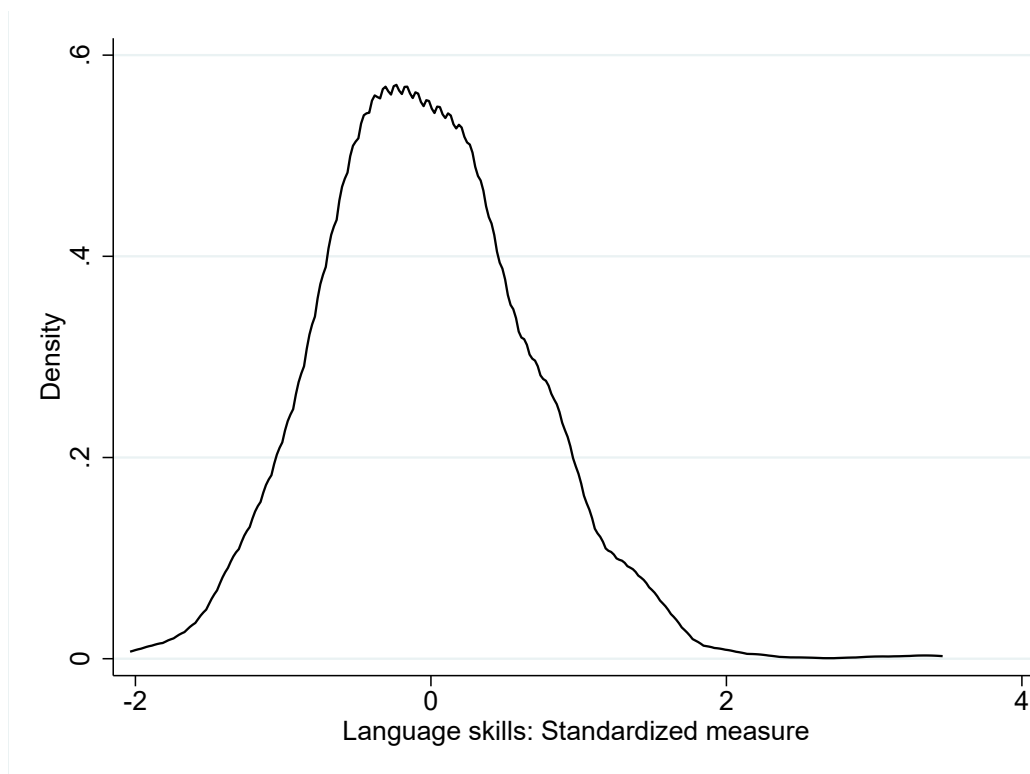


Figure Note: Data from survey, using the FUNA test and a standardized measure for a combined measure of language skills. Only results between -4 and 4 are included. (n = 1950)

Figure 5: Distribution of the standardized measure of mathematical skills.

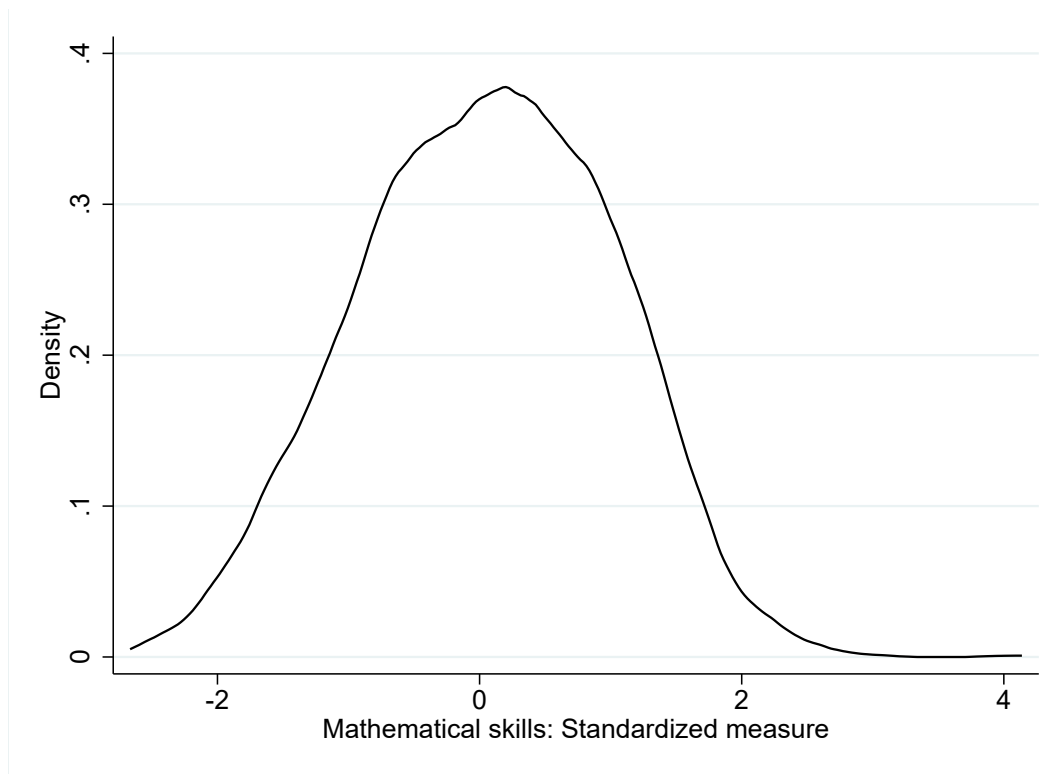


Figure Note: Data from survey, using the FUNA test and a standardized measure for a combined measure of mathematical skills. Only results between -4 and 4 are included. (n = 1982)

3.5 Methods

The main goal of this thesis is to, using the previously described collected data, calculate an index of segregation and estimate the correlation between the index and some known characteristics of the students. This thesis does not attempt to find or explain any causality, and the methods used are thus not so advanced. In short, a measure based on the number and identity of students nominating the specific student as a friend will provide an estimate of segregation and a fixed effects regression model will be used for the correlation estimation. This chapter will specify the computation of the measure of segregation and the regression model used and shortly describe the data modification process needed to calculate the segregation level.

3.5.1 Measuring Segregation

The measure used relies on very simple mathematics and gives a measure of segregation at individual level as well as an index of group-level segregation.

The measure is calculated using information on the friends of each student i . The friend information stems from two sources: self-reported friends and persons reporting the student to be their friend. The latter is in what follows referred to as friendship nominations. These two sources form two separate ways to calculate the measure. The friends are divided into two mutually excluding groups, students with immigrant background (I_i) and students with non-immigrant background (N_i). The total number of friends is denoted T_i . This number includes all friends of student i . Given this information, we can calculate the proportion of friends with a specific background. To allow for comparison between students and classes, the friend proportion value is normalized using the share of immigrants in the class, c_i . The formula for this is presented in Equation 1.

$$q_i = \frac{I_i}{T_i} - c_i \quad (1)$$

The measure achieved ranges from -1 to 1, with -1 being the impossible extreme where everyone in the class are immigrants, but none of your friends are. The value 1 indicates full segregation, where all your friends are immigrants, but the class share equals 0. This is similarly not possible. The value 0 occurs in the case where the share of immigrants among your friends equals the share of immigrants in the class. In other words, the higher value, the higher share of the students friends have immigrant background, and the more segregated is the student. A value below 0 indicates less friends with immigrant background than what could be expected based on class composition. A value above 0 stands respectively for the opposite, more friends with immigrant backgrounds than the class composition would suggest.

3.5.2 The Regression Model

The second research question of this thesis explores a possible relationship between academic skills and segregation, and age at arrival to Finland and segregation. This correlation will be studied for students with immigrant backgrounds using a regression model. In an optimal setting, the goal would be to find the factors leading to and creating segregation. However, this study does not in this stage allow for causal interpretations. It is easy to find more than one identification threat, including at least reverse causality and endogenous variables, ruling out the possibility for causal conclusions. A regression model can still be used to identify correlation and estimate the relationship between variables, which is what is done.

The focus will be on the correlation of three background variables with the outcome variable defining individual segregation, the proportion of friends with an immigrant background calculated using Equation 1. This will allow us to identify these variables correlation with segregation, if some. Equation 2 presents the regression used, a model allowing for fixed in-class and in-school effects. This is based on the assumption that school- or class-specific factors such as the teacher or the size of the school may have an effect on segregation that in this case is to be held outside the analysis.

$$q_i = \beta X_i + \gamma_c + \delta_s + \epsilon_i \quad (2)$$

q_i is the segregation index for individual i , computed as described above. X_i represents the independent variable of interest, in this case age at arrival or academic skills, the latter separated into either language skills or mathematical skills. The variable γ_c and δ_s are fixed effects determined by the class or school respectively. The actual estimation is done using the `reghdfe`-command in Stata, allowing for fixed effects without any additional data modification.

4 Results

As the two-parted research question suggests, the results will be presented in two parts. Using theories and methods described previously, the segregation level is measured using data collected on friendships. This provides an estimate of the individual segregation level for all the students participating in the survey. Using the level of segregation, correlation with other factors can be studied and described. This chapter presents the results from the analysis conducted, focusing on the level of segregation, and the correlation of individual segregation among students with immigrant backgrounds with age at arrival, skills in mathematics, and skills in the Finnish language.

The detailed data on the social networks of the students allows for an exact analysis of the individual level segregation based on friendship. As mentioned, due to bureaucratic issues, access to administrative data has at the time of writing the thesis not been provided. This creates uncertainty on immigrant background and age at arrival, as the results are dependent on the self reported immigrant backgrounds of the students, which suffers from a relatively large share of missing data. The fact that the data is self-reported should not be an issue in this case, as the students responding can be expected to have a good idea of the birth countries of them selves and their parents. This thesis thus studies segregation using friendship and background data gathered from students who did answer to the whole survey. To be able to conduct the analysis, students with for us unknown background are totally excluded. These are the students that did not answer to the questions on background, due to unknown reasons. This also implies that students with unknown backgrounds are removed from the friendship networks of the students included in the analysis. There is no reason to assume that the group left out of the analysis is equal in unknown factors to the group with known background data included in the analysis, and some bias is included in the measures. However, the class share of immigrants is counted based on the full number of class members and the known number of students with immigrant backgrounds. As the patterns found regarding segregation levels are clear, it is reasonable to assume that it is not

only caused by bias in the data, and the results can be interpreted thereby.

Figure 6: Differences in segregation measure depending on direction of reporting friendships, visualized through density differences.

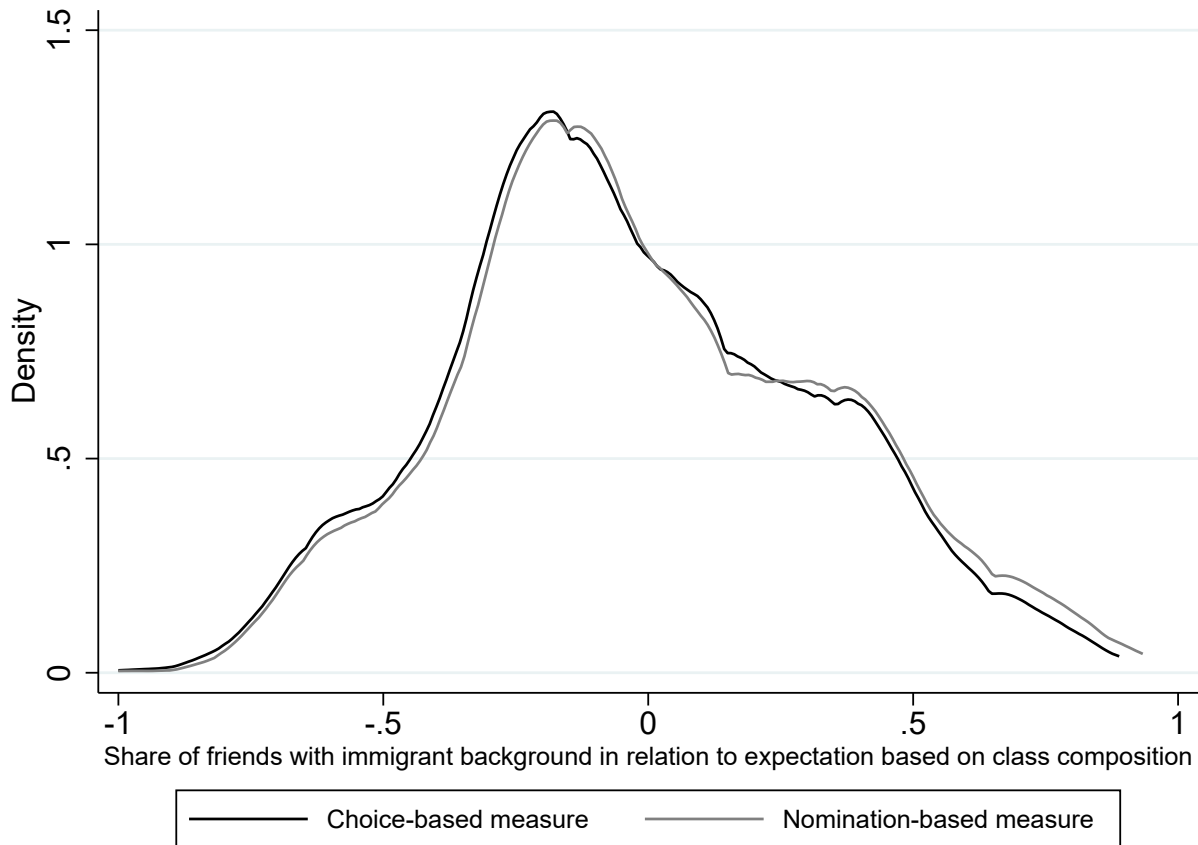


Figure Note: The density of the share of friends with immigrant backgrounds in relation to expectation based on class composition, using two different ways of reporting friendship. The choice-based measure relies on reported close friends by the students themselves. The nomination-based measure relies on the number of other students choosing the student in question as their close friend. (n = 2789)

4.1 Segregation Levels

The main question this thesis tries to answer is whether immigrants in Finnish schools experience segregation on an individual level. This is done using a segregation measure based directly on the reported friends of the students, and the background of these. The friends reported can be counted in two ways, either based on the friends each student reports, or based on the friends that report a specific student as their friend, here referred to as friend-

ship nominations. To add reliability, the measure of segregation is first calculated using both ways of interpretation. There are no significant differences visible between the segregation found using the two ways of measuring, as is seen in Figure 6 and Figure 7. Even though the friends chosen and the students choosing you are not completely identical, the difference is so small that only one of the measures will be needed. In line with Alan, Duysak, Kubilay and Mumcu (2021) using similar methods, the nomination measure is chosen.

Figure 7: Differences in segregation measure depending on direction of reporting friendships, visualized through a scatterplot.

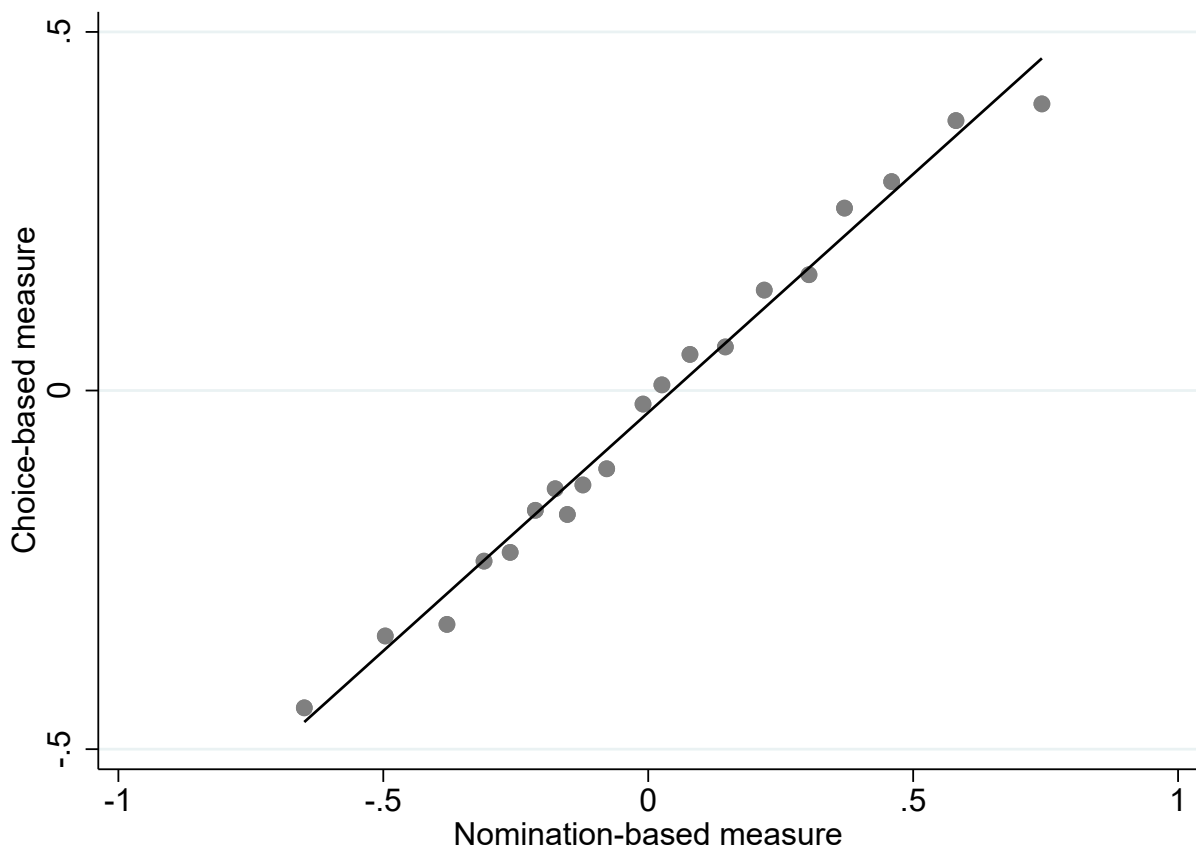


Figure Note: A scatterplot of the share of friends with immigrant backgrounds in relation to expectation based on class composition, comparing two different ways of reporting friendship. The choice-based measure relies on reported close friends by the students themselves. The nomination-based measure relies on the number of other students choosing the student in question as their close friend. (n = 2789)

Thus, using the friendship nominations and a division of students into the two groups: *Students with immigrant backgrounds* and *Students with non-immigrant backgrounds* (according

to the criteria set in chapter 3.4.1), the proportion of friends with immigrant backgrounds is calculated using Equation 1. This accounts for differences in class share of immigrants and can therefore be used to compare students in different classes. Table 3 shows the mean normalized proportions for the two groups, including also a value for all participating students. The two groups are mutually exclusive, and the full sample consists only of students belonging in either of the two groups. For students with immigrant background, the measure indicates quite intuitively segregation. For students with non-immigrant backgrounds, this is a measure of the share of total friends having immigrant background, in comparison to what could be expected based on the class share of students with immigrant backgrounds. For this group, the index value does not represent segregation. The main contribution of the segregation level among students with non-immigrant backgrounds is to deliver something to compare the segregation among students with immigrant backgrounds to. It is obvious that a difference exists, but it is also clear that there is great variation among both groups. Students with immigrant backgrounds seem to be friends with other students with immigrant backgrounds to a greater extent than students with non-immigrant backgrounds are, but the standard deviation is of such magnitude that no definite answers can be given based on these numbers.

Table 3: The share of friends with immigrant backgrounds in different groups, in relation to expectation based on class composition

	mean	SD	Obs
All students	-0.01	(0.35)	2,127
Non-immigrant background	-0.07	(0.33)	1,318
Immigrant background	0.10	(0.36)	809

Table Note: Mean values and standard deviation for share of friends with immigrant backgrounds in two different groups, taking into account the class composition, calculated using Equation 1 (Section 3.5.1).

The mean of the segregation measure indicates difference between the groups. This difference is further confirmed when the density level of the measure of segregation is scrutinized graphically. This is seen in Figure 8. As previously mentioned, segregation grows as the measure of segregation grows. A segregation value below zero implies having fewer friends with

immigrant backgrounds than the class composition would suggest. A positive segregation value is given to students with more friends with immigrant backgrounds than expected based on the students in the class. Given this, the right tail consists of students being more individually segregated, and it is clear that students with immigrant backgrounds are over-represented here. As the mean values suggested, variation inside the groups is large. Many students are located just below zero, having a little smaller share of friends with immigrant backgrounds than the class share of students with immigrant backgrounds. However, an even larger share of the students with immigrant backgrounds, represented by the red full line, is found on a segregation level above 0.5. This finding indicates that the students with immigrant backgrounds actually are segregated, at least at some level, a finding supported by previous studies, as discussed in chapter 2.

A fixed effect regression model similar to the one presented in Equation 2 further adds validity to the hypothesis that students with immigrant backgrounds have more friends with similar backgrounds and thus are more segregated. In Table 4 we see that, accounting for differences within schools and classrooms, having an immigrant background increases the value of the segregation index with 20 %, a statistically significant increase. This gives numerical support to the tendency shown in Figure 8. Based on the sample used, there is segregation on an individual level.

Table 4: Relation between immigrant status and share of friends with immigrant backgrounds in relation to expectation based on class composition.

	Segregation
Immigrant background	0.209*** (0.017)
Observations	2,126

Table Note: Result of regressing share of friends with immigrant background (see Equation 1, Section 3.5.1) on immigrant background of the student in question.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Figure 9 aims to replicate previous research, estimating the correlation between class com-

Figure 8: Density distribution of nomination-based segregation measure, over three groups.

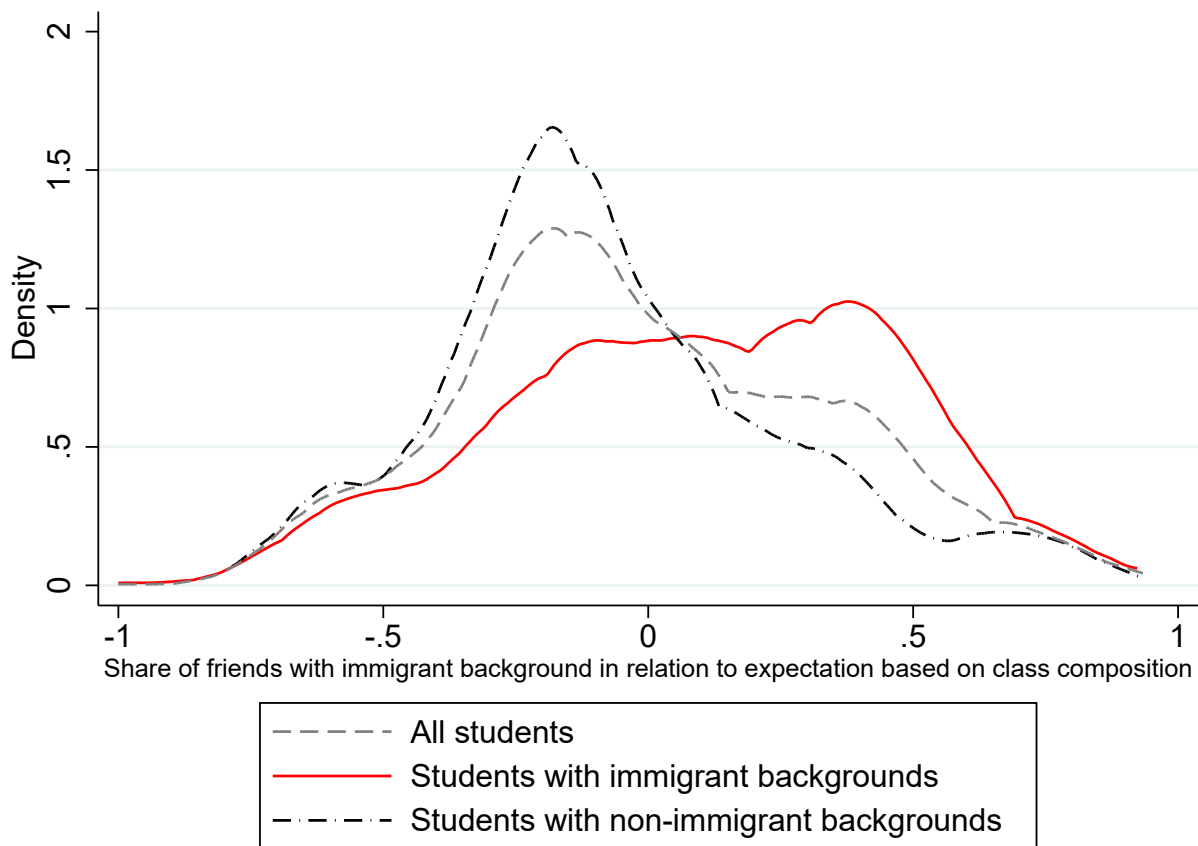


Figure Note: Density distribution of the segregation measure computed using Equation 1 (Section 3.5.1), for two groups and the whole sample. (n = 2789)

position and individual segregation among the students in the class. This has been done in several previous studies (Echenique and Fryer Jr., 2007; Holfve-Sabel, 2015; Mele, 2020), with partly different results. It seems that individual segregation in Finnish schools decrease as the proportion of immigrants in the class increases. This trend is statistically significant. As segregation is difficult to measure as an absolute matter, the change in segregation as some outside factor, such as class composition, changes is of great interest, both for the matter of knowledge and understanding, and for policy interest. As previously mentioned, results from other countries have shown different tendencies in individual level segregation as the class composition changes (Echenique and Fryer Jr., 2007; Holfve-Sabel, 2015; Mele, 2020). Data and methods used in this thesis provide results closer to what is previously found in the Nordic countries (Holfve-Sabel, 2015) than to American results (Mele, 2020;

Echenique and Fryer Jr., 2007). It seems that class composition does not matter much when it comes to individual segregation, at least not in the same way as it does in an American context, where segregation rises rapidly as the minority's share in the class grows, to achieve almost full segregation already at an early stage. Mele (2020) conclude that classes with an equal share of students with different background would be the best for segregation, which intuitively makes sense. Our results show that individual segregation decreases as the class share of immigrants increases. This can be a consequence of how the measure is constructed. With more students with immigrant backgrounds in the class, the expected proportion of students with immigrant backgrounds among friends rises and it becomes more difficult to have a higher share of own friends than the class share. It could also be explained by differences in the age of the students studied, where the Nordic students have been younger than the high school students in the U.S. This said, the conclusion must still be that class composition does not matter much for individual segregation, and the pattern in Finland is different than the pattern found in various studies in the U.S.

To further validate the finding of individual segregation among students in the sample, Table 5 present the expected composition of friends given the composition of different groups of Finnish 11 year old. On a population level, the expected share of immigrant friends for a non-immigrant child is 10 %. The expected share of non-immigrant friends is thus naturally 90 %. Columns 1 and 2 in Table 5 present the expected and observed share of non-immigrant and immigrant background friends respectively for four levels of sample data. Columns 3 and 4 repeat the calculation for students with immigrant background. The share of immigrants increases when we only look at our sample in comparison to the full population, as the schools chosen for the study has a higher immigrant share than the population. Approximately 20 % of Finnish 11 year old with immigrant backgrounds are included in the sample, whereas the share of all students is roughly 10 %. However, the composition continues to change when moving to school and class level, indicating that segregation exists between schools and classes. A student with immigrant background is expected to have 40 % immigrant friends and 60 % friends with non-immigrant background based on the composition

Figure 9: Correlation between individual segregation level and class share of students with immigrant backgrounds.

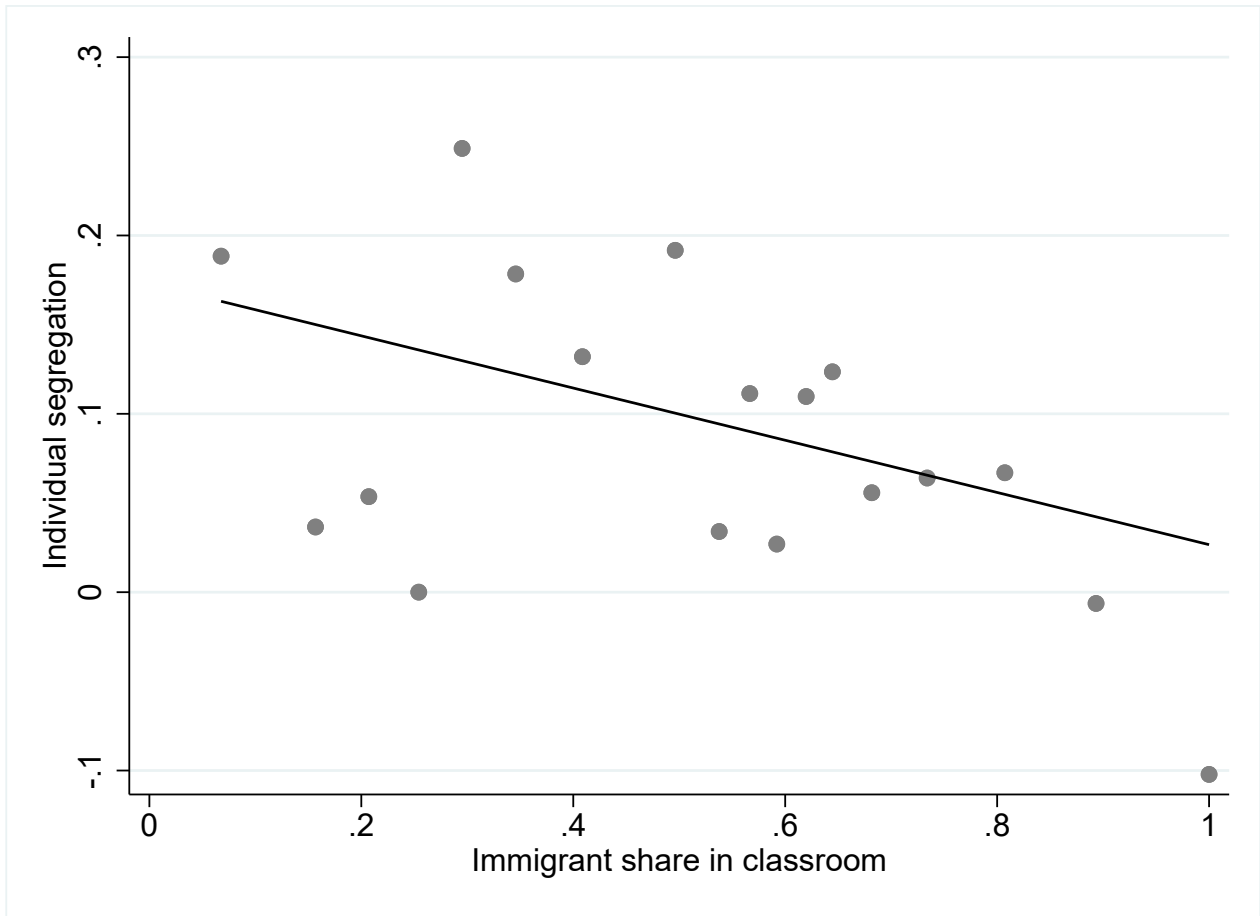


Figure Note: Correlation between class composition and share of friends with immigrant background (Equation 1, Section 3.5.1), including only students with immigrant backgrounds. As only students with immigrant backgrounds are included, the normalized share of friends with immigrant background can be interpreted as a measure of individual segregation. (n = 1139)

of the full sample, but taking into account class composition, the expected share of friends with immigrant background rises to 45 %.

The last row in Table 5 gives clear indication that segregation increases as individual students make friends. A student with immigrant background is observed to have a 61 % share of friends with immigrant backgrounds. Compared to the expected sample share of 40 % this is a high increase, and it can not be explained by between-school or between-class segregation, as the expected shares for schools and classes are closer to the full sample share than the observed share. A rough estimation of the share of immigrant friends of students

Table 5: Expected and observed immigrant share of friends among 11 year old on different levels in the society.

	Non-immigrant - non immigrant (1)	Non-immigrant - immigrant (2)	Immigrant - non-immigrant (3)	Immigrant - immigrant (4)
Expected values				
Full sample	0.60	0.40	0.60	0.40
Between-school	0.64	0.36	0.52	0.48
Between-class	0.69	0.31	0.55	0.45
Observed values				
Individual level	0.73	0.27	0.39	0.61

Figure Note: The table presents expected friend compositions assuming random friend choice. The full sample simply calculates the share of students with different backgrounds in the sample. Between-school shares take into account the actual composition between schools, randomizing friend choice inside schools. Between-class shares repeat this on a classroom level, taking into account the distribution of students between classes, assuming random friend choice within the class. Individual level shares show the observed friend group compositions. (n=2789)

with immigrant backgrounds on a population level would be based on these figures and population size be 20 %. Compared to the full sample composition, this is a significant increase. To validate this, further research must be undertaken. Despite the inexactness, it serves as an indicator of individual segregation and shows that the phenomena exists. The population estimate is very rough, but the observed values are exact enough to prove that segregation exists on an individual level.

4.2 Segregation in Relation to Other Factors

If segregation did not affect other factors in life and society, the measuring of it would be of a much lower interest. However, it is clear that in the big picture, segregation leads to lower quality of life through several aspects (Galster and Sharkey, 2017). Therefore, the second research question this thesis tries to answer is the question of a possible correlation between segregation level and academic performance, and segregation level and time spent in Finland, in other words age at arrival to Finland, in the case of immigrant students. This section searches for answer to that question in two parts, focusing solely on students with immigrant backgrounds. Firstly, the correlation between segregation and age at arrival to

Finland is investigated. Secondly, we study the possible correlation between segregation and academic skills, more exactly mathematical skills and verbal skills.

Table 6: Correlation between segregation and other factors, with standard error in parenthesis.

	Individual Segregation		
	(1)	(2)	(3)
Age at arrival	0.020 (0.013)		
Verbal skills		-0.013 (0.015)	
Mathematical skills			-0.001 (0.013)
Observations	344	586	595

Table Note: Regression coefficients from regression model presented in Equation 2 (Section 3.5.2), regressing age at arrival, verbal skills and mathematical skills respectively on the share of friends with immigrant background in relation to expectation based on class composition. The regressions include only students with immigrant backgrounds, letting us interpret the friend share as individual segregation.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.2.1 Correlation with Age at Arrival

Relying on previous research, among them Bleakley and Chin (2010), it is easy to expect a strong positive correlation between age at arrival to Finland and segregation among students with immigrant backgrounds. Using the data gathered for this study, there is however little or no evidence of this. As can be seen in Table 6, no statistically significant relation between the school grade to which the student enrolls to, based on their age, and the level of segregation the way it is measured in this thesis exists. The estimated effect of arriving one year later is an added segregation of 0.02, but with a standard error of 0.013, and a small sample, the effect could as well be non-existing. Figure 10 indicates a similar trend, but is also extremely shattered. As it is based on the same regression model, it is also clear that the trend visible is not statistically significant. Obviously, the lack of exact background data affects this part of the analysis severely, and no correlation is found.

The problem with missing data is influencing all parts of the results, but the greatest loss is

Figure 10: Correlation between individual segregation level and the grade during which the student arrived to Finland.

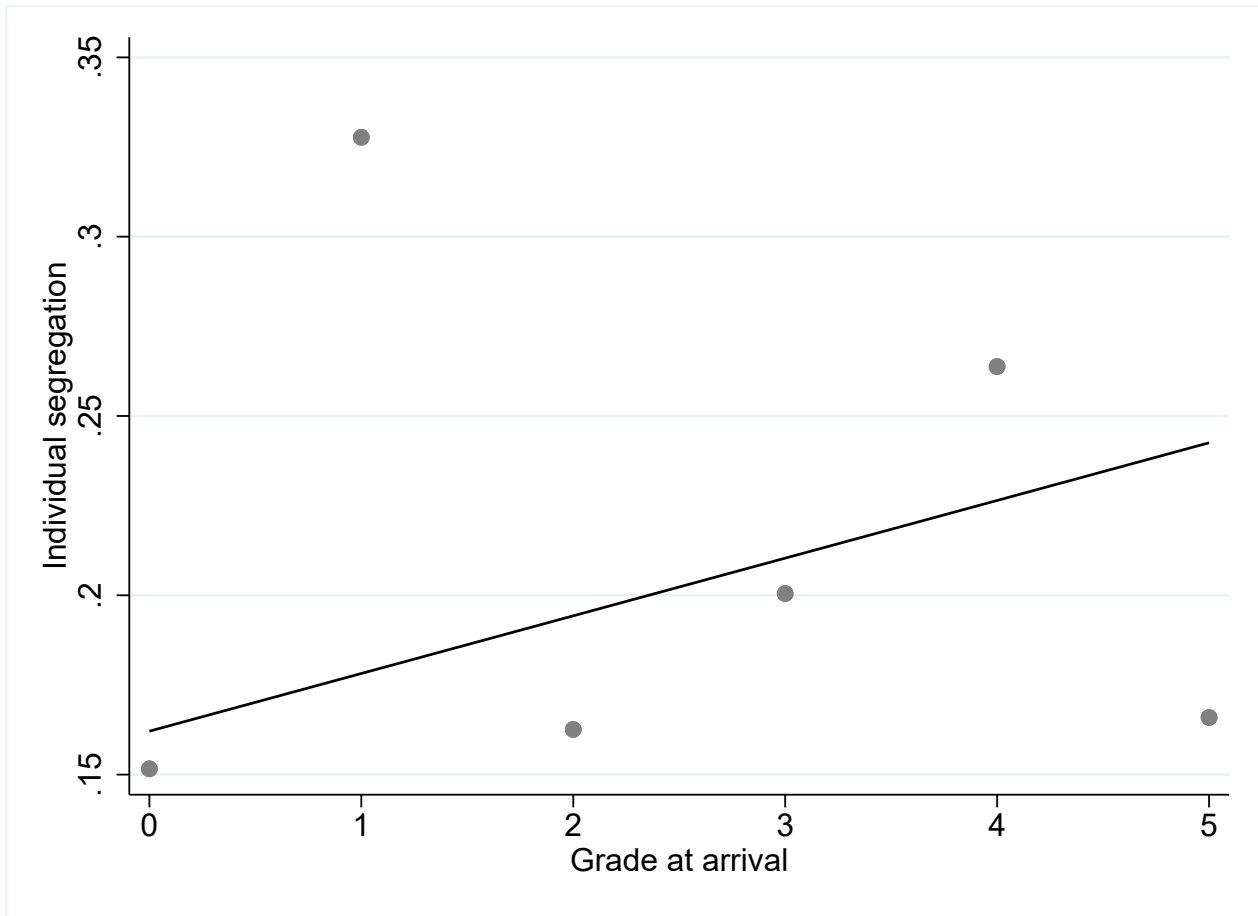


Figure Note: Visual presentation of the correlation between age at arrival and individual segregation. The individual segregation is measured as share of friends with immigrant background according to Equation 1 (Section 3.5.1), which can be interpreted as individual segregation as only students with immigrant backgrounds are included. The correlation is based on the regression model presented in Equation 2 (Section 3.5.2). (n = 344)

perhaps seen in this part, studying the effects of age at arrival to Finland. Finnish administrative data includes detailed information on arrival age. As this data has not been accessed, information on age at arrival is taken from the survey. Apart from risks with self-reported data, which might be more serious regarding this question than regarding the question of birth country, the exact survey question gathering information on this has also been simplified and asks only during which grade the student arrived to Finland. For students arriving before first grade, the answer to this question is "Before school". Obviously, there is still a great difference between coming as a newborn or as a five year old preschooler (Bleakley and Chin, 2010). Apart from this aspect being left out, the scarce information also leads to

a very small sample. A majority of the students with immigrant backgrounds has arrived before school or are born in Finland by parents with immigrant backgrounds. This leaves us with less than 400 students with appropriate data for this question. It would be of high interest to repeat the analysis with broader data, as this could add validity and allow us to make conclusions on the effect of arriving at different ages. This is especially interesting as Bleakley and Chin (2010) show significant correlation between age at arrival and segregation, and the overall effects of age at arrival are widely acknowledged (Ansala, Hämäläinen and Sarvimäki, 2019; Hermansen, 2017; Gonzalez, 2003; Lee and Edmonston, 2011). It would also be interesting to repeat the analysis with a more homogeneous group, such as students with immigrant backgrounds from a specific country or area. This would remove the current problem of heterogeneity in the group, as some of the students have moved from countries with cultures very similar to the Finnish culture, and the individual experiences in the group of students studied are likely to differ significantly.

4.2.2 Correlation with Academic Skills

Finally, this thesis aims to investigate any possible correlation between academic skills and segregation, in particular basic mathematical and verbal skills. Because of the relative easiness for the individual or the school to affect academic skills, results indicating correlation between school results and segregation could be of high policy interest. However, based on the data available for this study, no correlations are found. It seems that no correlation exists between mathematical skills and individual segregation. This is visible in Table 6 as well as in Figure 11. The mathematical score is fairly evenly distributed in relation to the x-axis and the individual segregation level also shows no obvious patterns. Thus, mathematical skills do not correlate with individual segregation.

Similarly as with correlation with mathematical skills, no statistically significant relation between individual segregation and language skills is found. In this case, the slight negative relation is larger than in the case of mathematics, but the standard error is even larger, im-

Figure 11: Correlation between individual segregation level and mathematical skills.

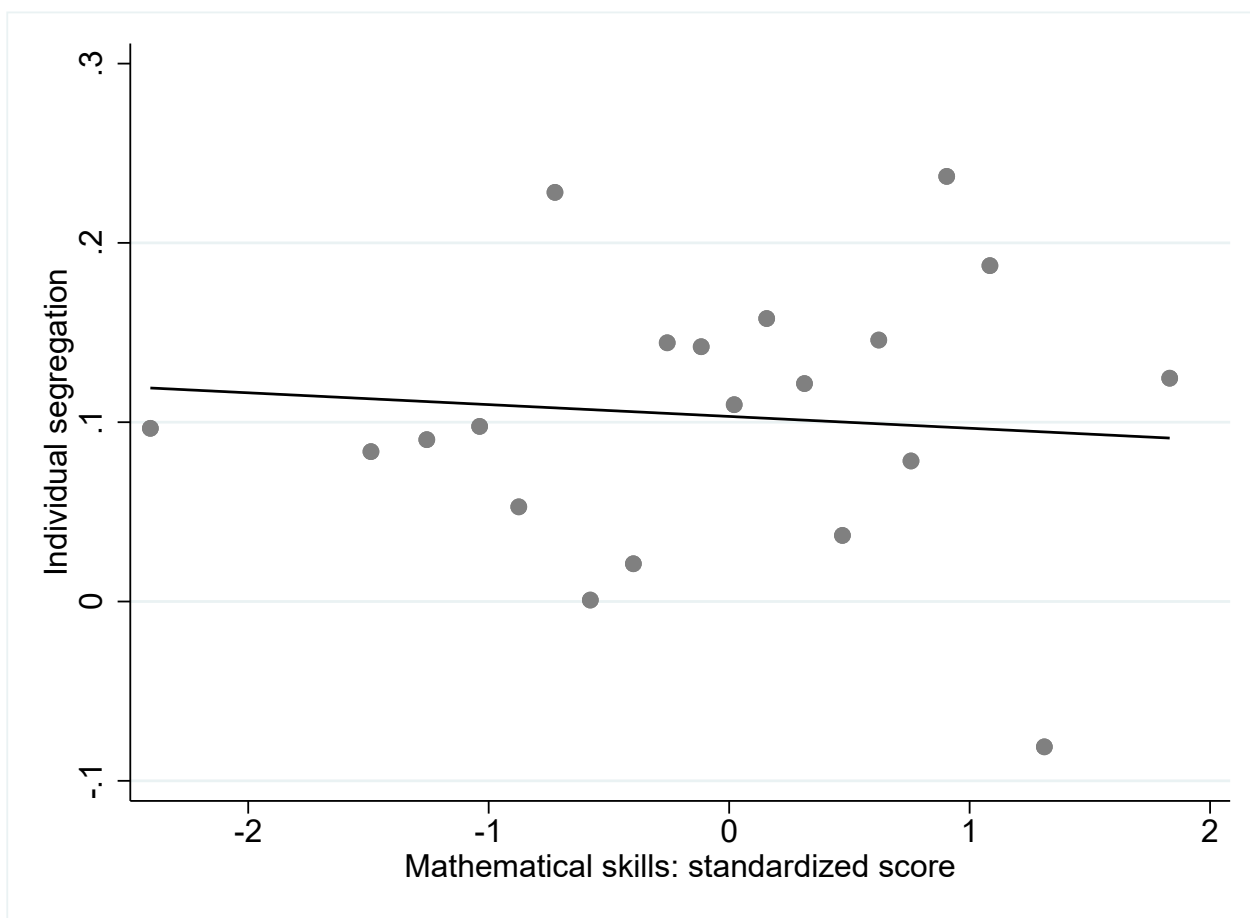


Figure Note: Visual presentation of the correlation between mathematical skills and individual segregation. The individual segregation is measured as share of friends with immigrant background according to Equation 1 (Section 3.5.1), which can be interpreted as individual segregation as only students with immigrant backgrounds are included. The correlation is based on the regression model presented in Equation 2 (Section 3.5.2). (n = 595)

plying no statistical difference from zero (see Table 6). Figure 12 emphasizes this. It is clear, as seen in section 3 previously, that language skills are more even and the differences from zero are lower compared to the mathematical score. However, the differences in individual segregation persist and if drawing by hand the line indicating the correlation could be put close to anywhere in the picture.

Results on correlations with segregation found in this thesis are not significant and would most likely gain remarkably from increased data access. That would however not make them completely simple to interpret either way. The study undertaken has focused only on

Figure 12: Correlation between individual segregation level and language skills.

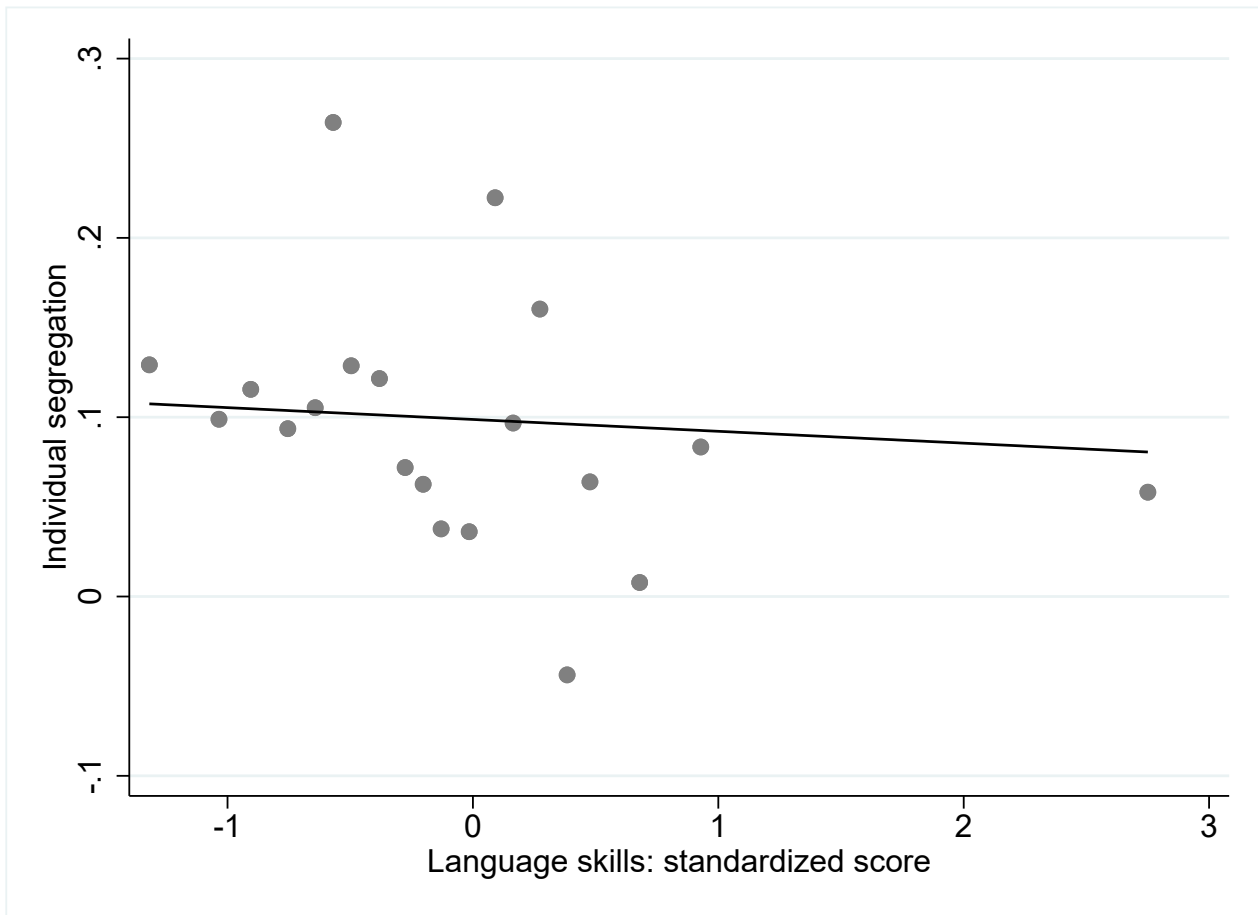


Figure Note: Visual presentation of the correlation between verbal skills and individual segregation. The individual segregation is measured as share of friends with immigrant background according to Equation 1 (Section 3.5.1), which can be interpreted as individual segregation as only students with immigrant backgrounds are included. The correlation is based on the regression model presented in Equation 2 (Section 3.5.2). (n = 586)

correlation between segregation and other factors. This is an informed choice, in order to avoid the difficulties following causal interpretation. It would be of large value to be able to present results saying for example that good language skills decrease segregation. This might be the case, but is in no way proven in this thesis, and cannot be proven with the methods used. As tempting it is to interpret the correlations causally, as easy it is to find reasons why this cannot be done. The main threat is reverse causality. Having good Finnish language skills could make it easier to find friends with non-immigrant backgrounds, compared to a case where the skills in Finnish are poor. However, language skills improve in use, and it is also very likely that a student with poor Finnish language skills could experience great

improvement of the skills, if this student spent time with students possessing a great Finnish language. The same holds at least partly for other academic skills, such as mathematical skills studied in this thesis. To avoid this, some identification method has to be used, ruling out all other effects and impacts of language skills and other individual features, to be able to find a causal relationship. This has not been done, and this thesis studies only correlation.

It is obvious that differences in individual segregation exist, but it is far from clear what they stem from. This thesis finds little or no correlation with all factors tested, including class composition, age at arrival to Finland, language skills and mathematical skills. The difference in segregation is too large to be explained by random friend choice, thus there must exist some reasons behind it. This thesis did not aim to find the reasons, but has not succeeded in finding even a correlating factor.

5 Conclusion

The aim of this thesis was to answer the two research questions:

1. Are immigrants within Finnish schools segregated on an individual level?
2. How does the within-school segregation of immigrants relate to their academic performance and spent time in Finland?

After conducting the analysis, it is clear that the answer to the first research question is simply *Yes*. Students with immigrant backgrounds show segregation on an individual level, meaning that they are more likely to be friends with other students with immigrant backgrounds than their fellow students with non-immigrant background, also when ruling out possible effects of the class and school environment. As segregation is rather complicated to measure accurately, and the ways to measure it vary heavily, the exact level of segregation is more difficult to provide. Comparing to results by Alan, Duysak, Kubilay and Mumcu (2021), students with immigrant backgrounds in Finnish schools seem to be a little less segregated than refugee students in Turkish schools. However, individual segregation does exist. To fully understand its level and the experienced impact of this segregation, further research must be conducted, not only with more accurate data, but perhaps also from a qualitative perspective.

Regarding the second research question, this thesis did not find any significant relation between the used measure of individual segregation and academic performance of students with immigrant backgrounds. The reason for this could be the difficulties met due to the missing data. If that is not the case, it might be that peer effects in friend groups not are that large, given that fixed effects of school and class choice are taken into account. Similarly, the study undertaken did not show any relation between time spent in Finland and individual segregation. In this case it is likely that the small sample makes it impossible to identify any such relation, and that a relation in fact does exist. Further research can provide more

exact results.

It was clear that little research on this field had been done prior to this study. This thesis can therefore contribute with results on individual segregation in Finnish schools not previously found. The knowledge of the existence of segregation between friend groups can be valuable for improving teaching and inclusion policies, all in order to provide better education and more equal settings for all students. That said, there could, and should, still be more research done on this subject. The most important research technical improvement needed is the conducting of a similar analysis using administrative data. This should not be difficult and will most likely be possible in a near future. With that done, it will be possible to study the details of individual segregation regarding gender, background country, and family background, to name a few factors.

Further knowledge on individual within-school segregation could help create policies to improve the academic results of immigrants in Finnish schools and further increase employment and improve equality on the Finnish labor market. This thesis has shown that individual level segregation exists, but it would still be of great value to understand more about the reasons behind it and its impact.

References

- Ahmad, Akhlaq. 2020. "Ethnic discrimination against second-generation immigrants in hiring: empirical evidence from a correspondence test." *European Societies* 22(5):659–681.
- Alan, Sule, Ceren Baysan, Mert Gumren and Elif Kubilay. 2021. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking." *Quarterly Journal of Economics* 136(4):2147–2194.
- Alan, Sule, Enes Duysak, Elif Kubilay and Ipek Mumcu. 2021. "Social Exclusion and Ethnic Segregation in Schools: The Role of Teacher's Ethnic Prejudice." *Review of Economics and Statistics* (Forthcoming).
- URL:** <https://static1.squarespace.com/static/5b75fc713917ee2070a01bfb/t/6001b48b5b22df1761e34a6a/1610724509849/AlanDuysakKubilayMumcu.pdf>
- Allen, Rebecca and Anna Vignoles. 2007. "What should an index of school segregation measure?" *Oxford Review of Education* 33(5):643–668.
- Ansala, Laura, Ulla Hämäläinen and Matti Sarvimäki. 2019. "Age at arrival, parents and neighborhoods: understanding the educational attainment of immigrants' children." *Journal of Economic Geography* 20(2):459–480.
- Ballester, Coralio and Marc Vorsatz. 2014. "Random Walk-Based Segregation Measures." *The Review of Economics and Statistics* 96(3):383–401.
- Berisha, Anna-Kaisa and Piia Seppänen. 2017. "Pupil selection segments urban comprehensive schooling in Finland: composition of school classes in pupils' school performance, gender, and ethnicity." *Scandinavian Journal of Educational Research* 61(2).
- Bernelius, Venla and Heidi Huilla. 2021. Educational equity, social and spatial segregation and opportunities for targeted support. Finnish Government 2021:7 Finnish Government.
- Bleakley, Hoyt and Aimee Chin. 2010. "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants." *American Economic Journal: Applied Economics* 2(1):165–92.

-
- Bojanowski, Michal and Rense Corten. 2014. "Measuring segregation in social networks." *Social Networks* 39:14–32.
- Bursztyn, Leonardo, Thomas Chaney, Tarek Alexander Hassan and Aakaash Rao. 2021. The Immigrant Next Door: Exposure, Prejudice, and Altruism. Working Paper 28448 National Bureau of Economic Research.
- CEDEFOP. 2020. Finland 2020. NQF Country Report. CEDEFOP European Center for the Development of Vocational Training.
- Charles, Camille Z, Gniesha Dinwiddie and Douglas S Massey. 2004. "The Continuing Consequences of Segregation: Family Stress and College Academic Performance." *Social science quarterly* 85(5):1353–1373.
- Coleman, James. 1958. "Relational Analysis: The Study of Social Organizations with Survey Methods." *Human organization* 17(4):28–36.
- Currarini, Sergio, Matthew O. Jackson and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica* 77(4):1003–1045.
- Duncan, Otis Dudley and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20(2):210–217.
- Echenique, Federico and Roland G. Fryer Jr. 2007. "A measure of segregation based on social interactions." *Quarterly Journal of Economics* 122(2):441–485.
- Freeman, Linton. 1978. "Segregation in Social Networks." *Sociological Methods & Research* 6:411–429.
- Galster, George and Patrick Sharkey. 2017. "Spatial Foundations of Inequality: A Conceptual Model and Empirical Overview." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3(2):1–33.
- Gonzalez, Arturo. 2003. "The education and wages of immigrant children: the impact of age at arrival." *Economics of Education Review* 22(2):203–212.

-
- Gutiérrez, Gabriel, John Jerrim and Rodrigo Torres. 2020. "School segregation across the world: Has any progress been made in reducing the separation of the rich from the poor?" *The Journal of Economic Inequality* 18(2):157–179.
- Hermansen, Are Skeie. 2017. "Age at Arrival and Life Chances Among Childhood Immigrants." *Demography* 54(1):201–229.
- Holfve-Sabel, Mary-Anne. 2015. "Students' Individual Choices of Peers to Work with During Lessons May Counteract Segregation." *Social Indicators Research* 122(2):577–594.
- James, David R. and Karl E. Taeuber. 1985. "Measures of Segregation." *Sociological Methodology* 15:1–32.
- Kosunen, Sonja, Venla Bernelius, Piia Seppänen and Miina Porkka. 2020. "School Choice to Lower Secondary Schools and Mechanisms of Segregation in Urban Finland." *Urban Education* 55(10):1461–1488.
- Kupiainen, Sirkku; Hautamäki, Jarkko; Karjalainen Tommi. 2009. The Finnish education system and PISA. Ministry of Education Publications, Finland 2009:46 Ministry of Education and Culture.
- Lee, Sharon M. and Barry Edmonston. 2011. "Age-at-Arrival's Effects on Asian Immigrants' Socioeconomic Outcomes in Canada and the U.S." *International Migration Review* 45(3):527–561.
- Matikka, Anni, Pauliina Luopa, Hanne Kivimäki, Jukka Jokela and Reija Paananen. 2015. The well-being of eight-year and ninth-year pupils with an immigrant background - School Health Promotion Study 2013. Discussion Paper 6/2015 National Institute for Health and Welfare.
- Mele, Angelo. 2020. "Does School Desegregation Promote Diverse Interactions? An Equilibrium Model of Segregation within Schools." *American Economic Journal: Economic Policy* 12(2):228–57.

-
- OECD. 2017. Finding the way: A discussion of the Finnish migrant integration system. Technical report International Migration Division.
- OECD. 2019. *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD Publishing.
- Oppimisanalytiikan Keskus. 2022. "FUNA."
URL: <https://www.oppimisanalytiikka.fi/ville/funa/>
- Philipson, Tomas. 1993. "Social Welfare and Measurement of Segregation." *Journal of Economic Theory* 60(2):322–334.
- Räsänen, Pekka, Pirjo Aunio, Anu Laine, Airi Hakkarainen, Eija Väisänen, Jonatan Finell, Teemu Rajala, Mikko-Jussi Laakso and Johan Korhonen. 2021. "Effects of Gender on Basic Numerical and Arithmetic Skills: Pilot Data From Third to Ninth Grade for a Large-Scale Online Dyscalculia Screener." *Frontiers in Education* 6.
- Statistics Finland. 2021. "11vy – Population according to origin and background country, age and sex, 1990-2020."
URL: https://pxnet2.stat.fi/PXWeb/pxweb/en/Maahanmuuttajat_ja_kotoutuminen/Maahanmuuttajat_ja_kotoutuminen__Maahanmuuttajat_ja_kotoutuminen/maakoto_pxt_11vy.px/
- Strohmeier, Dagmar, Antti Kärnä and Christina Salmivalli. 2011. "Intrapersonal and Interpersonal Risk Factors for Peer Victimization in Immigrant Youth in Finland." *Developmental Psychology* 47(1):248–258.
- Säkkinen, Salla and Tuula Kuoppala. 2021. "Varhaiskasvatus 2020 : Varhaiskasvatukseen osallistuneista lapsista enemmistö osallistui kunnan kustantamaan päiväkotitoimintaan." *THL TILASTORAPORTTI 32/2021 (27.9.2021)*.
- Söderström, Martin and Roope Uusitalo. 2010. "School Choice and Segregation: Evidence from an Admission Reform." *The Scandinavian Journal of Economics* 112(1):55–76.

Varjo, Janne, Ulf Lundström and Mira Kalalahti. 2018. "The governors of school markets? Local education authorities, school choice and equity in Finland and Sweden." *Research in Comparative and International Education* 13(4):481–498.

Yao, Jing, David W.S. Wong, Nick Bailey and Jonathan Minton. 2019. "Spatial Segregation Measures: A Methodological Review." *Tijdschrift voor Economische en Sociale Geografie* 110(3):235–250.